



HAL
open science

Traitement de données incertaines : application au tri des matières plastiques

Lucie Jacquin

► **To cite this version:**

Lucie Jacquin. Traitement de données incertaines : application au tri des matières plastiques. Autre [cs.OH]. IMT - MINES ALES - IMT - Mines Alès Ecole Mines - Télécom, 2021. Français. NNT : 2021EMAL0001 . tel-04076948

HAL Id: tel-04076948

<https://theses.hal.science/tel-04076948>

Submitted on 21 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR ÉCOLE NATIONALE SUPÉRIEURE DES MINES D'ALÈS (IMT MINES ALÈS)

En Informatique

École doctorale I2S – Information, Structures, Systèmes
Portée par l'Université de Montpellier

Unité de recherche EuroMov Digital Health in Motion

Traitement de données incertaines : Application au tri des matières plastiques

Présentée par Lucie Jacquin
Le 25 mars 2021

Sous la direction de Jacky MONTMAIN
et Didier PERRIN

Devant le jury composé de

Christophe MARSALA, Professeur, Sorbonne Université - LIP6	Rapporteur
Frédéric PICHON, Professeur, Laboratoire de Génie Informatique et d'Automatique de l'Artois	Rapporteur
Anne-Laure JOUSSELME, directrice de recherche, NATO STO Centre for Maritime Research and Experimentation	Examinatrice
Sébastien DESTERCKE, Chargé de recherche CNRS, HDR, Université de Technologie de Compiègne, CNRS	Examineur, président du jury
Jacky MONTMAIN, Professeur, IMT Mines Ales	Directeur de thèse
Didier PERRIN, Professeur, IMT Mines Ales	Directeur de thèse
Abdelhak IMOUSSEN, Maître de conférences, IMT Mines Ales	Encadrant de proximité
François TROUSSET, Maître assistant, IMT Mines Ales	Encadrant de proximité



Avec l'augmentation des flux de déchets, le tri industriel est devenu un enjeu majeur. Le principal défi consiste à minimiser les erreurs de tri afin d'éviter une dégradation importante de la qualité du produit recyclé final conduisant à une recyclabilité médiocre de matériaux de seconde vie. Dans cette thèse, nous disposons de données fournies par une technologie optique dans un contexte industriel (société Pellenc ST). Cette technologie est munie d'une caméra moyen infra-rouge capable de fournir des informations quel que soit la teinte des plastiques (les caméras précédentes n'étaient pas adaptées aux plastiques sombres et noirs). Cependant, même avec les technologies d'acquisition les plus récentes basées sur l'imagerie spectrale, la reconnaissance des plastiques reste une tâche difficile pour certains types de plastique. En effet, la présence, d'une part de l'incertitude sur la nature du matériau liée à des variabilités des mesures comme le vieillissement des plastiques, l'hygroscopie de surface, l'addition d'anti-UV opacifiants (noirs de carbone), etc. associée à la présence, d'autre part, d'imprécision due à l'incomplétude de l'information, i.e. absence de certaines caractéristiques, font obstacle à l'utilisation des algorithmes classiques de classification. Au-delà de l'étape de reconnaissance, le projet s'intéresse à la phase de tri : dans quel bac le fragment doit-il être envoyé compte tenu de l'information imprécise et incertaine sur la nature de ce fragment de plastique d'une part, des objectifs et des contraintes du processus de tri d'autre part. Par ailleurs, des seuils d'impuretés sont fixés pour chaque bac.

Dans ce travail de thèse nous proposons une approche de tri adaptée aux données imparfaites comme celles obtenues dans le contexte industriel de la société Pellenc ST. L'approche proposée se compose de deux phases : une phase de classification et une phase de tri. Pour la phase de classification, nous avons proposé un algorithme de classification prudente dans le cadre des fonctions de croyance basée sur : 1) une re-labellisation partielle automatique des exemples d'apprentissage "difficiles", 2) la construction d'une fonction de masse comme information sur le type de plastique et 3) une règle de décision où le décideur pourrait introduire ses contraintes de prudence et de précision. Concernant la phase de tri, nous avons posé un problème d'optimisation dans le cadre des fonctions de croyance pour choisir le bac approprié pour un fragment donné. Ce problème consiste à minimiser la quantité de fragments rejetés (perdus pour le recyclage), maximiser le gain, i.e., pureté des plastiques chers, tout en respectant des contraintes sur les proportions d'impuretés. Les expérimentations que nous avons menées sur le tri de plastiques montrent le gain en qualité de produit recyclé que notre approche prudente apporte en comparaison d'approches déterministes ou probabilistes.

With the increase in waste flows, industrial sorting has become a major issue. The main challenge is to minimise sorting errors in order to avoid a significant degradation in the quality of the final recycled product leading to poor recyclability of second life materials. In this thesis, we have data provided by an optical technology in an industrial context (Pellenc ST company). This technology is equipped with a medium infra-red camera capable of providing information regardless of the colour of the plastics (previous cameras were not suitable for dark and black plastics). However, even with the most recent acquisition technologies based on spectral imaging, plastic recognition remains a difficult task for certain types of plastic. Indeed, the presence, on the one hand, of uncertainty about the nature of the material due to measurement variabilities such as the ageing of plastics, surface hygroscopy, the addition of anti-UV opacifiers (carbon blacks), etc. associated with the presence, on the other hand, of imprecision due to incomplete information, i.e. the absence of certain characteristics, hamper the use of classic classification algorithms. Beyond the recognition stage, the project focuses on the sorting phase : to which container should the fragment be sent given the imprecise and uncertain information on the nature of this plastic fragment on the one hand, and the objectives and constraints of the sorting process on the other hand. In addition, impurity thresholds are set for each container.

In this thesis we propose a sorting approach adapted to imperfect data such as those obtained in the industrial context of the Pellenc ST company. The proposed approach consists of two phases : a classification phase and a sorting phase. For the classification phase, we proposed a conservative classification algorithm in the framework of belief functions based on : 1) an automatic partial re-labelling of "difficult" learning examples, 2) the construction of a mass function as information on the type of plastic and 3) a decision rule where the decision maker could introduce his constraints of prudence and precision. Concerning the sorting phase, we posed an optimisation problem in the framework of belief functions to choose the appropriate container for a given fragment. This problem consists in minimizing the quantity of rejected fragments (lost for recycling), maximizing the gain, i.e., purity of expensive plastics, while respecting constraints on the proportions of impurities. The experiments we have conducted on plastics sorting show the gain in quality of the recycled product that our conservative approach provides compared to deterministic or probabilistic approaches.

Remerciements

A travers les quelques lignes suivantes, je souhaite exprimer ma reconnaissance à tous ceux qui ont contribué à l'aboutissement de cette thèse.

Ma première pensée est adressée à l'ensemble de mes directeurs et encadrants de thèse. Mes directeurs, Jacky Montmain et Didier Perrin, m'ont accordé leur confiance et se sont montrés enthousiastes dès les premiers moments de ma thèse. Jacky, je te remercie pour tes précieux conseils pour l'orientation de mes travaux. Didier, merci de m'avoir enseigné avec patience les enjeux du contexte applicatif de ma thèse. Mes encadrants Abdelhak Imoussaten et François Troussset ont également été d'une grande aide. Je vous remercie tous les deux pour votre implication et votre bienveillance. Je tiens particulièrement à remercier Abdelhak Imoussaten pour sa disponibilité et son sang-froid dans mes moments d'interrogation. Enfin je vous remercie tous les quatre, pour m'avoir permis de participer à des échanges scientifiques interdisciplinaires enrichissants et m'avoir formée au métier de chercheur.

En second lieu, je tiens à remercier Sébastien Destercke pour son apport scientifique dans le cadre de sa collaboration avec mon encadrement de thèse. Cette collaboration a été fortement favorisée par le programme de visites de doctorants du Gdr IA qui m'a donnée l'opportunité d'effectuer une visite au laboratoire Heudiasyc de Université technologique de Compiègne.

Je tiens aussi à remercier vivement les rapporteurs Christophe Marsala et Frédéric Pichon et les examinateurs Anne-Laure Jousset et Sébastien Destercke, pour avoir accepté de se consacrer à l'évaluation de mon travail et pour leur commentaires et questions lors de ma soutenance. Je remercie particulièrement les rapporteurs pour leurs nombreuses remarques qui laissent envisager de nouvelles perspectives de recherches, et Sébastien Destercke pour m'avoir fait l'honneur de présider ce jury.

Ma reconnaissance va également à Antoine Bourelly, directeur scientifique de la société Pellenc st, pour m'avoir donné l'occasion de participer à un projet enthousiasmant dans le domaine du recyclage. Je souhaite également remercier Nathalie Pedoussaut et Kévin Alazet qui m'ont permis de travailler avec des données réelles.

J'aimerais remercier chaleureusement les membres des centres d'enseignement et de recherche CERIS et C2MA de l'école des Mines d'Alès. En particulier, Charles Signoret, qui a lui aussi mené à bien une thèse sous la direction de Didier Perrin en collaboration avec la société Pellenc st, et qui m'a aidée à saisir de nombreuses notions liées au contexte applicatif de ma thèse. Par ailleurs, j'étais heureuse de partager des moments conviviaux avec mes collègues et amis, notamment, Cécile l'Héritier, Frank, Alexandre, Quentin, Jihane, Yu, Rayane, Kaoutar,

Emilie, Behrang, Valentina et Aliénor.

Je souhaite aussi remercier ma professeure de master et amie Cécile Mercadier pour son grand soutien et son dévouement ! Elle m'a donné les clés pour réussir et m'a démontré les vertus de la persévérance dans la recherche.

Enfin, je remercie tous mes proches pour leur soutien. Notamment ma grand-mère Janine pour sa recette de la crème aux œufs (page [115](#)), et Yusuke !

Résumé		iii
Abstract		v
Remerciements		vii
Introduction générale		1
1 Cadre général		5
1.1	Enjeux du tri/de l'identification de plastiques	6
1.2	Technologies optiques pour le tri/l'identification de plastiques	9
1.3	Acquisition des données et objectifs du tri	12
1.3.1	Contexte d'acquisition des données	12
1.3.2	Qualité des données	14
1.3.3	Objectifs du processus du tri	17
1.4	Orientation de la thèse	17
1.4.1	Attitude prudente dans le tri des plastiques	19
1.4.2	Problème d'optimisation du tri	20
1.5	Conclusion	21
2 Classification à partir de données imprécises et/ou incertaines		23
2.1	Théories de l'incertain	26
2.1.1	Théorie des probabilités	27
2.1.2	Probabilités imprécises	30
2.1.3	Fonctions de croyance	31
2.1.4	Théorie des possibilités	36
2.1.5	Espérance mathématique et mesures floues	37
2.1.6	Logique floue	38
2.2	Classification automatique supervisée	40
2.2.1	Principe de la classification supervisée	40
2.2.2	Mesures d'évaluation	41
2.2.3	Algorithmes de classification supervisée usuels	43
2.2.4	Optimisation des classifieurs	46
2.2.4.1	Ré-équilibrage de classes	46

2.2.4.2	Gestion des exemples aberrants	46
2.2.4.3	Extraction de caractéristiques	47
2.3	Gestion d'incertitudes en classification automatique supervisée	48
2.3.1	Classifieur probabiliste	49
2.3.2	Classifieur évidentiel	50
2.3.3	Classifieur flou	53
2.4	Illustration	54
2.4.1	Evaluation de classifieurs sur des jeux de données	54
2.4.2	Illustration de sorties probabilistes, évidentielles et floues	55
2.5	Conclusion du chapitre	57
3	Classification imprécise	59
3.1	Principe	61
3.1.1	Définition et les attentes de la classification imprécise	61
3.1.2	Illustration de l'intérêt de la classification imprécise	65
3.1.3	Distinction avec la classification multi-labels	68
3.1.4	Classification imprécise versus la classification hiérarchique	69
3.1.5	Classification imprécise et option de rejet	70
3.2	Méthodes de classification imprécise	71
3.2.1	Classification imprécise basée sur la théorie des probabilités	71
3.2.1.1	Méthodes basée sur l'espérance de gain	71
3.2.1.2	Classification conforme	72
3.2.2	Classification imprécise basée sur les probabilités imprécises	73
3.2.3	Classification imprécise basée sur les fonctions de croyance	74
3.2.4	Classification imprécise basée sur la logique floue	75
3.2.5	Classificaion imprécise basée sur la théorie des possibilités	75
3.2.6	Autres méthodes de classification imprécise sans hypothèse de modèle	75
3.3	Évaluation d'un classifieur imprécis	75
3.3.1	Justesse affaiblie	77
3.3.2	Mesure d'évaluation F_β	78
3.3.3	Mesures d'évaluation g_{65} et g_{80}	80
3.3.4	Lien avec l'évaluation de classifieurs hiérarchiques	81
3.4	Positionnement	85
3.5	Conclusion du chapitre	87
4	Proposition d'un classifieur imprécis évidentiel	89
4.1	Proposition d'une approche de classification imprécise évidentielle	90
4.1.1	Construction générique d'une fonction de masse via un ré-étiquetage im- précis	91
4.1.1.1	Ré-étiquetage imprécis et extraction de caractéristiques	94
4.1.1.2	Ré-étiquetage imprécis et déséquilibre des classes	95
4.1.2	Choix d'une sortie imprécise à partir d'une fonction de masse et d'un critère de décision	97
4.2	Discussion	99
4.3	Illustration de l'approche proposée ECLAIR	100
4.4	Performances d'ECLAIR par rapport à l'état de l'art	106
4.5	Conclusion du chapitre	109

5 Manipulation de vecteurs de proportions imprécises et incertaines : Application au tri des plastiques sombres.	113
5.1 Éléments théoriques pour la gestion d'information évidentielle sur des vecteurs de proportions	115
5.1.1 Vecteurs de proportions imprécises et/ou incertaines	116
5.1.2 Vecteur de proportions d'un mélange	119
5.1.2.1 Cas déterministe	120
5.1.2.2 Cas probabiliste	120
5.1.2.3 Cas imprécis sous forme d'intervalles	121
5.1.2.4 Cas évidentiel	122
5.1.3 Inférence sur le respect des contraintes	124
5.2 Stratégies de tri	126
5.2.1 Mise à jour de la connaissance concernant la composition du bac	128
5.2.2 Inférence sur le respect des contraintes dans le bac positif	129
5.2.3 Complexité	131
5.2.3.1 Procédure naïve	131
5.2.3.2 Procédure prudente ou flexible	131
5.3 Illustration	132
5.3.1 Présentation des données disponibles pour le problème de tri	132
5.3.2 Simulations	133
5.4 Discussion	139
5.4.1 Généralisation au cas de fragments qui ne sont pas monomatières	139
5.4.2 Généralisation au cas de plusieurs bacs	140
5.5 Conclusion du chapitre	142
Conclusion générale	145
Publications durant la thèse	149
A Jeux de données	151
B Détails concernant les expérimentations	153
C Loi multinomiale et loi de Dirichlet	155
D Classifieur crédal naïf	157
E Détails des résultats du chapitre 4	161
Bibliographie	163

Table des figures

1.1	Évolution du nombre 1.1a et du tonnage 1.1b d'équipements mis sur le marché entre 2006 et 2018 (figures tirées de (ADEME, 2019))	6
1.2	Répartition des tonnages traités en 2018 par mode de traitement en France (figure tirée de (ADEME, 2019))	7
1.3	Composition des DEEE (figure tirée de (ADEME, 2019))	7
1.4	Structure hiérarchique des plastiques considérés	8
1.5	Domaines de longueurs d'onde électromagnétiques http://www.paultownsend.co.uk/research/fundamentals/terahertz-radiation/	10
1.6	Spectre laboratoire ATR d'un plastique ABS 1.6a, HIPS 1.6b, PE 1.6c, PP 1.6d.	10
1.7	Illustration d'un dispositif de tri positif ABS et PP	11
1.8	Spectre industriel de réflectance d'un plastique ABS 1.8a, HIPS 1.8b, PE 1.8c, PP 1.8d.	12
1.9	Image hyperspectrale d'un fragment de plastique	13
1.10	Photos de fragments de plastiques ABS 1.10a, HIPS 1.10b, PE 1.10c, PP 1.10d.	14
1.11	Spectres des quatre types de plastiques comparés à leur spectre de référence ATR correspondant	16
1.12	Illustration de la <i>procédure par vote majoritaire</i> pour le traitement d'un fragment <i>fr</i>	18
1.13	Illustration d'un dispositif de tri avec des options de tri prudentes	20
2.1	Exemple de fonction de masse, m et fonctions Bel et Pl associées.	33
2.2	Interprétation de l'ensemble flou "tiède"	39
2.3	Répartition des exemples d'un ensemble de validation lors du traitement du j -ième bloc.	42
2.4	Classification k -nn d'un exemple avec $k = 3$ et $k = 10$	44
2.5	Arbre CART entraîné sur le jeu de données <i>Iris</i>	44
2.6	Transformation de l'espace initial.	45
2.7	Représentation d'un neurone	45
2.8	Caption for LOF	47
2.9	Répartition des exemples d'apprentissage et trois nouveaux exemples \mathbf{x}^1 , \mathbf{x}^2 et \mathbf{x}^3 , dans l'espace des deux caractéristiques. (Figure inspirée de (Hüllermeier et Brinker, 2008))	49
3.1	Bases d'apprentissage pour les configurations facile (a) et difficile (b)	66

3.2	Classification standard pour les configurations facile (a) et difficile (b)	67
3.3	Classification imprécise pour les configurations facile (a) et difficile (b)	67
3.4	Illustration de partitions pour un problème de classification (figure inspirée de (Ha, 1997))	70
3.5	Récompenses fournies par F_β en fonction de β	79
3.6	Représentation des fonctions g_{50} , g_{65} et g_{80} inspirée des auteurs (Zaffalon et al., 2012).	81
3.7	Deux structures de hiérarchie	82
4.1	Manipulation de la base d'apprentissage pour un exemple fictif	92
4.2	Détection des exemples à ré-étiqueter	92
4.3	Pré-traitements de la base d'apprentissage dans le cas sans extraction (a) / avec extraction (b)	95
4.4	Création d'exemples synthétiques avec la méthode SMOTE	96
4.5	Étapes de la classification imprécise proposée par ECLAIR	99
4.6	Base d'apprentissage initiale	101
4.7	Sélection des exemples d'apprentissage à ré-étiqueter dans des ensembles de classes	101
4.8	Base d'apprentissage ré-étiquetée	102
4.9	Base d'apprentissage définitive (après ré-étiquetage et ré-équilibrage des classes)	103
4.10	Nouvel exemple x' à classifier dans la base de test et sa fonction de masse associée $m(. x')$	103
4.11	Sorties ECLAIR pour les exemples de test	105
4.12	Sorties NONDET pour les exemples de test	105
4.13	Récompenses acc contre $acc_{discounted}$ obtenues sur le jeu de données <i>wine</i> pour les approches NONDET (Δ), ECLAIR (+) avec les classifieurs Naives Bayes, k -nn et SVM, et l'approche NCC (\circ) - résultats moyens obtenus sur 10 expériences de validations croisées à 10 blocs	108
4.14	Récompenses u_{50} , u_{80} , et acc obtenues par les approches NONDET, ECLAIR (avec le classifieur Naive Bayes, $\beta = 1$) et NCC sur 8 jeux de données - résultats moyens obtenus sur 10 expériences de validations croisées à 10 blocs	110
5.1	Représentation du vecteur de proportions de la crème aux œufs $\tau \simeq (0.65, 0.13, 0.22)$ dans le simplexe unitaire	115
5.2	Représentation des vecteurs de proportions imprécises dans le simplexe unitaire.	119
5.3	Représentation du vecteur de proportions imprécises décrivant la composition du bac 5.3a, et de la nouvelle pièce 5.3b dans le simplexe.	122
5.4	Représentation du vecteur de proportions imprécises décrivant la composition du mélange (bac + nouvelle pièce) dans le simplexe.	122
5.5	Représentation des éléments focaux décrivant, le bac 5.5a, la nouvelle pièce 5.5b et leur mélange 5.5c dans le simplexe unitaire.	124
5.6	Représentation des éléments focaux du mélange et de l'ensemble A_α dans le simplexe.	126
5.7	Estimation de la composition finale du bac d'après notre hypothèse simple	130
5.8	Estimation de la composition finale du bac d'après notre hypothèse simple, au début du remplissage (k petit), 5.8a et sur la fin du remplissage (k grand), 5.8b avec $\nu = 0.05$	131
5.9	Procédure utilisée pour obtenir une fonction de masse concernant un fragment f_r à partir de ses 6 pixels x^1, \dots, x^6	133

5.10	Détails des informations sur les fonctions de masse et les poids du flux de fragments simulé (5.10a, 5.10c) et du bac positif (5.10b, 5.10d) pour une simulation de remplissage avec la procédure naïve.	134
5.11	Suivi au cours du remplissage du bac dédié aux plastiques styréniques avec la stratégie naïve , 5.11a de la croyance que la contrainte est respectée dans le bac, 5.11b du poids du contenu du bac W_k^{bac}	136
5.12	Suivi au cours du remplissage du bac dédié aux plastiques styréniques avec la stratégie prudente , 5.12a de la croyance que la contrainte est respectée dans le bac, 5.12b du poids du contenu du bac W_k^{bac}	137
5.13	Suivi au cours du remplissage du bac dédié aux plastiques styréniques avec la stratégie flexible , 5.13a de la croyance que la contrainte est respectée dans le bac, 5.13b du poids du contenu du bac W_k^{bac}	138

1.1 Répartition des catégories de plastiques dans les données disponibles	14
1.2 Exemples de valeurs monétaire et de capacité pour quelques de bacs de tri	21
2.1 Qualités de quelques exemples d'information pour répondre à la question : <i>Quelle est la nature exacte du fragment de plastique ?</i>	26
2.2 Matrice de confusion	41
2.3 <i>acc</i> moyens obtenus sur 10 répétitions de validations croisées à 10 blocs.	55
2.4 Temps moyens en secondes pour la classification d'un bloc de validation croisée (moyenne sur tous les jeux de données et les 10 répétitions des validations croisées)	55
2.5 Prédictions obtenues pour l'exemple de test \mathbf{x}' (zone de chevauchement des classes)	56
2.6 Prédictions obtenues pour l'exemple de test \mathbf{x}'' (zone avec peu d'exemples)	56
3.1 Récompenses de sorties imprécises	76
3.2 <i>acc_{discounted}</i>	77
3.3 F_β	79
3.4 Récompenses de la mesure d'évaluation obtenues avec g_{65}	81
3.5 Evaluations hP et hR obtenues considérant la hiérarchie de la Fig. 3.7a.	83
3.6 Exemples d'articles proposant des approches de classification imprécise (répartis en fonction du cadre théorique et du type de décision utilisés)	86
4.1 Répartitions des exemples d'apprentissage	95
4.2 Utilité F_β généralisée.	98
4.3 Ré-étiquetage de deux exemples d'apprentissage	102
4.4 Effectifs des exemples dans les nouvelles classes	102
4.5 Perte L_β	104
4.6 Résultats de classification imprécise pour ECLAIR et NONDET	104
5.1 Les deux informations évidentielles	123
5.2 Calcul de $Bel(A_\alpha)$ et $Pl(A_\alpha)$	126
5.3 Base de données	132
5.4 Valeurs moyennes de $Bel_{final}(A_\alpha)$ et <i>taux de tris positifs final</i> sur 30 simulations	135
5.5 Taux de justesse (taux de prédictions exactes) <i>acc</i> de classification de fragments	142

E.1	Récompenses <i>acc</i> NONDET et ECLAIR basés sur Naive Bayes avec $\beta = 1$ et NCC avec $s = 1$ - résultats moyens obtenus sur 10 expériences de validations croisées à 10 blocs	161
E.2	Récompenses <i>u₆₅</i> NONDET et ECLAIR basés sur Naive Bayes avec $\beta = 1$ et NCC avec $s = 1$ - résultats moyens obtenus sur 10 expériences de validations croisées à 10 blocs	161
E.3	Récompenses <i>u₈₀</i> NONDET et ECLAIR basés sur Naive Bbayes avec $\beta = 1$ et NCC avec $s = 1$ - résultats moyens obtenus sur 10 expériences de validations croisées à 10 blocs	162
E.4	Récompenses <i>acc_{discounted}</i> NONDET et ECLAIR basés sur Naive Bayes avec $\beta = 1$ et NCC avec $s = 1$ - résultats moyens obtenus sur 10 expériences de validations croisées à 10 blocs	162

Avec l'essor des capacités de calcul et de stockage basés sur les premiers travaux d'Alan Turing sur l'intelligence artificielle en 1950 (Turing et Haugeland, 1950), de plus en plus de problèmes industriels comme sociétaux bénéficient d'outils de traitement de données, que ce soit à des fins d'analyse descriptive ou de prédiction. L'apprentissage automatique est l'un des champs de l'intelligence artificielle permettant de résoudre des problèmes de prédiction en exploitant des données pour construire des modèles automatiquement. Parmi ces modèles, certains permettent d'affecter une classe à un vecteur d'attributs, ils sont appelés classifieurs supervisés lorsqu'ils sont construits à partir de données annotées par la classe que l'on cherche à caractériser dans le cadre d'une base d'apprentissage. Après la phase d'apprentissage, i.e, la construction à partir des données annotées, un classifieur supervisé est capable de prédire les annotations, i.e, les classes, de nouveaux échantillons. La présence d'incertitude (lorsqu'on n'est pas certain qu'une information soit vraie ou fausse) dans la majorité des applications réelles est généralement traitée en faisant appel à la théorie des probabilités. Cependant dans certaines applications, la prise en compte de la seule incertitude est insuffisante ou inadaptée pour appréhender toute l'imperfection des données. Par exemple, il peut s'avérer que le contenu informationnel des données soit insuffisant pour répondre à la question posée. A titre d'exemple, apercevoir une "baie" cachée au fond d'un buisson épais ne permet pas d'affirmer qu'il s'agisse assurément d'une "fraise des bois" ou d'une "framboise" mais nous pouvons être sûrs, à partir de sa couleur, qu'il s'agit d'une "baie rouge". Nous parlerons dans ce cas d'*informations imprécises*. Le besoin de gérer imprécision et incertitude concomitamment s'est traduit par le développement de théories modernes de l'incertain. Parmi elles, la théorie des fonctions de croyance est une généralisation de la théorie des probabilités permettant de représenter à la fois l'incertitude et l'imprécision. L'exploitation de la prise en compte de l'incertitude et de l'imprécision dans la représentation de l'information, la gestion de l'équilibre, du paradoxe entre ces deux dimensions de l'imperfection de l'information permet d'introduire la notion de prudence dans la prise de décision des classifieurs supervisés.

Cette thèse fait partie d'un projet de collaboration entre l'IMT Mines Ales, SUEZ et la société Pellenc ST, qui conçoit, produit et commercialise des machines de tri optique pour les déchets ménagers et industriels. La collaboration a pour objectif la conception d'une machine de tri en ligne pour tout type de plastiques y compris les plastiques sombres et noirs issus de la filière Déchets d'Equipements Electriques et Electroniques (DEEE). L'application de cette thèse concerne l'algorithme de tri qui doit être intégré à la machine pour effectuer la décision du tri en temps réel. Plus particulièrement, l'objet de cet algorithme est d'identifier la nature de fragments de plastiques issus de DEEE et arrivant en continu sur un tapis roulant parmi

quatre catégories de plastiques et de prendre une décision de tri pour chacun de ces fragments (choix du bac où sera envoyé chaque fragment). Pour cela, l'algorithme dispose pour chaque fragment de plastiques arrivant sur un tapis roulant, d'images hyperspectrales, i.e., des images pour lesquelles chaque pixel correspond à un spectre renseignant sur la nature du matériau à l'emplacement du pixel. En raison des conditions d'acquisition et de la technologie utilisée, ces images hyperspectrales sont sujettes à des imperfections qui entravent leur traitement. Cependant, les classifieurs supervisés classiquement utilisés dans l'industrie ne tiennent pas compte de ces imperfections dans leur traitement, ce qui nuit à la qualité du tri. De ce point de vue applicatif, notre ambition est de montrer qu'une gestion appropriée de l'imperfection des données combinée à des options de tri prudentes peut être avantageuse pour le tri en ligne appliqué ici à des fragments de plastiques issus de DEEE.

Nous avons choisi d'inscrire nos travaux de recherche dans la théorie des fonctions de croyance en raison des spécificités du problème applicatif. En effet, les données disponibles dans l'application, i.e., les images hyperspectrales des pixels de fragments, entachées d'imprécision et d'incertitude, peuvent être représentées et pleinement exploitées via les fonctions de croyance. Par ailleurs, le passage de l'étude des pixels à l'étude d'images hyperspectrales composées de plusieurs pixels, peut être effectué avec de nombreux outils de combinaison et de fusion proposés par les fonctions de croyance.

Nous nous sommes également intéressés à des algorithmes de classification particuliers exploitant les incertitudes et les imprécisions sur les informations disponibles, pour fournir des sorties exprimant le doute/l'hésitation. Nous nous référons à ce type de classification par le terme *classification prudente*. Notre intérêt pour la classification prudente est justifié par l'exigence du contrôle des erreurs de tri qui est une des attentes du problème de tri des plastiques. En effet, afin d'être recyclés les fragments de plastiques doivent être séparés par type de matières avec une grande fiabilité malgré la présence d'incertitudes et d'imprécisions. Nous souhaitons montrer dans cette thèse que des sorties exprimant le doute/l'hésitation sont profitables pour limiter les erreurs de tri ayant de graves conséquences. A titre d'exemple, s'il est impossible d'identifier précisément un fragment de plastiques au vu de l'information disponible, mais qu'en revanche il n'y a pas de doute quant à sa famille, il peut s'avérer judicieux de l'envoyer dans un bac moins sélectif dédié à une famille de plastiques et donc de valeur marchande moindre, en étant certain en revanche de la légitimité de l'action, plutôt que de prendre le risque de l'envoyer dans un bac sélectif qui garantit une rémunération avantageuse sur le marché du recyclage, mais qui n'exclut pas une erreur de tri et par suite l'adjonction d'impuretés dans un bac de tri.

Nous allons suivre l'organisation suivante dans ce manuscrit. Le chapitre 1 est dédié à la présentation du contexte et du problème industriel du tri de fragments de plastiques qui a motivé cette thèse. Nous montrons en quoi nos orientations de recherche sont intéressantes pour cette application.

Dans le chapitre 2, nous introduisons le cadre de notre état de l'art à l'intersection des domaines de la classification supervisée et des théories de l'incertain. Cela nous permet dans le chapitre 3, de nous concentrer sur une forme de classification prudente, la classification imprécise, dont les approches proposées dans la littérature s'appuient sur différentes théories de l'incertain.

Dans le chapitre 4, nous présentons notre proposition de classification imprécise basée sur la théorie des fonctions de croyance, que nous comparons avec deux classifieurs imprécis de l'état de l'art sur plusieurs jeux de données. Le chapitre 5 de ce manuscrit est dédié à l'intégration des spécificités du tri des plastiques, i.e., aspect séquentiel et contraintes sur les erreurs de tri, dans le problème de décision du tri des fragments. A cette fin, nous sommes amenés à manipuler des informations concernant les proportions des différents types de plastiques dans des bacs de

tri. Nous présentons des éléments théoriques permettant de manipuler ces informations dans le cadre des fonctions de croyance, puis nous proposons des stratégies de tri en exploitant les éléments théoriques.

Sommaire

1.1	Enjeux du tri/de l'identification de plastiques	6
1.2	Technologies optiques pour le tri/l'identification de plastiques	9
1.3	Acquisition des données et objectifs du tri	12
1.3.1	Contexte d'acquisition des données	12
1.3.2	Qualité des données	14
1.3.3	Objectifs du processus du tri	17
1.4	Orientation de la thèse	17
1.4.1	Attitude prudente dans le tri des plastiques	19
1.4.2	Problème d'optimisation du tri	20
1.5	Conclusion	21

Dans ce premier chapitre, nous présentons le contexte applicatif qui a motivé cette thèse dans le projet de collaboration entre l'IMT Mines Ales, SUEZ et la société Pellenc st. Il s'agit du problème du tri/de l'identification de plastiques présents dans les flux de type *déchets d'équipements électriques et électroniques* (DEEE). Nous faisons d'abord état des verrous technologiques qui empêchent de traiter les plastiques sombres et noirs, et en quoi la conception d'une nouvelle machine munie d'une technologie *moyen infrarouge* (MIR) permettrait de lever ces verrous. Bien que la technique permette de traiter les nouveaux plastiques visés, nous sommes confrontés à de nouvelles difficultés liées à des imperfections qui entachent les nouvelles données fournies par la technologie MIR. Nous présentons les défis rencontrés pour traiter ces données imparfaites et nous formalisons la problématique industrielle du tri de fragments de plastiques qui traduit des attentes en termes de rendement (quantité de plastiques triés disponibles pour le recyclage) et de qualité du tri (présence limitée d'impuretés parmi les fragments triés). Nous verrons également comment les problèmes de l'identification et du tri sont traités par les approches habituelles et nous identifions leurs limites. Ce constat nous permettra de donner des orientations de recherche pour améliorer la prise en compte de la qualité des données et la stratégie de tri de façon à améliorer la réponse à la problématique industrielle en terme de rendement et de qualité du tri. L'organisation de ce chapitre est la suivante. Dans la section 1.1, nous présentons les enjeux du tri/de l'identification de plastiques. Dans la section 1.2, nous introduisons certaines technologies candidates pour être intégrées à un dispositif de tri automatique des DEEE. Nous montrons ensuite, section 1.3, dans quel contexte nos données ont été recueillies et comment cela influence

leur qualité et, par conséquent, l'objectif de tri/ de l'identification des plastiques. Enfin, dans la section 1.4, nous présentons l'orientation générale de la thèse.

1.1 Enjeux du tri/de l'identification de plastiques

La production de masse d'appareils technologiques dont l'obsolescence est programmée par simplement la production de nouveaux modèles et qui n'ont pas été écoconçus¹, a entraîné l'émergence d'un des flux de déchets à la croissance la plus rapide au monde : les déchets d'équipements électriques et électroniques (DEEE). La catégorie des équipements électriques et électroniques² regroupe de nombreux objets comme les machines à laver, téléphones portables, télévisions, perceuses, lampes, outils d'analyses, etc. Les estimations du taux de croissance des DEEE vont de 3 à 5 % par an (Cucchiella et al., 2015). La Fig. 1.1 montre l'évolution de la mise sur le marché français de ces équipements. En 2018, la quantité d'équipements mis sur le marché représentait 1 928 995 tonnes, soit 192 fois le poids de la tour Eiffel. Nous remarquons que la hausse en terme de nombre d'équipements est plus marquée que la hausse du tonnage. Cela s'explique notamment par l'essor des petits appareils plus légers comme les téléphones, ordinateurs, etc..

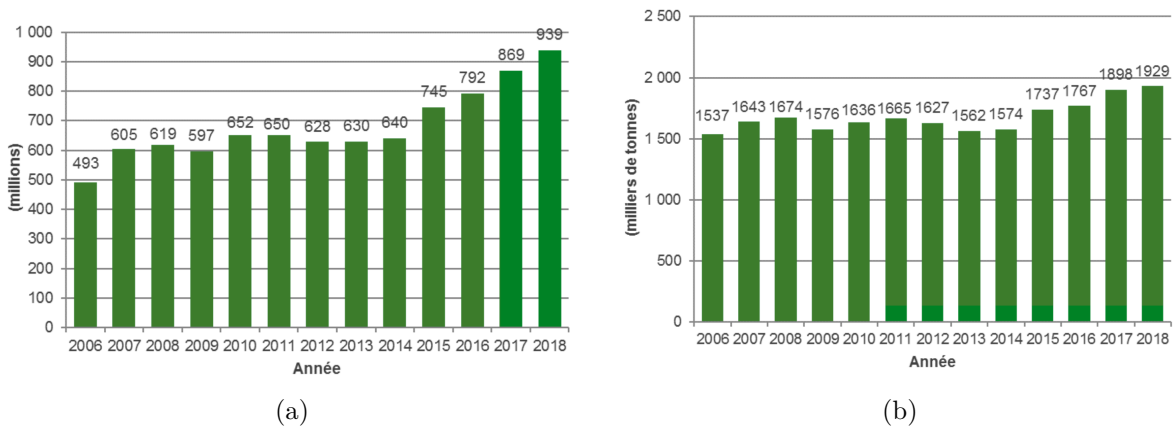


FIG. 1.1 Évolution du nombre 1.1a et du tonnage 1.1b d'équipements mis sur le marché entre 2006 et 2018 (figures tirées de (ADEME, 2019))

Arrivés en fin de vie, ces équipements peuvent être traités de différentes façons que nous présentons ci-dessous par ordre de priorité selon la réglementation (ADEME, 2019) :

1. Préparation à la réutilisation,
2. Réutilisation de pièces,
3. Recyclage matière : transformation en matières premières secondaires,
4. Valorisation énergétique : incinération des matériaux pour produire de l'énergie,
5. Élimination : enfouissement et incinération sans récupération énergétique.

La Fig. 1.2 montre la répartition des tonnages des DEEE en France en 2018 par les différents modes de traitement.

1. Conçus de façon à assurer un impact écologique réduit sur l'environnement pour l'ensemble du cycle de vie du produit et notamment la fin de vie du produit.

2. "équipements fonctionnant grâce à un courant électrique ou à un champ électromagnétique, ou un équipement de production, de transfert ou de mesure de ces courants et champs, conçu pour être utilisé à une tension ne dépassant pas 1 000 volts en courant alternatif et 1 500 volts en courant continu." (ADEME, 2019)

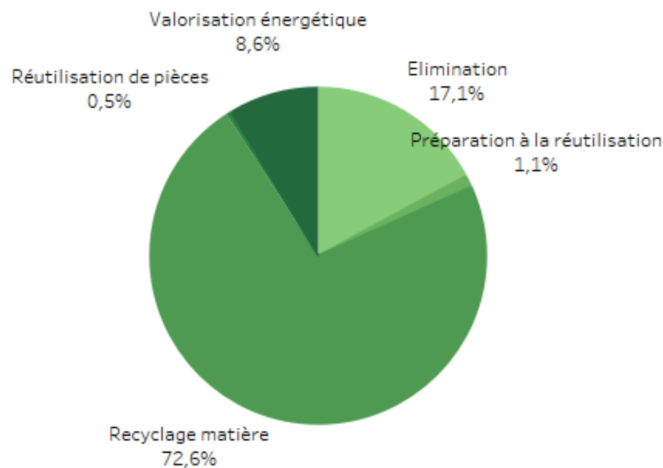


FIG. 1.2 Répartition des tonnages traités en 2018 par mode de traitement en France (figure tirée de (ADEME, 2019))

Nous constatons, que le mode de traitement par élimination est encore très présent (17,1%). Ce mode de traitement est complexe car les DEEE contiennent de nombreuses matières dangereuses qui présentent des risques pour la santé (Matarazzo et al., 2019). En raison de ces problèmes environnementaux et sanitaires croissants, les DEEE font l'objet d'une attention particulière de la part des autorités. La Commission européenne a renforcé les objectifs minimaux de valorisation des DEEE dans la directive 2012/19/UE (UE, 2012), selon laquelle, pour 10 catégories de DEEE, au moins 80% doivent être préparés pour la réutilisation et le recyclage. Afin de se conformer à cette réglementation, des avancées technologiques dans le processus de recyclage sont nécessaires.

Les DEEE sont principalement composés de métaux ferreux (51%) et de matières plastiques (18%) ou de verre (8%). La Fig. 1.3 montre les principaux matériaux qui constituent les DEEE.

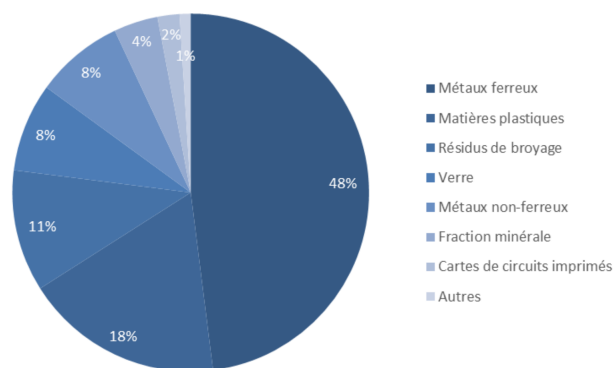


FIG. 1.3 Composition des DEEE (figure tirée de (ADEME, 2019))

Dans le processus du recyclage, une phase importante est la séparation de ces matériaux, qui a une influence directe sur tout traitement ultérieur et par conséquent sur la qualité finale des produits recyclés. La séparation doit être effectuée avec une grande fiabilité et une capacité massive pour faire face à l'énorme quantité de DEEE.

Dans cette thèse nous nous concentrons sur le flux des plastiques, qui représentent $\sim 21\%$ en poids des matériaux trouvés dans les DEEE (Ongondo et al., 2011), et 18% selon (ADEME, 2019). Le tri de ces plastiques devient de plus en plus complexe en raison de la grande variété des types de plastiques existants qui partagent des caractéristiques visuelles similaires. Le tri des plastiques est d'autant plus important que les différentes catégories de plastiques sont très peu miscibles lors de leur retransformation sous forme de mélanges et les impuretés provoquent des grandes difficultés pour le recyclage avec une baisse significative des propriétés mécaniques. De plus, des produits d'équipement électriques et électroniques sont constitués de plusieurs matières et souvent de plusieurs catégories de plastiques. Il est donc nécessaire d'identifier/de trier les plastiques afin de les séparer en monomatière. La séparation de ces plastiques est donc effectuée *a posteriori*, c'est-à-dire que tous les plastiques sont collectés ensemble, avant d'être déchiquetés, i.e., fragmentés en petits morceaux de plastiques, puis triés. Pour assurer une grande cadence et efficacité du processus de tri dans une telle configuration, c'est-à-dire des quantités considérables de plastiques de natures très différentes à trier dans un délai très court, de nombreux industriels envisagent des dispositifs de tri *optique en ligne* (section 1.2).

Notre étude se concentre plus particulièrement sur le tri de polymères styréniques et polyoléfiniques spécifiques qui sont très utilisés dans les processus industriels et qui représentent les matériaux majoritaires dans les DEEE (Signoret et al., 2019c). Les types de matériaux et de plastiques examinés dans cette thèse sont les quatre types de déchets de polymères d'origine pétrolière :

- *ABS* : *acrylonitrile-butadiène-styrène*, il s'agit d'un plastique de la famille des *styréniques*, historiquement utilisé par le fabricant de jouets Lego. Ce plastique est utilisé principalement dans des équipements électroménagers et dans l'ameublement.
- *HIPS* : *polystyrène à haute résistance aux chocs*, aussi appelé *polystyrène choc*. Il fait également partie de la famille des *styréniques*. Les principales applications de ce plastique sont les équipements électroménagers.
- *PE* : *polyéthylène*, est un plastique bon marché et très répandu de la famille des *polyoléfines*. Nous retrouvons ce plastique notamment dans les films alimentaires, les produits ménagers (emballages), les corps creux (bouteille de shampoing, cosmétiques, etc.), l'ameublement, le matériel de manutention, des pièces industrielles etc.
- *PP* : *polypropylène*. C'est un plastique de la famille des *polyoléfines*. Nous pouvons rencontrer ce plastique, entre autres, dans les emballages alimentaires et dans les équipements automobiles.

La Fig. 1.4 représente les précédents plastiques dans la structure hiérarchique de leurs familles de plastiques. Dans ce manuscrit, nous référons à l'ensemble des plastiques $\{ABS, HIPS\}$ par *styrénique* et à l'ensemble des plastiques $\{PE, PP\}$ par *polyoléfine*.

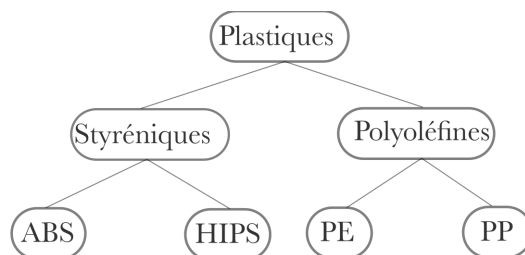


FIG. 1.4 Structure hiérarchique des plastiques considérés

Notons qu'outre les technologies optiques, il existe de nombreuses autres techniques de séparation. Parmi elles, la *flottaison* permet d'exploiter les différences de masse volumique entre les matériaux à trier, la *séparation par électromagnétisme* permet d'extraire les éléments ferromagnétiques³ (Signoret, 2019), ou encore, le *tri par la triboélectricité* qui consiste à tirer partie des différences de charges électriques accumulées sur la surface des matériaux (Signoret, 2019). Cependant, pour des raisons de cadence, d'efficacité, et de reconnaissance plus aisée, c'est une technologie à tri optique en ligne qui a été retenue dans le projet de machine de tri de l'entreprise Pellenc ST.

1.2 Technologies optiques pour le tri/l'identification de plastiques

Les technologies de tri/d'identification optiques déjà très répandues dans l'industrie alimentaire s'avèrent très appropriées pour le tri des déchets plastiques (Chen et Sun, 1991; Massen, 1998). Par rapport au tri manuel, elles offrent des avantages significatifs de rendement et de coût de main-d'œuvre. Pour le tri/l'identification des plastiques, les procédés infrarouges sont les technologies optiques les plus utilisées, car ils permettent de collecter rapidement des informations sur les vibrations inter-atomiques à la surface des matériaux (Signoret, 2019). Les informations spectrales (provenant des mesures d'absorption⁴ ou de transmission⁵) utilisées par les industriels pour distinguer les plastiques, sont principalement dérivées des gammes spectrales *proche infrarouge, near-infrared* (NIR) et des gammes spectrales *moyen infrarouge, mid-infrared* (MIR). La Fig. 1.5 montre comment sont situées les gammes NIR et MIR dans les domaines des longueurs d'ondes électromagnétiques. Notons qu'une longueur d'onde s'exprime en cm alors qu'un nombre d'onde s'exprime en cm^{-1} . Nous voyons que la gamme MIR est décrite pour des longueurs d'onde entre 2.5 et $25\mu m$. L'usage est d'utiliser les nombres d'onde exprimés en microns. Le MIR est donc décrit sur $4000 - 400cm^{-1}$. La technique de spectroscopie MIR "relate différents modes de vibrations des liaisons covalentes selon les atomes qu'elles relie, la nature de la liaison (simple, double ou triple) et la force de la liaison qui est influencée par son environnement chimique." (Signoret, 2019). Quant à la gamme NIR, exprimée sur la plage $12500 - 4000 cm^{-1}$, elle "relate les harmoniques de ce qui se passe en MIR" (Signoret, 2019). Dans la suite, nous distinguons deux gammes de longueurs d'ondes sur la gamme MIR (Signoret, 2019) : le MWIR (généralement $5000 - 2000cm^{-1}$) et le LWIR (généralement $1250 - 800cm^{-1}$).

La Fig. 1.6 présente les spectres d'absorbance sur la gamme MIR complète $4000 - 400cm^{-1}$ pour des matières plastiques *ABS, HIPS, PE, PP*, obtenus avec la technologie de laboratoire appelée la réflectance totale atténuée (ATR). Les courbes ATR ont été obtenues dans des conditions idéales (statique et en contact avec le fragment), de sorte que les pics sont très nets et qu'elle contient des informations parfaites pour l'identification de la nature du plastique en question. Pour les quatre spectres d'absorbance présentés, nous pouvons noter que les informations discriminantes entre les matières plastiques (pics différents e.g., les positions et intensité des pics) sont plus nombreuses sur la gamme LWIR que sur la gamme MWIR.

En utilisant la technologie NIR (longueurs d'onde : $700 - 1400nm$), certains industriels ont conçu des machines de tri capables de discriminer entre plusieurs types de plastiques. Cependant, l'utilisation de ces machines n'est pas adaptée aux plastiques sombres et noirs, qui sont très

3. Le ferromagnétisme est le mécanisme par lequel certains matériaux sont attirés par des aimants.

4. Une valeur d'absorption a est l'intensité du rayonnement électromagnétique absorbée par une substance à une longueur d'onde donnée.

5. Une valeur de transmittance t à une longueur d'onde donnée est associée à la valeur d'absorbance à cette longueur d'onde $t = \exp(-a)$

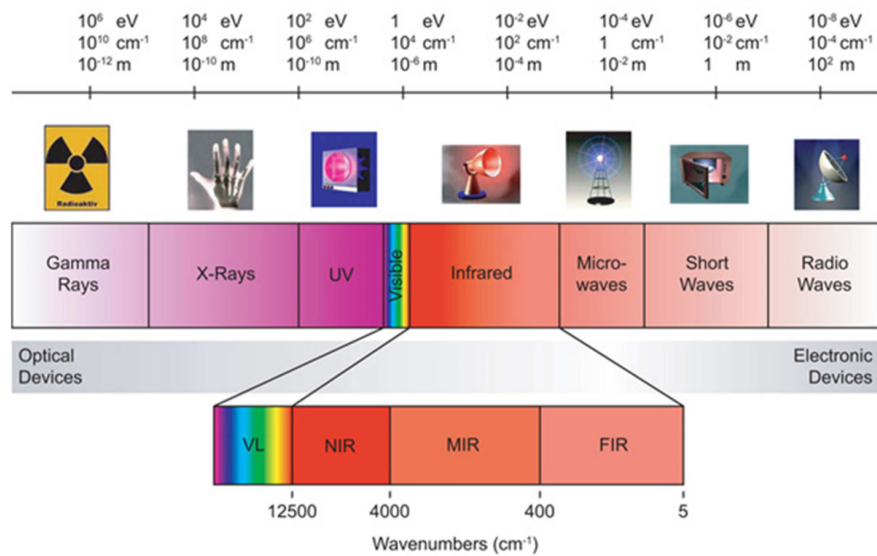


FIG. 1.5 Domaines de longueurs d'onde électromagnétiques <http://www.paultownsend.co.uk/research/fundamentals/terahertz-radiation/>

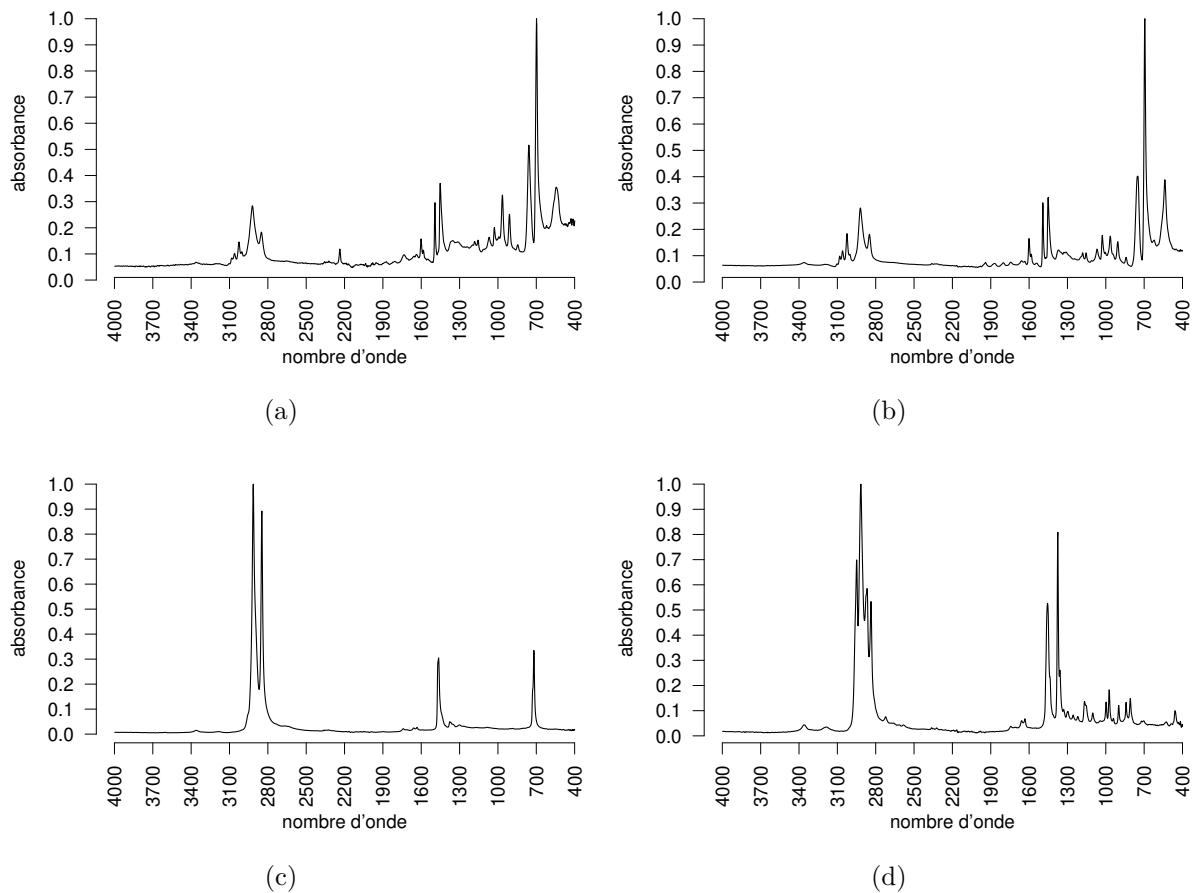


FIG. 1.6 Spectre laboratoire ATR d'un plastique ABS 1.6a, HIPS 1.6b, PE 1.6c, PP 1.6d.

courants dans les DEEE (près de 50% du gisement). La raison de cette limitation est que les plastiques sombres et noirs contiennent principalement du noir de carbone, qui absorbe une proportion si conséquente du signal NIR qu'il rend la signature du plastique méconnaissable.

Afin de surmonter cet obstacle, la technologie MIR (longueurs d'onde : 3000 – 8000nm) est une alternative prometteuse au NIR. L'impact du noir de carbone sur la signature spectrale est beaucoup moins problématique à ces longueurs d'onde plus élevées (Signoret *et al.*, 2020). La technologie MIR a déjà fait l'objet de plusieurs études. Par exemple, dans (Rozenstein *et al.*, 2017) les caractéristiques d'absorption du MIR pour plusieurs plastiques noirs ont été identifiées. Par ailleurs, les auteurs de (Kassouf *et al.*, 2014) ont présenté une étude de faisabilité pour la discrimination entre certains plastiques y compris des plastiques noirs, en utilisant l'ATR pour les caractériser. D'autres technologies telles que le TéraHertz (THz), la spectrométrie d'émission atomique de plasma induite par laser (LIBS) et la spectroscopie Raman sont des solutions alternatives potentielles pour le traitement des plastiques noirs (Barbier *et al.*, 2013). Cependant, la technologie THz ne semble pas assez mature pour être incorporée dans un dispositif industriel et la technologie LIBS est limitée par le nombre restreint d'éléments chimiques qu'elle peut mesurer (Barbier *et al.*, 2013). Concernant le Raman, une étude comparative de l'ATR et de la spectroscopie Raman a suggéré que la technologie Raman peut également produire des résultats très compétitifs pour la discrimination des déchets de plastiques noirs (Bae *et al.*, 2019). Cependant, il est difficile de fournir une acquisition en ligne compétitive en utilisant la technologie Raman. Enfin, la technologie ATR n'est pas encore adaptée pour une configuration de tri en ligne. En se basant sur ces comparaisons, il nous semble que le MIR possède les qualités pour être la technologie à appliquer industriellement pour le tri de plastiques. C'est cette technologie qui a été retenue pour trier les plastiques dans le contexte applicatif de cette thèse.

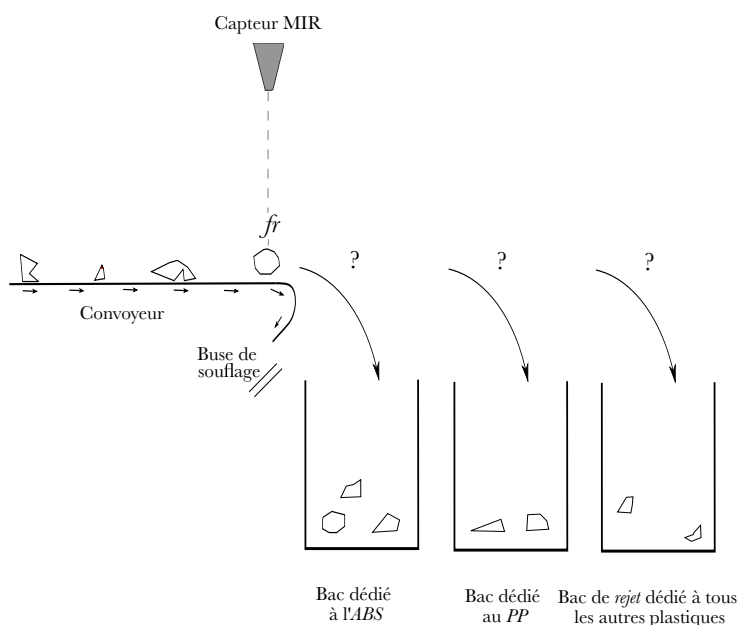


FIG. 1.7 Illustration d'un dispositif de tri positif ABS et PP

En pratique, le principe du tri en ligne des plastiques est le suivant. Des fragments de plastiques ou des objets sont acheminés en continu sur un convoyeur avant d'être enregistrés par une caméra MIR. Pour chaque fragment à trier, la caméra fournit un ensemble de spectres où chaque spectre correspond à un *pixel* dans l'image hyperspectrale du fragment identifié. Ces spectres sont ensuite traités par un algorithme intégré afin de déterminer le bac de tri le plus approprié où sera soufflé, i.e., envoyé par air comprimé, le fragment.

La figure 1.7, illustre un tel dispositif pour le tri de plastiques *ABS* et *PP* parmi les plastiques *ABS*, *HIPS*, *PE*, *PP*. Notons que le bac *rejet* est un bac dédié à tous les fragments qui n'ont pu être reconnus comme de l'*ABS* ou du *PP*, c'est à dire de l'*HIPS* ou du *PE* dans le dispositif présenté. Nous donnons dans la prochaine section des précisions sur la méthode d'acquisition utilisée par ce dispositif et le type des données qu'elle fournit et que nous utilisons dans ce manuscrit.

1.3 Acquisition des données et objectifs du tri

Dans cette section, nous présentons le contexte dans lequel nous avons acquis les données (la sous-section 1.3.1). Puis, nous exposons les objectifs que doit remplir une procédure de tri pour répondre aux exigences de valorisation dans la sous-section 1.3.3.

1.3.1 Contexte d'acquisition des données

Présentons les informations acquises industriellement concernant les fragments de plastiques considérés dans le problème applicatif de cette thèse. Il s'agit de spectres de réflectance enregistrés en mode dynamique par une caméra Specim MWIR sur la gamme $3687 - 896 \text{ cm}^{-1}$. La Fig. 1.8 présente des spectres de réflectance acquis avec la camera MWIR dans des conditions industrielles (fragments de plastiques transportés sur un tapis roulant à la vitesse de 2 m.s^{-1}). Un prétraitement a été appliqué à ces données, afin de réduire la diffusion de la lumière et les effets de bruit spectral, tous les spectres des pixels ont été corrigés par la technique de la variation normale standard (SNV)⁶.

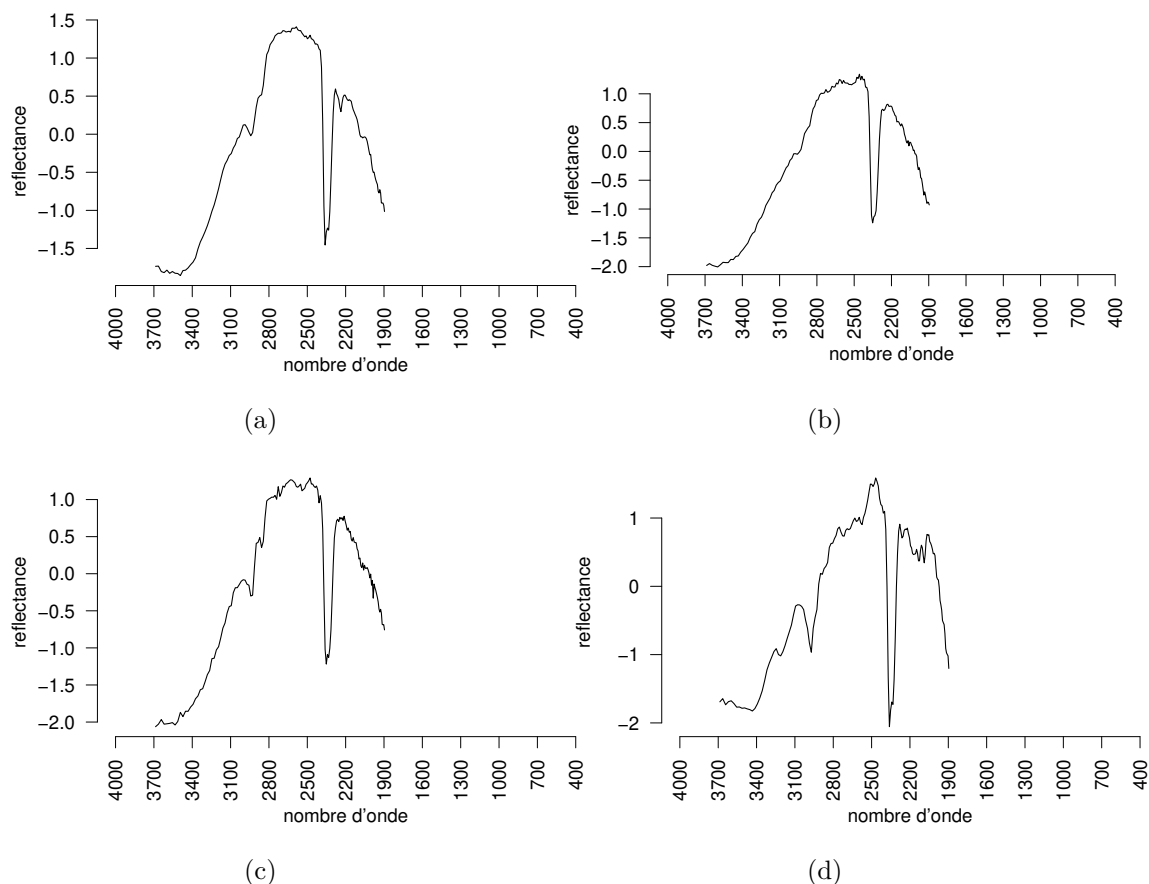


FIG. 1.8 Spectre industriel de réflectance d'un plastique *ABS* 1.8a, *HIPS* 1.8b, *PE* 1.8c, *PP* 1.8d.

6. Cette méthode consiste simplement à effectuer un recentrage et une réduction du spectre (variance unitaire sur l'ensemble du spectre). Chaque spectre peut donc être traité isolément. Il est à noter que les spectres corrigés par la technique SNV présentent des valeurs négatives dues à leur recentrage.

Les données dont nous disposons sont organisées sous la forme d'images hyperspectrales définies comme suit.

Définition 1. Image hyperspectrale L'image hyperspectrale d'un fragment de plastiques est un objet en trois dimensions. Les deux premières dimensions indiquent l'emplacement du pixel sur la largeur et la longueur du convoyeur. La troisième dimension de l'image correspond à la valeur de la longueur d'onde/du nombre d'onde, dans la pratique, la caméra fournit 154 valeurs par pixel réparties de manière équidistante sur $2712 - 5274 \text{ nm}$ (correspondant à des points non équidistants sur $3687 - 896 \text{ cm}^{-1}$). La figure 1.9 représente l'image hyperspectrale en trois dimensions acquise par la caméra pour un fragment donné.

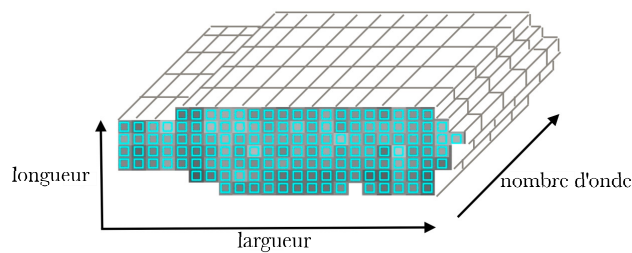


FIG. 1.9 Image hyperspectrale d'un fragment de plastique

Dans un souci de simplification, dans la suite du manuscrit, un vecteur spectral composé de ses 154 valeurs de réflectance est appelé *pixel*.

Notons que nous faisons l'hypothèse dans notre étude que les fragments à tirer sont monomatériaux, i.e., qu'ils ne sont composés que d'une matière plastique par fragments. Cette hypothèse a en réalité plus de chances d'être vérifiée lorsque les fragments sont petits. En effet, les DEEE sont souvent composés de plusieurs plastiques soudés ou collés entre eux. Le fait de fragmenter finement les DEEE est donc la seule solutions pour la séparation par matières.

Les données disponibles pour cette thèse concernent 280 fragments de plastiques provenant du groupe français Suez et comprenant des plastiques foncés, noirs et colorés. La Fig. 1.10 présente des fragments de plastiques utilisés pour constituer la base de données.



(a)



(b)



FIG. 1.10 Photos de fragments de plastiques ABS 1.10a, HIPS 1.10b, PE 1.10c, PP 1.10d.

La taille des fragments est environ $2 \times 2 \text{ cm}^2$. Chaque fragment de plastiques de notre base de données a été étiqueté sur la base d'identifications d'experts fiables à partir de spectres ATR. Après avoir été enregistrée par la caméra, chaque région d'intérêt des fragments de plastiques a été sélectionnée manuellement pour obtenir une image hyperspectrale, comme celle représentée dans la Fig. 1.9. La TABLE 1.1 présente la répartition des données fournies par l'entreprise Pellenc st. Précisons que parmi les données fournies par l'entreprise, nous avons constitué cette base de données bien équilibrée en terme de nombre de fragments et de pixels par classe en sélectionnant le même nombre de fragments en fonction de leur nombre de pixels.

	ABS	HIPS	PE	PP	total
nombres de fragments	70	70	70	70	280
nombres de pixels	5835	5859	5862	5833	23389

TABLE 1.1 – Répartition des catégories de plastiques dans les données disponibles

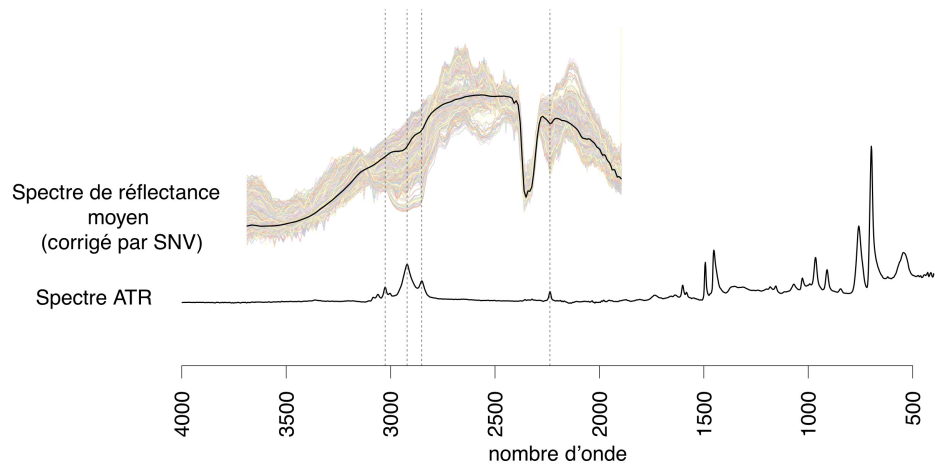
La figure 1.11 illustre certains spectres de la base données (correspondant aux pixels de plusieurs fragments) enregistrés par la caméra pour chacune des quatre catégories de plastiques (courbes colorées) et les spectres moyens (première courbe noire), comparés aux spectres de référence ATR correspondants (seconde courbe noire en bas de la figure) sur une échelle arbitraire pour l'axe des ordonnées. Notons que les pics des spectres de réflectance sont vers le bas alors que les pics des spectres ATR en absorbance sont vers le haut.

Sur chaque figure, les lignes verticales en pointillés montrent les pics les plus intenses en ATR (Signoret et al., 2019a,b).

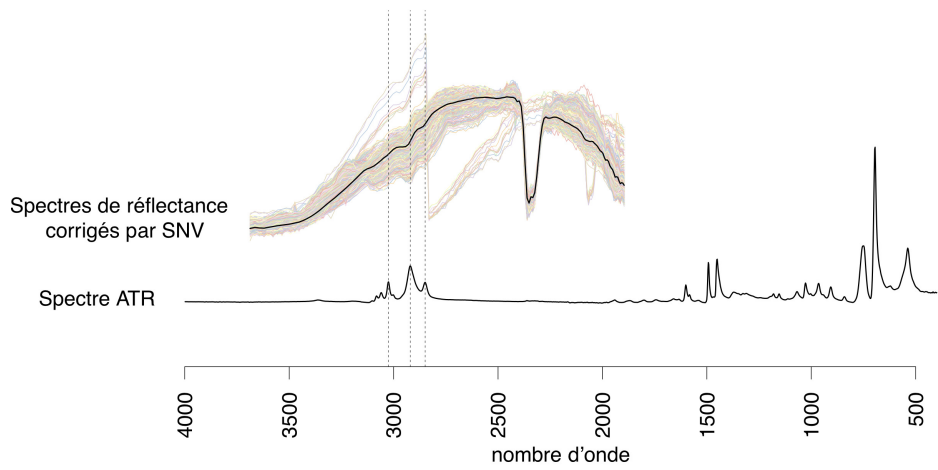
Lors de l'utilisation d'une technologie d'identification des plastiques dans un contexte industriel, la qualité des données acquises peut être affectée de manière significative de plusieurs façons, qui sont détaillées dans la suite.

1.3.2 Qualité des données

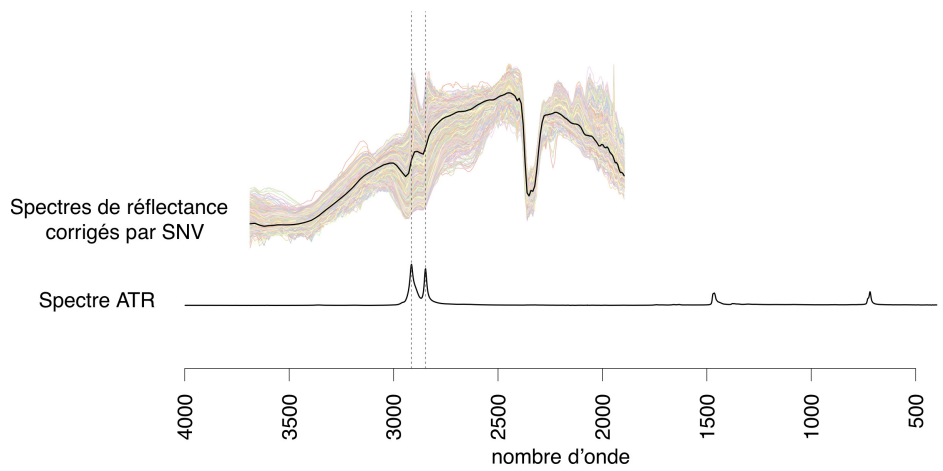
Comme le montre la figure 1.11 les conditions industrielles d'acquisition des spectres donnent des résultats très différents de ceux obtenus dans les conditions d'acquisition de laboratoire ATR.



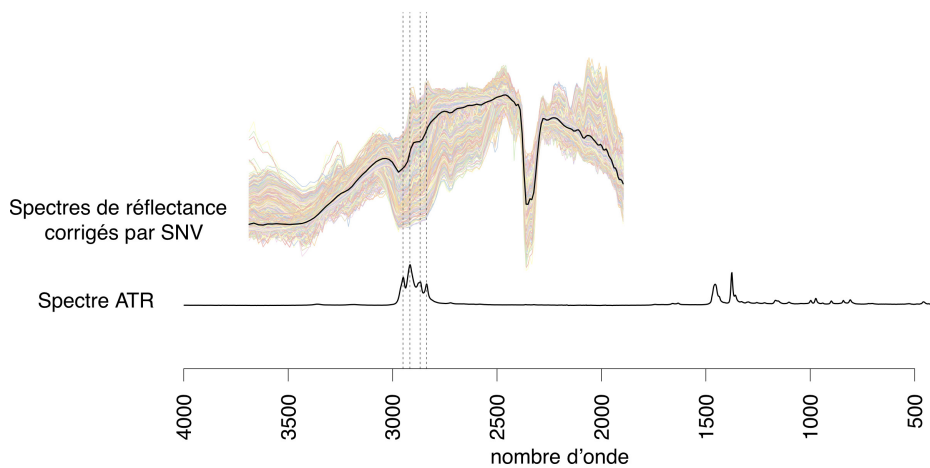
(a) Spectres ABS



(b) Spectres HIPS



(c) Spectres PE



(d) Spectres PP

FIG. 1.11 Spectres des quatre types de plastiques comparés à leur spectre de référence ATR correspondant

En effet, comme nous l'avons vu dans la section 1.2, les conditions ATR sont idéales (statique et en contact avec le fragment) alors que la caméra hyperspectrale fonctionne en dynamique, dans des conditions industrielle et a une gamme spectrale MIR moins complète. Nous donnons dans la suite les explication de la baisse de la qualité des données en conditions hyperspectral par rapport aux conditions idéales de laboratoire :

- avec la caméra hyperspectrale, la gamme spectrale disponible peut être insuffisante, elle n'est que de $1896 - 3687\text{cm}^{-1}$ (MWIR) pour les spectres industriels alors que la gamme de $400 - 4000\text{cm}^{-1}$ (LWIR et MWIR) est disponible pour les spectres ATR, or l'essentiel des pics discriminants pour les plastiques considérés sont dans la gamme LWIR ;
- les perturbations atmosphériques peuvent provoquer du bruit, ce qui explique une partie de la grande variance des spectres pour les mêmes matériaux dans la figure 1.11. Dans le cas de ATR ces perturbations sont bien moins importantes puisque l'acquisition se fait en contact direct avec le fragment.
- le vieillissement des plastiques et les additifs plastiques sont connus pour modifier les informations spectrales. Avec la technologie ATR, un laborantin est tout de même capable d'obtenir un spectre de bonne qualité en coupant le fragment de plastiques ou en le ponçant étant donné que le vieillissement des plastiques affecte seulement la surface des plastiques. Cette opération n'est pas possible dans les conditions industrielles, où la caméra hyperspectrale n'enregistre que la surface (partie la plus altérée par le vieillissement) des fragments.
- les impuretés comme les dépôts de poussière ou les restes d'étiquettes produisent également un bruit spectral, etc. En ATR, le laborantin peut prendre le temps de nettoyer les fragments, retirer les étiquettes, etc.
- les types de plastiques à recycler sont structurellement très similaires, et donc spectralement proches, par exemple ABS et HIPS ne peuvent physiquement être distingués que par le pic de l'acrylonitrile à 2236cm^{-1} . Ce pic clairement détectable en ATR l'est moins

avec la caméra hyperspectrale pour les raisons énoncées ci-dessus.

Ces conditions difficiles du contexte industriel conduisent à l’acquisition de données entachées d’imperfections. En effet, nous avons une *incertitude* sur la nature du matériau représentée par les spectres due à plusieurs raisons : nous ne sommes pas certains que le spectre est celui de la matière principale du fragment ou celui d’une impureté, nous ne savons pas à quel point le vieillissement a modifié le spectre, nous ne connaissons pas l’effet des perturbations atmosphériques, ect. Nous avons aussi de l’imprécision sur la nature du matériau. En effet le spectre d’un plastique peut fortement ressembler au spectre d’un autre plastique.

Notons que les auteurs ([Lachaize et al., 2018](#)) ont utilisé la théorie des fonctions de croyance pour effectuer une segmentation et une estimation de la classe de matériaux à partir d’images RVB, d’un profilomètre laser et d’une imagerie NIR. Par ailleurs, les auteurs ([Roh et al., 2017](#)) ont appliqué un classifieur par réseaux de neurones à fonction de base radiale floue pour identifier les plastiques noirs en se basant sur les spectres Raman.

1.3.3 Objectifs du processus du tri

Le tri des plastiques doit répondre à un double objectif traduisant des objectifs environnementaux et économiques : valorisation des déchets avec des produits de qualité mais aussi compétitifs en terme de coût. D’une part il est nécessaire de respecter certaines contraintes sur les proportions d’impuretés dans les bacs de tri. Si ces contraintes ne sont pas respectées, les contenus des bacs de tri ne pourront pas être exploités pour l’application visée. Par exemple, un travail précédent ([Perrin et al., 2016](#)), indique que sans agents de compatibilisation⁷, la tolérance maximale d’*ABS* dans l’*HIPS* ne peut être que de 4%. Du point de vue des fournisseurs de matières premières secondaires (résultats de la valorisation d’un bac trié), si les plastiques ne respectent pas certaines contraintes sur les taux d’impuretés, ces plastiques ne pourront pas être vendus pour cette application, ce qui représente une perte financière. Plus l’application visée sera exigeante, plus les contraintes sur les impuretés sont fortes. D’autre part, la procédure de tri devrait maximiser la quantité de plastiques soufflés dans les bacs de tri les plus valorisables. Il s’agit typiquement d’éviter de terminer le tri avec une trop grande quantité de plastiques dans le bac de rejets, i.e., le bac dédié aux plastiques qui n’ont pas été soufflés dans les bacs dédiés à des plastiques visés par le tri. Plus généralement, nous pouvons imaginer que certains bacs sont à prioriser dans certaines occasions. Nous pouvons résumer la problématique du problème du tri des plastiques de la façon suivante :

Sur la base d’informations imparfaites, effectuer un tri optimisé de fragments de plastiques respectant des contraintes de qualité sur les proportions d’impuretés dans les bacs de tri.

1.4 Orientation de la thèse

Dans cette section nous présentons les orientations de recherche que nous avons choisies de suivre pour répondre à la problématique du tri des plastiques. Pour déterminer ces orientations, nous nous sommes d’abord intéressés aux pratiques existantes pour effectuer le tri des plastiques. Dans la suite nous donnons quelques précisions sur une procédure habituelle de tri que l’on rencontre dans l’industrie, i.e., *le vote majoritaire* ([Paclik et Duin, 2005](#); [Paclik et al., 2006](#); [Lorente et al., 2012](#); [Blanch-Perez-del Notario et al., 2019](#)).

7. agents capables d’améliorer les affinités chimiques entre des plastiques qu’on ne pourrait pas recycler ensemble

Procédure de tri par vote majoritaire À notre connaissance, les algorithmes de tri basés sur l'imagerie hyperspectrale souvent utilisés dans l'industrie effectuent un vote majoritaire pour fusionner les informations provenant de chaque pixel d'un objet (dans notre cas fragment de plastique). Dans ce cas, une matière plastique est assignée à chaque pixel (phase de classification standard). Ensuite la matière assignée au fragment est celle assignée au plus grand nombre de ses pixels. Enfin, le fragment est soufflé dans le bac correspondant à la catégorie qui lui a été assignée (phase de tri). La Fig. 1.12 illustre cette procédure que nous appelons *procédure de tri par vote majoritaire* dans la suite du manuscrit.

Traitant des données imparfaites, les procédures habituelles comme celle *par vote majoritaire*, sont confrontées à deux problèmes différents qui peuvent affecter les performances de tri.

Le premier problème concerne l'absence de représentation adéquate de l'imperfection des données. Par conséquent, des données imparfaites pourraient être mal traitées et provoquer des erreurs de tri.

Exemple 1. *Considérons un fragment à trois pixels. Supposons que nous connaissons les probabilités que la nature de chaque pixel soit ABS, HIPS, PE ou PP. Par exemple supposons que les deux premiers pixels aient la même probabilité 0.55 d'être de l'ABS et la même probabilité 0.45 d'être de l'HIPS. Supposons aussi que le troisième pixel ait une probabilité d'être de l'HIPS égale à 1. Avec la procédure par vote majoritaire, nous assignons la catégorie de plastiques ABS au deux premiers pixels, et HIPS au troisième. Comme ABS reçoit plus de votes au niveau du fragment, la catégorie ABS est retenue comme identification du fragment. Pourtant, les probabilités sont en moyenne en faveur de la catégorie HIPS.*

Le deuxième problème concerne le type de sortie fournie par la procédure. Les fragments sont systématiquement assignés à une catégorie précise de plastiques, même lorsque les informations sont insuffisantes pour prendre une décision. Par conséquent, tous les fragments dont l'information sur leur matière est entachée d'imperfections, seront quand même soufflés dans des bacs de tri dédiés à des catégories de plastiques précises, ce qui entraînera inévitablement des erreurs de tri. Ce deuxième problème est étroitement lié au premier car la représentation des informations a un impact sur le résultat : si les imperfections ne sont pas représentées, elles ne peuvent pas être prises en compte pour obtenir une sortie plus riche concernant la nature du fragment. En particulier, la représentation probabiliste exploitée par les classificateurs standards, ne permet pas une représentation fidèle à l'imprécision des données.

Le premier défi abordé dans cette thèse est de traiter des données imparfaites pour fournir des informations plus prudentes sur la matière des fragments. Le second défi est d'intégrer les informations prudentes dans la procédure de tri des plastiques. Dans les sous-sections suivantes, nous donnons plus de détails sur l'intérêt de la prudence dans la procédure de tri (sous-section

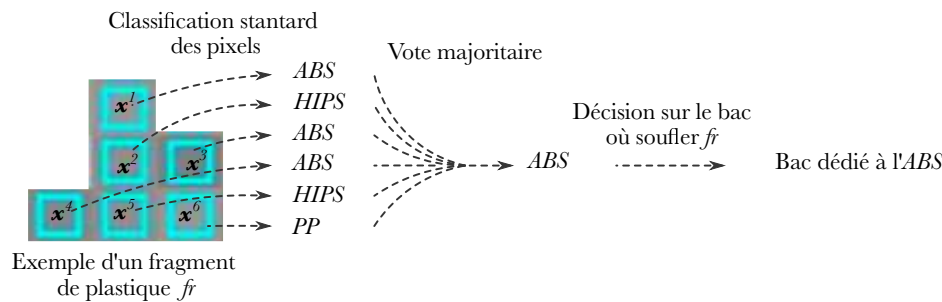


FIG. 1.12 Illustration de la procédure par vote majoritaire pour le traitement d'un fragment f_r

1.4.1). Ensuite, nous posons le cadre du problème d’optimisation pour répondre à la problématique du tri (sous-section 1.4.2).

1.4.1 Attitude prudente dans le tri des plastiques

Pour répondre à la problématique énoncée dans la sous-section 1.3.3, nous avons choisi d’inscrire cette thèse dans le cadre des théories modernes de l’incertain, et plus particulièrement, les travaux qui visent à mettre en place des algorithmes de classification *prudents* quand les données sont entachées d’incertitudes importantes. Nous entendons par *prudence*, dans ce manuscrit, l’attitude consistant à intégrer dans la représentation des données l’imperfection qui les entache et d’élargir l’ensemble des sorties d’un algorithme pour prendre le moins de risque possible. Dans les approches de tri les plus répandues (voir Fig. 1.12), nous sommes confrontés à une attitude risquée à plusieurs niveaux. D’abord lors de l’assignation d’une catégorie de plastiques précise pour chaque pixel. En effet, les pixels, certains plus que d’autres, sont entachés des imperfections énumérées dans la section 1.3.1. Ces imperfections rendent plus difficile le traitement de ces fragments. Sans la possibilité de s’abstenir, nous sommes obligés de parier sur une catégorie précise de plastiques. Ensuite lors de la fusion des informations provenant des pixels par le vote majoritaire, nous sommes obligés de retenir une unique catégorie de plastiques (même lorsque les résultats du vote sont serrés). Finalement lors de la décision du bac, en raison des prises de risques précédentes, nous agissons de façon risquée en soufflant le fragment de plastiques dans le bac dédié à une catégorie de plastiques précise. Lorsque l’information spectrale obtenue sur les pixels d’un fragment de plastiques est particulièrement entachée d’imperfections, nous risquons de commettre des erreurs de tri, et donc d’introduire des impuretés qui affecteront la qualité du bac. Certes dans certains cas, une procédure plus risquée aboutira au bon tri du fragment, mais lors du tri d’un flux de plastiques, les erreurs de tri sont inévitables. Notons que le paris sur une matière à associer à un pixel conduit à une perte d’information pour les éventuels traitements dans le processus du tri.

Dans la suite, nous présentons comment le dispositif de tri peut s’adapter à une attitude de tri prudente et nous précisons les intérêts de cette attitude dans le tri des plastiques.

Pour éviter de créer des impuretés trop importantes dans les bacs de tri, nous pouvons envisager des options de tri plus prudentes lorsque le fragment à trier est difficile. Une façon qui pourrait être compatible avec les dispositifs actuels de tri est d’ajouter des bacs de tri dédiés à plusieurs catégories de plastiques. La Fig. 1.13, illustre un dispositif de tri permettant un tri plus prudent. Par exemple, considérons un fragment de plastique, pour lequel, les pixels sont tels que la zone du pic *acrylonitrile* (qui permet la distinction claire entre l’*ABS* et l’*HIPS*), est fortement entachée d’imperfections. Pour ce fragment, le tri par l’approche du vote majoritaire conduit à souffler le fragment dans le bac dédié à l’*ABS* ou dans le bac dédié à l’*HIPS* sans réel fondement. Avec les nouvelles options de tri, nous avons l’occasion d’agir de façon plus prudente en soufflant le fragment dans le bac 3 dédiés aux plastiques *styreniques* (*ABS* ou *HIPS*).

Sur le plan technique, l’adaptation du dispositif de tri pour intégrer l’ensemble de toutes les options de tri possibles telles que représentées dans la Fig. 1.13 n’est pas toujours possible. Toutefois, il est possible de considérer un ensemble réduit de bacs. La solution la plus évidente est de considérer uniquement les bacs ayant un intérêt pour l’utilisateur de la machine (i.e., mélanges ayant encore une valeur marchande). L’intérêt d’un tel dispositif de tri, permettant une procédure de tri prudente pour les cas de fragments difficiles :

- Diminution des erreurs de tri dans les bacs dédiés à des plastiques précis : cela permettrait d’atteindre les normes exigées de qualité du tri pour les bacs les plus spécifiques, qui ont un impact conséquent sur la qualité des futurs matériaux recyclés.

- Exploitation des contenus des bacs moins spécifiques, dédiés à plusieurs catégories de plastiques : les contenus de ces bacs pourraient être utilisés pour un recyclage moins exigeant ou être intégrés dans une procédure de tri basée sur plusieurs passages.

1.4.2 Problème d'optimisation du tri

En autorisant les options de tri prudentes, nous pouvons éviter de dépasser les proportions d'impuretés dans les bacs de tri dédiés aux catégories précises de plastiques. L'inconvénient est qu'en soufflant des fragments de plastiques dans des bacs dédiés à plusieurs plastiques, nous diminuons la quantité finale de fragments finement triés (soufflés dans des bacs dédiés à des catégories de plastiques précises).

Exemple 2. Prenons l'exemple d'une procédure de tri excessivement prudente consistant à souffler systématiquement tout fragment dans le bac 15 dédié à tous les plastiques (voir Fig. 1.13). Cette procédure est excessivement prudente car elle garantit trivialement qu'aucune erreur de tri ne sera commise. En revanche, trier les plastiques avec cette procédure n'a aucun intérêt car tous les autres bacs sont vides.

La procédure de tri ne doit donc pas tomber dans la prudence excessive. Une attente mentionnée dans la problématique formulée dans la section 1.3.3, est que le tri doit être optimisé de façon à maximiser la quantité de plastiques triée dans les bacs visés pour le recyclage (ces bacs sont généralement ceux dédiés aux catégories précises de plastiques). Comme nous sommes en phase de construction de la machine, nous ne connaissons pas encore les préférences des décideurs, futurs acheteurs des machines. Pour cela, dans cette thèse, nous avons décidé de considérer des valeurs monétaires fictives associées au remplissage de chaque bac de tri. La TABLE 1.2 présente des exemples de valeurs monétaires M^b d'une unité de masse de plastiques trié dans un bac $b \in \{1, \dots, 15\}$ et la capacité W^b d'un bac de tri b . Ces données fictives doivent être déterminées en fonction des conditions de tri et des valeurs du marché lors de l'application réelle au tri de plastiques.

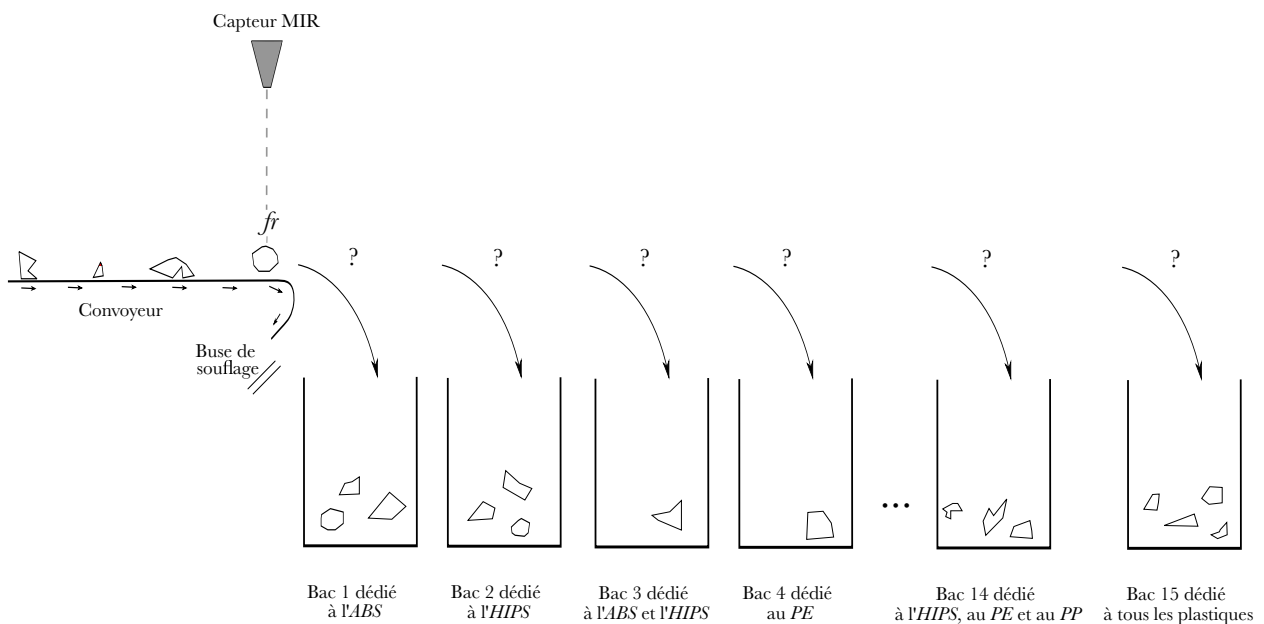


FIG. 1.13 Illustration d'un dispositif de tri avec des options de tri prudentes

Bac de tri b	1	2	3	15
dédié à l'ensemble de plastiques	$\{ABS\}$	$\{HIPS\}$	$\{ABS, HIPS\}$	$\{ABS, HIPS, PE, PP\}$
Valeur monétaire M^b	100€	90€	20€	1€
Capacité du bac W^b	1kg	1kg	1kg	1kg

TABLE 1.2 – Exemples de valeurs monétaire et de capacité pour quelques de bacs de tri

Nous avons intérêt à bien remplir les bacs de valeur monétaire élevée. Par exemple, avec les données de la TABLE 1.2, nous devons souffler le maximum de fragments de plastiques dans les bacs dédiés à l' ABS et à l' $HIPS$, les bacs spécifiques à plus grande valeur marchande. Cependant nous devons également veiller à ce que les contraintes sur les proportions d'impuretés soient respectées dans les bacs de tri. Nous utilisons des contraintes fictives pour illustrer nos travaux dans cette thèse. Nous notons $\tau^b = (\tau_{ABS}^b, \tau_{HIPS}^b, \tau_{PE}^b, \tau_{PP}^b)$ le vecteur des proportions de plastiques de catégorie ABS , $HIPS$, PE et PP à la fin du remplissage du bac b . Nous formalisons les contraintes caractérisant le bac de tri b comme

$$g^b(\tau^b) < 0, \quad (1)$$

où g^b est une fonction définie sur $[0, 1]^4$ à valeurs dans \mathbb{R} .

Exemple 3. Donnons un exemple de contrainte pour le bac 1 dédié à l' ABS . Supposons que les proportions de toutes les impuretés ne doivent pas dépasser 2% dans ce bac. Alors nous considérons la fonction g^1 définie comme $g^1(\tau_{ABS}^1, \tau_{HIPS}^1, \tau_{PE}^1, \tau_{PP}^1) = \tau_{HIPS}^1 + \tau_{PE}^1 + \tau_{PP}^1 - 0.02$.

Le problème applicatif adressé dans cette thèse est de proposer une procédure de tri assez prudente pour respecter les contraintes sur les proportions d'impuretés, tout en maximisant le remplissage des bacs avec la valeur marchande la plus élevée.

$$\sum_{b=1}^{15} W_{final}^b \cdot M^b, \quad (2)$$

où W_{final}^b est la quantité finale de plastiques dans le bac b .

Dans cette thèse, nous nous intéressons à la possibilité d'exploiter des informations prudentes recueillies sur les fragments précédemment triés pour prendre la décision du bac où souffler le fragment en cours de traitement pour maximiser la quantité (2) en respectant les contraintes (1).

1.5 Conclusion

Dans ce chapitre nous avons présenté le problème industriel du tri des plastiques issus de DEEE qui a motivé cette thèse. Le tri doit répondre à des problématiques environnementales et économiques : valorisation des déchets avec des produits de qualité et compétitivité en terme de coût. Les attentes du tri sont les suivantes : en terme d'optimisation, il faut maximiser le remplissage des bacs à plus haute valeur marchande. Concernant les matières triées, il faut respecter les contraintes sur les proportions d'impuretés (erreurs de tri). Pour répondre à ces attentes, nous disposons d'informations concernant les fragments de plastiques sous la forme de données affectées par des imperfections liées au contexte industriel. Nous avons exposé certaines limites que pourraient engendrer l'utilisation des approches classiques de tri de plastique qui ne sont pas capables de prendre en compte certaines imperfections dans leurs représentation de l'information. Ce constat nous a poussé à envisager un autre traitement de l'information capable

de fournir des informations plus prudentes concernant la nature des fragments. Pour fournir ces informations prudentes, nous devons adopter des représentations adéquates de l'information imparfaite.

Nous présentons dans le chapitre suivant des cadres adaptés à divers types d'informations imparfaites. Certains de ces cadres ont déjà été utilisés dans le contexte du tri des plastiques ([Roh et al., 2017](#); [Lachaize et al., 2018](#)).

 Classification à partir de données imprécises et/ou incertaines

Sommaire

2.1 Théories de l'incertain	26
2.1.1 Théorie des probabilités	27
2.1.2 Probabilités imprécises	30
2.1.3 Fonctions de croyance	31
2.1.4 Théorie des possibilités	36
2.1.5 Espérance mathématique et mesures floues	37
2.1.6 Logique floue	38
2.2 Classification automatique supervisée	40
2.2.1 Principe de la classification supervisée	40
2.2.2 Mesures d'évaluation	41
2.2.3 Algorithmes de classification supervisée usuels	43
2.2.4 Optimisation des classifieurs	46
2.3 Gestion d'incertitudes en classification automatique supervisée	48
2.3.1 Classifieur probabiliste	49
2.3.2 Classifieur évidentiel	50
2.3.3 Classifieur flou	53
2.4 Illustration	54
2.4.1 Evaluation de classifieurs sur des jeux de données	54
2.4.2 Illustration de sorties probabilistes, évidentielles et floues	55
2.5 Conclusion du chapitre	57

Dans le chapitre précédent, nous avons présenté le contexte général de cette thèse et la problématique industrielle considérée. Nous pouvons la résumer de la façon suivante : *Sur la base d'informations incertaines et/ou imprécises, effectuer un tri de fragments de plastiques respectant des contraintes de qualité sur les proportions d'impuretés admissibles dans les bacs de tri.* Cette problématique industrielle recouvre deux grandes problématiques de recherche. Il s'agit de la classification de données entachées d'imprécision et/ou d'incertitude et l'optimisation de mélanges sous incertitude. Pour résoudre ces problématiques générales, les travaux de l'état de l'art relèvent de plusieurs domaines : l'apprentissage automatique ; les théories de l'incertain ; la théorie de la décision ; la classification de données entachées d'imprécision et/ou d'incertitude ;

traitement et optimisation de mélanges sous incertitude. Ce chapitre se consacre à l'introduction des quatre premiers domaines.

L'*apprentissage automatique*, appelé *machine learning* en anglais, est un domaine de l'intelligence artificielle qui implique l'élaboration de modèles descriptifs ou prédictifs en exploitant des données disponibles appelées *données d'apprentissage*. Le terme "apprentissage" est employé pour insister sur le fait que le modèle est appris/entraîné, c'est-à-dire construit à partir de ces données d'apprentissage. Le modèle appris peut ensuite être appliqué pour prédire ou décrire de nouvelles données. Le qualificatif "automatique" nous permet d'insister sur l'aspect algorithmique de l'apprentissage automatique. En effet, l'apprentissage du modèle est automatisé (il suit des instructions pré-établies par des fonctions appelées *classifieurs*). L'apprentissage automatique a donc bénéficié des capacités de calcul et de stockage avec l'émergence de l'informatique. Depuis, l'apprentissage automatique est utilisé dans de très nombreux domaines parmi lesquels nous pouvons citer l'aide au diagnostic médical, la détection de fraudes, le traitement automatique de la langue ou encore l'interprétation d'images satellites. Nous distinguons deux principaux types de problèmes de classification :

1. La *classification non-supervisée* appelée aussi clustering qui vise à partitionner un ensemble d'objets en regroupant les plus similaires entre eux. Les classes induites par le partitionnement n'ont donc pas de signification a priori. De nombreux algorithmes ont été proposés pour résoudre cette tâche comme le *partitionnement en k-moyennes* (MacQueen et al., 1967) ou des algorithmes de *regroupement hiérarchique* (Rokach et Maimon, 2005).
2. La *classification supervisée* qui consiste à déterminer la classe d'un objet parmi un ensemble de classes connues a priori à l'aide de plusieurs exemples d'objets dont on connaît les classes.

Dans ce manuscrit, nous nous situons dans le cas de la classification supervisée étant donné que nous nous intéressons à la reconnaissance de nature de fragments de plastiques parmi un ensemble restreint de matériaux, i.e, (*ABS, HIPS, PE, et PP*), à l'aide de plusieurs exemples de ces matériaux. Notons que nous choisissons d'inscrire notre travail dans le cadre général de l'apprentissage automatique et que nous n'avons pas envisagé d'approches basées uniquement sur les connaissances des experts chimistes en reconnaissance des matériaux plastiques. En effet, la construction d'un modèle expert dans l'application de cette thèse aurait été fastidieuse puisque les règles des experts pour reconnaître des plastiques valent seulement pour des spectres de laboratoire bien contrôlés comme l'ATR. Ces règles ne sont plus valables ou sont insuffisantes lorsque nous disposons de données industrielles en raison des nombreuses imperfections de l'acquisition industrielle évoquées dans le chapitre précédent. Le cadre de l'apprentissage automatique nous permet de modéliser la reconnaissance des plastiques en exploitant une base données sans faire appel à ces connaissances expertes.

La présence des imperfections dans nos données, nous a conduit à utiliser les théories de l'incertain pour les représenter. Les différentes théories de l'incertain sont des formalismes pour manipuler des données incertaines et/ou imprécises. Historiquement, la théorie des probabilités a longtemps été le seul recours pour manipuler des données incertaines. Récemment, plusieurs autres cadres ont été proposés pour répondre à certaines limites de la théorie des probabilités. Plus précisément, elles viennent assouplir la condition d'additivité des probabilités qui ne correspond pas à la modélisation de certains problèmes réels. Le paradoxe d'Allais (Allais, 1953) montre que les axiomes de la théorie des probabilités peuvent être violés par le comportement de décideurs en situation réelle. D'autre part, le paradoxe d'Ellsberg (Ellsberg, 1961) montre que le comportement de décideurs est incompatible avec l'existence d'une unique probabilité subjective (Dubois et Prade, 2006). Dans ce chapitre nous proposons d'introduire les quatre

théories de l'incertain suivantes :

- Les *probabilités imprécises* (Walley, 1991) sont une généralisation de la théorie des probabilités permettant de manipuler des informations incertaines et imprécise. Cette théorie a été appliquée dans les domaines de l'intelligence artificielle (Antonucci et al., 2009), de la classification (Zaffalon, 1999), ou les études de risque (Destercke, 2008). Cette théorie offre de nouveaux outils et de nouvelles approches alternatifs à l'analyse Bayésienne robuste classique (Berger, 1990; Ramoni et Sebastiani, 2001).
- La *théorie des fonctions de croyance*, également appelée théorie de Dempster Shafer (Shafer, 1976; Smets et Kennes, 1994), est une extension de la théorie des probabilités qui permet d'allouer des masses à des sous-ensembles de l'univers des possibles ce qui permet, en plus de la représentation de l'incertitude, une représentation de l'imprécision. Les applications des fonctions de croyances sont nombreuses par exemple, la fusion d'information, la représentation de connaissances, l'apprentissage automatique ou l'analyse de risque.
- La *théorie des possibilités* (Dubois et Prade, 2012) représente l'incertitude et l'imprécision par le biais de deux mesures floues duales quantifiant le degré de possibilité et de certitude d'un événement. La *théorie des possibilités* peut représenter un cas particulier des fonctions de croyances.
- La *théorie de la logique floue* (Zadeh, 1965) généralise les ensembles classiques avec la notion d'ensemble flou. Un ensemble flou est un ensemble avec des frontières vagues comme on en trouve dans le langage naturel, par exemple "chaud", "tiède" ou "froid". Ainsi la manipulation des informations exprimées dans le langage naturel est une des applications de la logique floue. Plus généralement, des applications de la logique floue ont été proposées dans de nombreux domaines (Zimmermann, 2011; Kuncheva, 2000) tels que le contrôle de systèmes techniques, la théorie de la décision, les systèmes experts, la recherche opérationnelle ou encore la robotique. Les principaux outils de la logique floue à la base du raisonnement approximatif (très utilisés en intelligence artificielle) sont les règles floues (Mamdani, 1977).

L'objectif principal de ce chapitre est de dessiner un cadre pour notre état de l'art à l'intersection des domaines de la classification supervisée et des théories de l'incertain puisque notre tâche consiste en partie à la classification de données incertaines. Nous introduisons donc des méthodes de classification supervisée basées sur des théories de l'incertain. Dans l'application au tri des plastiques, notre intention est d'obtenir des informations, concernant la nature des fragments, riches en terme de représentation de l'incertitude et/ou de l'imprécision afin de maîtriser au mieux le processus du tri.

Dans ce chapitre, nous introduisons les concepts et les notations nécessaires à la présentation des solutions de l'état de l'art et de l'approche proposée dans cette thèse. L'organisation du chapitre est la suivante. Dans la première section 2.1 nous introduisons cinq théories qui proposent des cadres pour manipuler des informations incertaines et/ou imprécises. Dans la section 2.2, nous présentons des éléments de la classification supervisée, que nous rencontrons dans le manuscrit. Ensuite nous voyons comment la classification supervisée peut s'adapter à des contextes incertains en s'appuyant sur les théories de l'incertain. Nous terminons ce chapitre en comparant les performances et en illustrant les résultats de plusieurs méthodes de classification faisant appel à des représentations distinctes de l'incertitude.

2.1 Théories de l'incertain

Dans cette section nous introduisons les concepts élémentaires des cinq théories qui sont utilisées dans les approches de classification ou de tri discutées dans cette thèse. Nous voyons quels sont les types d'incertitudes qui sont considérés dans ces cadres (voir la référence suivante pour plus de détails sur ce point (Dubois et Prade, 2006)).

Commençons par préciser ce que nous appelons *information* au cours de cette thèse. D'après (Dubois et Prade, 2006), une information désigne "toute collection de symboles ou de signes produits soit par l'observation de phénomènes naturels ou artificiels, soit par l'activité cognitive humaine et destinée à comprendre le monde qui nous entoure, à aider à la prise de décision ou à communiquer avec des individus". Par exemple dans le problème industriel considéré dans cette thèse l'information disponible pour nous renseigner sur la nature d'un fragment est une image hyperspectrale (voir chapitre 1 pour la description de cette information).

Une information n'est pas toujours parfaite. Reprenant l'exemple du contexte applicatif de cette thèse, en raison des conditions d'acquisition de l'image hyperspectrale, les informations disponibles sont entachées d'imperfections. Il existe plusieurs types d'imperfections qui peuvent affecter la qualité d'une information et de nombreux termes sont utilisés dans la littérature pour les différencier. Dans certains travaux, *l'incertitude épistémique* (résultant d'un manque de connaissance ou de données) qui pourrait être levée en faisant des recherches plus poussées est distinguée de *l'incertitude aléatoire* (intrinsèque à l'aléa d'un phénomène) qui elle ne pourrait pas être levée (Der Kiureghian et Ditlevsen, 2009). Dans d'autres travaux, le terme *ambiguïté* est utilisé pour désigner le manque d'information (Camerer et Weber, 1992). Dans cette thèse nous adoptons les définitions suivantes pour les deux types d'imperfections de l'information *incertitude / imprécision* (Dubois et Prade, 2006) :

- **L'incertitude** : l'information est incertaine si l'on ne sait pas si cette information est vraie ou fausse.
- **L'imprécision** : l'information est imprécise lorsqu'elle ne permet pas de répondre complètement à la question que l'on se pose.

Exemple 1. *Illustrons les notions de l'incertitude et de l'imprécision avec l'exemple suivant. Considérons la question : Quelle est la nature exacte du fragment de plastique ? Si l'information disponible concernant la nature du fragment peut être vraie ou fausse, à cause par exemple de la confiance que nous accordons au capteur, alors nous sommes en présence d'incertitude. Si l'information disponible est insuffisante pour répondre à cette question, par exemple Le fragment est un styrénique¹ alors nous sommes en présence d'imprécision car le fragment pourrait bien être de l'ABS mais pourrait également être de l'HIPS. La Table 2.1 donne des exemples d'informations et leurs qualités pour répondre à la question posée.*

Information	Qualité de l'information
Le fragment est un ABS	certaine et précise
Le fragment est un styrénique	certaine et imprécise
Le fragment a des chances d'être un ABS	incertaine et précise
Le fragment a des chances d'être un styrénique	incertaine et imprécise

TABLE 2.1 – Qualités de quelques exemples d'information pour répondre à la question : *Quelle est la nature exacte du fragment de plastique ?*

1. Styrénique est une famille de plastiques comprenant l'ABS

Nous rencontrons dans notre état de l'art un autre type d'information appelée *information linguistique graduelle* (Dubois et Prade, 2006). Dans ce cas il ne s'agit plus d'informations qui peuvent être vraies ou fausses, mais qui peuvent être ni complètement vraies ni complètement fausses. Par exemple l'information "il fait froid" est perçue plus intensément lorsqu'il fait $5^{\circ}C$ que lorsqu'il fait $10^{\circ}C$.

Nous considérons, au fil de cette section, la valeur ν d'une variable θ . La variable θ peut être la variable de sortie dans un problème de régression, la variable discrète représentant les classes candidates dans un problème de classification, le vecteur de paramètres d'un modèle dans un problème d'inférence statistique, etc. Nous considérons ici que θ prend ses valeurs sur un univers Θ . Lorsque nous traitons du cas où Θ est un univers fini nous adoptons la notation $\Theta = \{\theta_1, \dots, \theta_n\}$, $n > 1$. Nous regardons comment chaque théorie de l'incertain propose de représenter l'incertitude et/ou l'imprécision concernant la valeur ν prise par θ . Nous présentons également les notions élémentaires qui sont utilisées dans des tâches d'estimation ou/et de décision à partir des informations recueillies sur ν .

2.1.1 Théorie des probabilités

Dans cette sous-section nous présentons les concepts de la théorie de probabilités que nous rencontrons dans cette thèse. Notamment, nous donnons les outils pour la représentation de l'incertitude (mesure de probabilité, variable aléatoire), pour l'estimation (maximum de vraisemblance, inférence bayésienne) et pour la décision (espérance mathématique). L'une des façons les plus répandues pour prendre une décision est en effet de minimiser l'espérance d'une fonction de perte (ou de maximiser l'espérance d'une fonction de gain). Cette méthode de décision est présentée plus loin dans ce chapitre, section 2.3.1.

Définition 1. Mesure de probabilité Formellement une *mesure de probabilité* notée P , est une fonction d'ensemble définie sur un espace mesurable (Θ, \mathcal{T}) où \mathcal{T} est une tribu² sur Θ et à valeurs dans $[0, 1]$ telle que :

$$\begin{aligned} P(\Theta) &= 1, \\ P(\emptyset) &= 0, \end{aligned}$$

et telle que $\forall A, B \subseteq \Theta$ tels que $A \cap B = \emptyset$, nous avons

$$P(A \cup B) = P(A) + P(B). \quad (3)$$

La dernière propriété appelée *additivité* est cruciale car elle permet de caractériser complètement P à partir des mesures des événements élémentaires de Θ qui sont représentés par la densité de probabilité notée f dans le cas continu ou par la distribution de probabilité discrète notée p dans le cas discret.

Définition 2. Variable aléatoire continue Le comportement de θ est décrit par sa densité de probabilité associée f définie sur Θ et à valeurs réelles positives comme suit :

$$P(\theta \in A) = \int_A f(u) du.$$

2. \mathcal{T} est un sous-ensemble non vide de 2^{Θ} , stable par passage au complémentaire et par l'union dénombrable. Les éléments de \mathcal{T} sont appelés événements.

Définition 3. Variable aléatoire discrète Lorsque θ est une variable aléatoire à valeurs discrètes $\{\theta_1, \dots, \theta_n\}$, son comportement est caractérisé par sa distribution de probabilités discrète p définie sur Θ et à valeurs dans $[0, 1]$ comme suit :

$$P(\theta \in A) = \sum_{\theta_i \in A} p(\theta_i).$$

Exemple 2. Afin d'illustrer la notion de variable aléatoire discrète, prenons l'exemple du lancement d'un dé à six faces. Considérons la variable aléatoire θ correspondant à la face obtenue lors du lancement du dé. Cette variable aléatoire prend ses valeurs dans l'univers fini $\Theta = \{\text{face } 1, \dots, \text{face } 6\}$. Supposons que nous ignorons si le dé est truqué ou non et que nous cherchons à caractériser le comportement de θ . Nous pouvons alors nous appuyer sur un échantillon de réalisations de θ (faces du dé obtenues lors de plusieurs expériences), pour construire une probabilité empirique. Notons que la valeur ν prise par θ est incertaine du fait de la variabilité des réalisations, i.e., plusieurs réalisations distinctes possibles pour la valeur que l'on cherche (plusieurs vérités). Par exemple si sur 100 expériences de lancement du dé nous avons obtenu 16 fois la face 1, nous posons $p(\text{face } 1) = 0.16$.

Il existe de nombreuses lois de probabilités usuelles pour caractériser le comportement d'une variable aléatoire. Les plus connues sont, dans le cas discret, la loi de Bernoulli (décrivant le résultat d'une expérience aléatoire à deux issues, succès ou échec), la loi Binomiale (pour décrire le nombre de succès lors de plusieurs expériences de Bernoulli), et dans le cas continu, nous pouvons citer la loi uniforme (caractérisée par une densité constante sur l'intervalle des valeurs possibles) et la loi normale aussi appelée loi gaussienne (qui d'après le théorème central limite, est une approximation de la somme d'une infinité de variables aléatoires indépendantes et identiquement distribuées suivant la même loi). Deux autres lois, la loi multinomiale et la loi de Dirichet sont présentées en annexe C.

Une fois que nous disposons d'une représentation de l'incertitude sous la forme d'une distribution de probabilité, nous basons généralement la décision sur la valeur espérée d'une variable aléatoire. Cette valeur espérée correspond intuitivement à la valeur que l'on s'attend à obtenir en moyenne pour un grand nombre d'expériences. Dans ce manuscrit nous rencontrons seulement le cas de calculs de valeurs espérées en fonction de variables aléatoires discrètes.

Définition 4. Espérance mathématique L'espérance mathématique de la variable aléatoire discrète θ prenant ses valeurs sur l'ensemble fini Θ avec une distribution de probabilités p est définie par :

$$\mathbb{E}[\theta] = \sum_i^n \theta_i p(\theta_i). \quad (4)$$

Dans certaines situations nous nous intéressons à l'espérance d'une valeur exprimée en fonction de θ . Par exemple dans la théorie de la décision, nous pouvons nous intéresser à l'espérance du gain associé à une action définie pour chaque valeur possible de θ (voir la sous-section 2.3.1 pour plus de précisions). Dans ce type de situation, nous pouvons calculer l'espérance mathématique de la variable aléatoire $\phi(\theta)$ sous certaines conditions sur la fonction ϕ (Théorème de transfert). Ainsi, la valeur espérée de $\phi(\theta)$ est définie par :

$$\mathbb{E}[\phi(\theta)] = \sum_{i=1}^n \phi(\theta_i) p(\theta_i).$$

Les définitions suivantes sont nécessaires pour appréhender les étapes d'estimation que nous rencontrons dans notre état de l'art.

Définition 5. Probabilité conditionnelle La probabilité conditionnelle d'un évènement $A \in \mathcal{T}$ sachant qu'un autre évènement $B \in \mathcal{T}$ a eu lieu est notée $P(A|B)$ et se définit comme suit :

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Le *théorème de Bayes* découle de cette définition et est donné par :

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

Définition 6. Estimation par maximum de vraisemblance La méthode du maximum de vraisemblance est utilisée pour estimer les paramètres d'une loi de probabilité d'une variable étant donné un échantillon de réalisations de cette variable. Considérons une variable aléatoire continue X et $l > 0$ réalisations de X : $\mathbf{x} = x^1, \dots, x^l$. Supposons que X suit une loi de probabilité paramétrée par un paramètre λ , i.e., f_λ . La fonction de vraisemblance de l'échantillon, en considérant l'indépendance des réalisations, est définie comme suit :

$$L(\mathbf{x}, \lambda) = \prod_{i=1}^l f_\lambda(x^i),$$

L'estimateur du maximum de vraisemblance est alors défini comme suit :

$$\hat{\lambda}(\mathbf{x}) = \underset{\lambda}{\operatorname{argmax}} L(\mathbf{x}, \lambda)$$

Dans le cas d'une variable X discrète, les paramètres sont les probabilités élémentaires $\{p_1, \dots, p_n\}$ dans le cas de n valeurs possibles pour θ .

Lorsque nous disposons de peu de réalisations d'une variable aléatoire, nous pouvons utiliser des informations disponibles *a priori* concernant les paramètres de la distribution de la variable aléatoire pour estimer certaines quantités d'intérêt en utilisant l'inférence bayésienne.

Définition 7. Inférence Bayésienne Considérons une variable aléatoire X de loi de probabilité connue paramétrée par λ . Dans l'inférence statistique classique, le paramètre λ est considéré comme une valeur inconnue à estimer, par exemple, par maximum de vraisemblance. L'originalité de l'inférence bayésienne consiste à considérer que le paramètre λ est aussi une variable aléatoire. La loi de λ , notée $\pi(\lambda)$, est appelée *loi a priori* en référence au fait que c'est l'information dont on dispose sur λ avant celle apportée par les observations. La loi du paramètre λ étant données les réalisations \mathbf{x} est appelée *loi a posteriori* et notée $\pi(\lambda|\mathbf{x})$. D'après le théorème de Bayes, cette dernière s'exprime comme suit :

$$\pi(\lambda|\mathbf{x}) = \frac{f(\mathbf{x}|\lambda) \times \pi(\lambda)}{f(\mathbf{x})}$$

où $f(\mathbf{x}|\lambda)$ est la fonction de vraisemblance des observations sachant la valeur du paramètre λ et $f(\mathbf{x}) = \int f(\mathbf{x}|\lambda) \pi(\lambda) d\lambda$ (théorème des probabilités totales).

Le problème majeur de l'inférence bayésienne est le choix de la loi a priori $\pi(\lambda)$. Dans la majorité des cas, en absence d'information, c'est la loi la moins informative possible qui est utilisée (la loi uniforme, la distribution de Jeffrey (Jeffreys, 1946), etc). On dit que la loi des données est *conjuguée* avec la loi a priori si la loi a priori et la loi a posteriori appartiennent à la même famille de lois. Les lois conjuguées sont souvent utilisées pour simplifier les calculs des intégrales. Par exemple, la loi de Dirichlet est conjuguée avec la loi multinomiale (les lois a priori et a posteriori sont toutes les deux des Dirichlet).

Problème de représentation de l'imprécision avec la théorie des probabilités La théorie des probabilités est capable de représenter l'incertitude concernant la valeur ν prise par une variable θ , en attribuant une distribution de probabilité sur l'univers Θ . Il s'agit bien d'incertitude car la réponse apportée à la question "Quelle est la valeur ν ?" est $\nu = \theta_i$ avec une probabilité de p_i . Ce qui signifie qu'on ne sait pas si la proposition $\nu = \theta_i$ est vraie ou fausse. Cependant la théorie des probabilités ne permet pas de représenter l'imprécision (lorsque les informations sont incomplètes pour répondre à la question).

Exemple 3. Prenons l'exemple d'une expérience consistant à tirer une balle dans une boîte contenant 10 balles dont 4 rouges, 2 bleues, 2 vertes, 1 noire et 1 balle dont la couleur est inconnue (mais est soit rouge, bleue, verte ou noire). Essayons de caractériser le comportement de la variable aléatoire θ de la couleur de la balle que l'on obtient lors du prochain tirage. Notons que l'univers considéré ici est $\Theta = \{\text{rouge, bleue, verte, noire}\}$. Nous devons affecter une probabilité à chaque valeur que peut prendre θ . Le problème est qu'en fonction de la couleur de la dernière balle, nous obtenons des probabilités différentes. Par exemple si la dernière balle est rouge nous avons $p(\text{rouge}) = 1/2$, $p(\text{bleue}) = 1/5$, $p(\text{verte}) = 1/5$, $p(\text{noire}) = 1/10$. Dans le cas où la dernière balle est bleue, nous avons une autre distribution de probabilité, $p(\text{rouge}) = 2/5$, $p(\text{bleue}) = 3/10$, $p(\text{verte}) = 1/5$, $p(\text{noire}) = 1/10$. Finalement une seule distribution de probabilité ne suffit pas pour caractériser le comportement de θ . Dans cet exemple l'incertitude sur la valeur ν que prend la variable θ découle de l'existence de plusieurs vérités possibles. L'imprécision vient de l'information sur la dernière balle, que l'on peut reformuler "la dernière balle est rouge, bleue, verte ou noire". Cette information est imprécise car elle ne nous permet pas de répondre complètement à la question de la nature précise de la balle.

Dans la pratique, en présence d'imprécision dans le cadre probabiliste, nous sommes obligés de faire des hypothèses pour construire une unique distribution de probabilité. En particulier, lorsque les réalisations de θ sont imprécises, l'usage est de faire l'hypothèse de l'équiprobabilité. Ainsi l'information imprécise est assimilée à de l'incertitude.

Exemple 4. Prenons l'exemple d'une variable θ représentant le futur gagnant d'une élection qui oppose deux candidats θ_1 et θ_2 . Nous avons $\Theta = \{\theta_1, \theta_2\}$. Nous nous posons la question "Qui sera ν , le gagnant de l'élection?". Considérons les deux situations suivantes :

1. Dans la première situation, nous disposons des résultats d'un sondage qui montre l'information "les deux candidats ont autant de chances de gagner". Cette information incertaine s'exprime avec la distribution de probabilité : $p(\theta_1) = 0.5$ et $p(\theta_2) = 0.5$.
2. Dans la seconde situation, nous savons seulement que le gagnant sera l'un des deux candidats. Cette information est imprécise, donc pour construire une distribution de probabilités, nous devons faire une hypothèse. Avec l'hypothèse de l'équiprobabilité, nous donnons à chaque événement la même probabilité. Nous obtenons la distribution de probabilités : $p(\theta_1) = 0.5$ et $p(\theta_2) = 0.5$.

Dans ces deux situations, nous disposons d'informations affectées par des types d'incertitudes différentes mais la distribution de probabilité utilisée pour les représenter est la même. Nous voyons avec cet exemple que les informations imprécises ne sont pas fidèlement représentées par les probabilités et qu'elles sont assimilées à des informations incertaines si l'on fait l'hypothèse de l'équiprobabilité.

2.1.2 Probabilités imprécises

Les probabilités imprécises proposent un cadre permettant de capturer l'incertitude et l'imprécision sur θ sous la forme d'un ensemble convexe de probabilités, appelé *ensemble crédal* (Walley et Peter, 1991) et défini comme :

$$\mathcal{P} = \{P : \forall A \subseteq \Theta, \underline{P}(\theta \in A) \leq P(\theta \in A) \leq \overline{P}(\theta \in A)\},$$

avec \underline{P} et \overline{P} étant des probabilités respectivement sous-additive et sur-additive appelées probabilités inférieure et supérieure, c'est-à-dire $\forall A, B \subseteq \Theta, A \cap B = \emptyset$,

$$\underline{P}(A) + \underline{P}(B) \leq \underline{P}(A \cup B) \text{ et } \overline{P}(A) + \overline{P}(B) \geq \overline{P}(A \cup B).$$

Nous obtenons l'égalité pour ces deux inéquations dans le cas de probabilités précises, i.e., $\underline{P}(A) = \overline{P}(A)$ quelque soit $A \subseteq \Theta$. Le cas extrême opposé $\underline{P}(A) = 0$ et $\overline{P}(A) = 1$ représente la situation d'ignorance complète concernant l'évènement A .

Exemple 5. Reprenons l'exemple 3 (section 2.1.1), du tirage d'une balle dont la couleur appartient à l'ensemble $\Theta = \{\text{rouge}, \text{bleue}, \text{verte}, \text{noire}\}$. Avec les probabilités imprécises, nous pouvons représenter l'incertitude et l'imprécision concernant la couleur de la balle tirée par l'ensemble crédal $\mathcal{P} = \{P_1, P_2, P_3, P_4\}$ où P_1, P_2, P_3 et P_4 correspondent aux mesures de probabilité que l'on obtient pour chaque possibilité pour la dernière balle : P_1 est caractérisée par la distribution p_1 :

$$p_1(\text{rouge}) = 1/2, p_1(\text{bleue}) = 1/5, p_1(\text{verte}) = 1/5, p_1(\text{noire}) = 1/10,$$

P_2 par la distribution p_2 :

$$p_2(\text{rouge}) = 2/5, p_2(\text{bleue}) = 3/10, p_2(\text{verte}) = 1/5, p_2(\text{noire}) = 1/10,$$

P_3 par la distribution p_3 :

$$p_3(\text{rouge}) = 2/5, p_3(\text{bleue}) = 1/5, p_3(\text{verte}) = 3/10, p_3(\text{noire}) = 1/10,$$

et P_4 par la distribution p_4 :

$$p_4(\text{rouge}) = 2/5, p_4(\text{bleue}) = 1/5, p_4(\text{verte}) = 1/5, p_4(\text{noire}) = 1/5.$$

Exemple 6. Dans l'exemple 4 (section 2.1.1), pour la seconde situation, alors que nous ne disposons d'aucune information sur le gagnant ν de l'élection à deux candidats, θ_1 et θ_2 , nous ne sommes plus obligés de faire l'hypothèse de l'équiprobabilité. Nous pouvons représenter l'incertitude et l'imprécision de la seconde situation concernant ν par l'ensemble crédal $\mathcal{P} = \{P : \forall \theta \in \{\theta_1, \theta_2\}, 0 \leq P(\theta) \leq 1\}$, (situation d'ignorance complète).

Les probabilités imprécises offrent un cadre pour représenter l'incertitude et l'imprécision mais en pratique déterminer \underline{P} et \overline{P} peut être compliqué. La modélisation de l'incertitude et de l'imprécision à propos d'une variable θ par un ensemble crédal est possible soit sur la base d'information experte, soit à partir de données empiriques en utilisant des techniques statistiques comme les intervalles de confiance (De Campos et al., 1994). Il existe également un modèle de probabilités imprécises basé sur le raisonnement bayésien proposé dans (Walley et Peter, 1991). Ce modèle est une version imprécise du modèle de Dirichlet.

2.1.3 Fonctions de croyance

La théorie des fonctions de croyance propose un cadre pour représenter l'incertitude et l'imprécision concernant la valeur ν prise par la variable θ . Deux niveaux sont considérés lors de l'introduction des fonctions de croyance : les niveaux *crédal* et *pignistique*. Au niveau crédal, les croyances sont appréhendées et quantifiées par des fonctions de croyance, tandis qu'au niveau pignistique ou au niveau décision, les croyances sont quantifiées à l'aide de distributions de probabilité.

Niveau crédal : représentation Une *fonction de masse*, également appelée *basic belief assignment* (bba) en anglais, est une fonction $m : 2^\Theta \rightarrow [0, 1]$ satisfaisant

$$\sum_{A \subseteq \Theta} m(A) = 1.$$

Pour un ensemble $A \subseteq \Theta$, la quantité $m(A)$ est interprétée comme la masse de croyance allouée exactement à l'ensemble A et non à des sous-ensembles plus spécifiques de A . En d'autres termes, $m(A)$ peut être interprétée comme la probabilité qu'on ne sache rien de plus que " $\theta \in A$ " (Shafer, 1976). Habituellement, les croyances ne peuvent pas être allouées à l'ensemble vide, m doit alors également satisfaire $m(\emptyset) = 0$. Cette condition traduit une hypothèse de monde fermé (Smets, 1988). Dans certains ouvrages (Smets et Kennes, 1994) $m(\emptyset) \geq 0$ est considéré et $m(\emptyset)$ quantifie un conflit. Dans ce manuscrit, nous considérons $m(\emptyset) = 0$.

Exemple 7. Revenons à l'exemple 3 (section 2.1.1) du tirage d'une balle dans une boîte contenant 4 balles rouges, 2 bleues, 2 vertes, 1 noire et 1 balle dont la couleur est inconnue (mais est soit rouge, bleue, verte ou noire). À partir de ces informations nous pouvons construire une fonction de masse m pour décrire le tirage : $m(\{\text{rouge}\}) = 1/5$, $m(\{\text{bleue}\}) = 3/10$, $m(\{\text{verte}\}) = 1/5$, et $m(\{\text{rouge, bleue, verte, noire}\}) = 1/10$.

En pratique la manipulation de fonctions de masse peut être coûteuse en calculs et stockage car à n'importe quel sous-ensemble de Θ peut être affectée une masse. Il est donc nécessaire d'adopter une notation plus concise :

- *Éléments focaux* : les éléments $A \in 2^\Theta$ tels que $m(A) > 0$ sont appelés éléments focaux et forment un ensemble noté \mathbb{F} . Ainsi une fonction de masse est complètement caractérisée par les masses données aux éléments focaux et dont la somme fait 1.
- *Corps d'évidence* : la paire (m, \mathbb{F}) est appelée corps d'évidence.

Il existe certaines fonctions de masses particulières :

- *Fonction de masse catégorique* : une fonction de masse est dite *catégorique* si elle n'a qu'un élément focal.
- *Fonction de masse vide* : une fonction de masse qui alloue une masse 1 à l'ensemble total Θ est appelée *fonction de masse vide*, (*vacuous mass function*, en anglais). Cette fonction de masse représente l'ignorance totale. Nous notons la fonction de masse vide $m_?$.
- *Fonction de masse bayésienne* : une fonction de masse dite *bayésienne* est telle que tous ses éléments focaux sont des singletons. Ce type de fonction de masse est équivalent à une distribution de probabilité.
- *Fonction de masse consonante* : une fonction de masse est dite *consonante* lorsque que les éléments focaux sont emboîtés (les éléments focaux A_1, A_2, A_3, \dots sont tels que $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$).
- *Fonction de masse à support simple* : Une fonction de masse est dite à *support simple* lorsque qu'elle n'a que deux éléments focaux et que l'un des deux est Θ .

Une fonction de masse induit deux autres fonctions d'ensemble qui peuvent être utilisées pour faire de l'inférence :

- La *fonction de croyance* $Bel : 2^\Theta \rightarrow [0, 1]$, qui quantifie la croyance totale en A comme la somme de toutes les masses des sous-ensembles de A :

$$Bel(A) = \sum_{B \subseteq \Theta, B \subseteq A} m(B).$$

- La *plausibilité* de A , $Pl : 2^\Theta \rightarrow [0, 1]$, qui quantifie la masse maximale qui pourrait être allouée à A :

$$Pl(A) = \sum_{B \subseteq \Theta, B \cap A \neq \emptyset} m(B).$$

Exemple 8. *Considérons l'exemple suivant afin d'illustrer les fonctions Bel et Pl . Prenons $\Theta = \{\theta_1, \theta_2, \theta_3\}$ et la fonction de masse suivante :*

$$m(A) = \begin{cases} 0.8 & \text{si } A = \{\theta_1, \theta_2\}, \\ 0.1 & \text{si } A = \{\theta_2\} \text{ ou } A = \{\theta_3\}, \\ 0 & \text{sinon.} \end{cases}$$

La Fig. 2.1 montre les fonctions Bel et Pl associées à m .

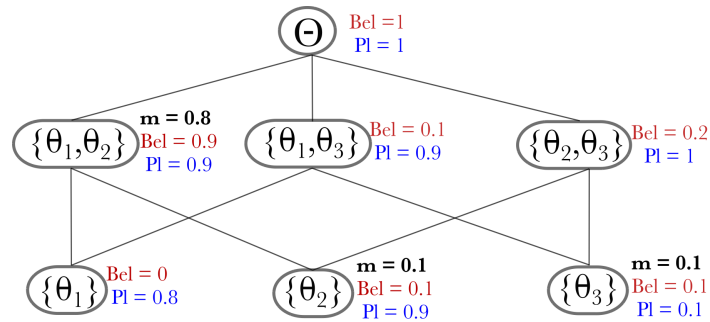


FIG. 2.1 Exemple de fonction de masse, m et fonctions Bel et Pl associées.

Niveau pignistique/décision Dans le modèle des croyances transférables (Smets, 1988), la décision est prise au niveau pignistique. Les informations évidentielles sont transférées dans un cadre probabiliste au moyen de la distribution de probabilité pignistique $betP_m$, pour $\theta \in \Theta$,

$$betP_m(\theta) = \sum_{A \subseteq \Theta, A \ni \theta} \frac{m(A)}{|A|}, \quad (5)$$

où $|A|$ désigne le cardinal de A .

Exemple 9. *Reprenant la fonction de masse de l'exemple 8, nous obtenons : $betP_m(\theta_1) = 0.4$, $betP_m(\theta_2) = 0.5$, $betP_m(\theta_3) = 0.1$. Nous constatons que la pignistique attribue une probabilité strictement positive au singleton $\{\theta_1\}$ alors que nous ne disposons pas d'information précise concernant ce singleton ($m(\{\theta_1\}) = 0$).*

Nous verrons dans la sous-section 2.3.1, comment décider à partir d'une distribution de probabilité. Notons déjà que baser la décision sur la distribution de probabilité pignistique est encore plus risqué puisque nous perdons la représentation de l'imprécision. Nous verrons d'autres méthodes de décision basées notamment sur les fonctions Bel ou Pl dans la sous-section 2.3.2.

Combinaison de fonctions de masse L'un des outils les plus importants proposé dans le cadre des fonctions de croyance est l'outil de combinaison d'informations différentes provenant de plusieurs sources. La procédure de combinaison la plus populaire utilisée dans le cadre des fonctions de croyance est la combinaison de Dempster (combinaison conjonctive) qui suppose que toutes les sources d'information sont indépendantes, fiables et qu'il n'y a pas de conflit entre elles. Pour combiner $k > 0$, fonctions de masse $m_1 \dots, m_k$, la combinaison de Dempster, notée \oplus est définie comme suit,

$$m_1 \oplus \dots \oplus m_k(A) = \begin{cases} 0 & \text{si } A = \emptyset, \\ \frac{m_1 \circledast \dots \circledast m_k(A)}{1 - m_1 \circledast \dots \circledast m_k(\emptyset)} & \text{sinon.} \end{cases}$$

où

$$m_1 \circledast \dots \circledast m_k(A) = \sum_{B_1 \cap \dots \cap B_k = A} \prod_{j=1}^k m_j(B_j).$$

La quantité $m_1 \circledast \dots \circledast m_k(\emptyset)$ est appelée *degré de conflit*. Lorsque ce degré de conflit est égal à 1, c'est qu'il y a au moins deux masses logiquement contradictoires et donc que la combinaison de Dempster n'est pas possible. Notons que la règle conjonctive est aussi utilisée dans certains travaux en considérant une masse non nulle pour l'ensemble vide.

Exemple 10. Prenons par exemple une fonction de masse vide m_1 , une fonction de masse m_2 à support simple et une fonction de masse catégorique m_3 définies sur $\Theta = \{\theta_1, \theta_2, \theta_3\}$ par :

$$m_1 = m_?, \quad m_2(A) = \begin{cases} 0.9 & \text{si } A = \{\theta_2\}, \\ 0.1 & \text{si } A = \Theta, \\ 0 & \text{sinon.} \end{cases} \quad m_3(\{\theta_1\}) = 1.$$

Voyons les masses résultantes des combinaisons de Dempster suivantes :

- La combinaison de Dempster de m_1 et m_2 (avec un degré de conflit égal à 0) est une fonction à support simple telle que :

$$\begin{aligned} m_1 \oplus m_2(\{\theta_2\}) &= 0.9, \\ m_1 \oplus m_2(\Theta) &= 0.1. \end{aligned}$$

- La combinaison de Dempster de m_1 et m_3 (avec un degré de conflit égal à 0) est une fonction catégorique :

$$m_1 \oplus m_3(\{\theta_1\}) = 1.$$

- La combinaison de Dempster de m_2 et m_3 (avec un degré de conflit égal à 0.9) est une masse catégorique telle que :

$$m_2 \oplus m_3(\{\theta_1\}) = 1.$$

- La combinaison de Dempster de m_1 , m_2 et m_3 (avec un degré de conflit égal à 0.9) est une masse catégorique telle que :

$$m_1 \oplus m_2 \oplus m_3(\{\theta_1\}) = 1$$

Remarquons que cette combinaison resserre la masse sur les intersections (perte d'information imprécise). Cette combinaison peut donc être risquée notamment lorsque le degré de conflit est important.

Une autre combinaison courante est la règle de combinaison disjonctive, qui suppose également que les sources d'information sont indépendantes, mais qui convient aux situations où il n'y aurait même qu'une source d'information fiable (Denceux, 2008). La combinaison disjonctive est donnée ci-dessous pour deux fonctions de masse.

$$m_1 \oplus_2(A) = \begin{cases} \sum_{B \cup C = A} m_1(B)m_2(C) & \text{si } A \neq \emptyset, \\ 0 & \text{sinon.} \end{cases}$$

Dans le cas de plusieurs sources (> 2), la combinaison disjonctive de toutes les masses est calculée en utilisant la propriété associative de la règle.

Exemple 11. Avec les fonctions de masses de l'exemple 10, nous obtenons les combinaisons disjonctives suivantes :

- La combinaison disjonctive de m_1 et m_2 est la fonction de masse vide :

$$m_1 \oplus_2 = m_?$$

- La combinaison disjonctive de m_1 et m_3 est la fonction de masse vide :

$$m_1 \oplus_3 = m_?$$

- La combinaison disjonctive de m_2 et m_3 est la fonction à support simple :

$$\begin{aligned} m_2 \oplus_3(\{\theta_1, \theta_2\}) &= 0.9, \\ m_2 \oplus_3(\Theta) &= 0.1. \end{aligned}$$

- La combinaison disjonctive de m_1 , m_2 et m_3 est la fonction de masse vide :

$$m_1 \oplus_2 \oplus_3 = m_?$$

Nous voyons avec le dernier exemple que la combinaison disjonctive a tendance à mettre la masse sur les grands ensembles. Nous notons également que nous pouvons obtenir de nouveaux éléments focaux. Dans cet exemple, l'ensemble $\{\theta_1, \theta_2\}$ n'était l'élément focal d'aucune des masses combinées.

Nous pouvons également combiner les fonctions de masse en prenant l'opérateur moyen arithmétique noté \oplus . La combinaison de $k > 0$, fonctions de masse $m_1 \dots, m_k$, est définie pour un sous-ensemble A de la façon suivante :

$$m_1 \oplus \dots \oplus m_k(A) = \frac{\sum_{i=1}^k m_i(A)}{k}.$$

Exemple 12. Prenant à nouveau l'exemple 10, nous obtenons les combinaisons de l'opérateur moyen suivantes :

- La combinaison de l'opérateur moyen de m_1 et m_2 est la fonction de masse à support simple :

$$\begin{aligned} m_1 \oplus m_2(\{\theta_2\}) &= 0.45, \\ m_1 \oplus m_2(\Theta) &= 0.55. \end{aligned}$$

- La combinaison de l'opérateur moyen de m_1 et m_3 est la fonction de masse à support simple :

$$\begin{aligned} m_1 \oplus m_3(\{\theta_1\}) &= 0.5, \\ m_1 \oplus m_3(\Theta) &= 0.5. \end{aligned}$$

- La combinaison de l'opérateur moyen de m_2 et m_3 est la fonction de masse :

$$m_2 \oplus m_3(A) = \begin{cases} 0.5 & \text{si } A = \{\theta_1\}, \\ 0.45 & \text{si } A = \{\theta_2\}, \\ 0.05 & \text{si } A = \Theta, \\ 0 & \text{sinon.} \end{cases}$$

- La combinaison de l'opérateur moyen de m_1 , m_2 et m_3 est la fonction de masse :

$$m_1 \oplus m_2 \oplus m_3(A) = \begin{cases} \simeq 0.333 & \text{si } A = \{\theta_1\}, \\ 0.3 & \text{si } A = \{\theta_2\}, \\ \simeq 0.366 & \text{si } A = \Theta, \\ 0 & \text{sinon.} \end{cases}$$

Remarquons que cette combinaison ne fait pas apparaître de nouveaux éléments focaux, contrairement la combinaison disjonctive, nous pouvons obtenir de nouveaux éléments focaux qui ne sont pas partagés avec ceux des masses d'origines. Dans l'exemple précédent, la masse combinée, $m_1 \oplus m_2 \oplus m_3$ a $\{\theta_1\}$ pour élément focal alors que ni m_1 ni m_2 n'ont cet élément focal. Par ailleurs, cette combinaison est adaptée dans le cas de conflit et a l'avantage d'avoir une faible complexité. D'autres combinaisons ont été proposées pour répondre à des situations particulières de sources d'information. Citons la combinaison conjonctive de Smets ((Smets, 1990), la combinaison conjonctive de Yager (Yager, 1987), la combinaison disjonctive de Dubois (Dubois et Prade, 1988), les combinaisons robustes (Florea et al., 2009) etc.

2.1.4 Théorie des possibilités

La théorie des possibilités propose un cadre pour représenter et manipuler des informations imprécises et incertaines par l'utilisation de deux mesures duales non additives. Considérons que nous disposons de l'information suivante concernant la valeur ν prise par $\theta : \nu \in E$ avec E un sous-ensemble de Θ . La théorie des possibilités s'intéresse aux notions de *possibilité* et de *nécessité* pour des événements :

- Un événement $A \subseteq \Theta$ est *possible* si et seulement si $A \cap E \neq \emptyset$.
- Un événement $A \subseteq \Theta$ est *nécessaire* (certain) si et seulement si $A \subseteq E$.

Formellement ces notions correspondent aux fonctions d'ensemble, Π (possibilité) et N (nécessité) suivante :

$$\Pi(A) = \begin{cases} 1 & \text{si } A \cap E \neq \emptyset, \\ 0 & \text{sinon,} \end{cases}$$

et

$$N(A) = \begin{cases} 1 & \text{si } A \subseteq E, \\ 0 & \text{sinon,} \end{cases}$$

Nous remarquons que si un évènement A est certain (resp. possible) alors l'évènement \bar{A} n'est pas possible (resp. pas certain), $N(A) = 1 - \Pi(\bar{A})$. Ces fonctions ont les propriétés suivantes :

$$\begin{aligned}\Pi(\Theta) &= 1, \\ \Pi(\emptyset) &= 0, \\ \Pi(A \cup B) &= \max(\Pi(A), \Pi(B)), \forall A, B \subseteq \Theta,\end{aligned}$$

et

$$\begin{aligned}N(\Theta) &= 1, \\ N(\emptyset) &= 0, \\ N(A \cap B) &= \min(N(A), N(B)), \forall A, B \subseteq \Theta.\end{aligned}$$

La théorie des possibilités généralise les fonctions Π et N en les autorisant à prendre leurs valeurs dans $[0, 1]$. Cette généralisation permet de considérer qu'un évènement est possible ou certain à un certain degré.

Il est possible de définir Π à partir de ses valeurs pour les singletons. Dans ce cas, nous obtenons une distribution de possibilité π :

$$\pi(\theta) = \Pi(\{\theta\}), \forall \theta \in \Theta.$$

Nous pouvons rencontrer les deux situations extrêmes suivantes. En cas d'ignorance totale, nous avons pour tout $\theta \in \Theta$, $\pi(\theta) = 1$. Dans l'autre cas extrême, en cas d'information précise, nous avons pour un $\theta_i \in \Theta$, $\pi(\theta_i) = 1$ et $\pi(\theta_j) = 0$, pour tous $j \neq i$.

2.1.5 Espérance mathématique et mesures floues

Dans les trois dernières théories de l'incertain présentées, probabilités imprécises, fonctions de croyance et possibilités, nous avons considéré des mesures non-additives (qui ne respectent pas la propriété d'additivité des probabilités (3)). Avec de telles mesures, nous ne pouvons plus appliquer l'espérance classique (équation (4)) définie pour les probabilités, sur laquelle se base généralement la décision. En effet, dans le cas classique, l'espérance est une intégrale définie pour une mesure additive. Dans cette sous-section nous présentons une alternative à l'espérance classique lorsque nous ne disposons pas de mesure additive, l'*intégrale de Choquet*. A cette fin, nous introduisons la notion de *mesure floue*.

Définition 8. Mesure floue Une mesure floue telle que considérée dans (Murofushi et Sugeno, 1989) est une fonction d'ensemble non-nulle μ définie sur un ensemble fini non vide Θ , $\mu : 2^\Theta \rightarrow [0, 1]$, telle que :

$$\mu(\emptyset) = 0,$$

et monotone, c'est-à-dire $\forall A, B \subseteq \Theta$ tels que $A \subseteq B$, nous avons

$$\mu(A) \leq \mu(B).$$

Les mesures non-additives rencontrées précédemment, Bel , Pl , Π et N sont des mesures floues.

L'intégrale de Choquet (Labreuche et Grabisch, 2003) est l'extension de l'intégrale classique aux mesures floues.

Définition 9. Intégrale de Choquet associée à une mesure floue Etant donné un vecteur de valeurs positives réelles $y = (y_1, \dots, y_n) \in \mathbb{R}^{+n}$, son intégrale de Choquet associée à une mesure floue μ est définie comme

$$C_\mu(y) = \sum_{i=1}^n (y_{\sigma(i)} - y_{\sigma(i-1)}) \mu(\{\theta_{\sigma(i)}, \theta_{\sigma(i+1)}, \dots, \theta_{\sigma(n)}\}) \quad (15)$$

où $0 = y_{\sigma(0)} \leq y_{\sigma(1)} \leq y_{\sigma(2)} \leq \dots \leq y_{\sigma(n)}$ (σ est une permutation sur $\{1, \dots, n\}$).

Remarque 1. Dans son application aux fonctions de croyances, l'intégrale de Choquet peut être interprétée comme la plus faible espérance prise sur toutes les probabilités P telles que $Bel \leq P \leq Pl$, c'est-à-dire toutes les probabilités délimitées par nos connaissances imprécises. Lorsque $Bel = Pl$, alors l'équation (15) est simplement l'opérateur d'espérance mathématique standard.

Exemple 13. Prenons un exemple de calcul d'espérance avec l'intégrale de Choquet dans le cas des fonctions de croyance. Considérons trois candidats $\theta_1, \theta_2, \theta_3$ à une élection. Supposons que nous nous intéressions au pourcentage du bond de la croissance économique qu'on obtiendra avec le gagnant de l'élection par rapport au prédécesseur et que nous connaissons le pourcentage du bond de la croissance économique qu'on obtiendra avec chaque candidat : Respectivement 181%, 168% et 176%. Supposons que nous disposions d'information concernant l'identité du gagnant de l'élection sous la forme de la fonction de masse : $m(\{\theta_1\}) = 0.2$, $m(\{\theta_1, \theta_2, \theta_3\}) = 0.3$, $m(\{\theta_3\}) = 0.4$, $m(\{\theta_1, \theta_2\}) = 0.1$. Avec ces informations, nous pouvons calculer le pourcentage du bond de la croissance économique espéré du gagnant en utilisant l'espérance de Choquet :

$$\begin{aligned} C_{Bel}(y) &= \sum_{i=1}^n (y_{\sigma(i)} - y_{\sigma(i-1)}) Bel(\{\theta_{\sigma(i)}, \theta_{\sigma(i+1)}, \dots, \theta_{\sigma(n)}\}) \\ &= y_{\sigma(1)} + (y_{\sigma(2)} - y_{\sigma(1)}) Bel(\{\theta_{\sigma(2)}, \theta_{\sigma(3)}\}) + (y_{\sigma(3)} - y_{\sigma(2)}) Bel(\{\theta_{\sigma(3)}\}) \\ &= 168 + (176 - 168) \times 0.6 + (181 - 176) \times 0.2 \\ &= 173.8\% \end{aligned}$$

Nous pouvons également calculer l'espérance classique à partir de la probabilité pignistique. Ainsi nous obtenons un pourcentage du bond de la croissance économique espéré plus important 176.55%.

2.1.6 Logique floue

La logique floue permet de manipuler des informations linguistiques graduelles concernant la valeur ν de la variable θ dans Θ . La représentation de ces informations est faite par la notion de sous-ensemble flou associé à une fonction d'appartenance et la notion de variable linguistique.

Définition 10. Sous-ensemble flou et fonction d'appartenance Formellement, un *sous-ensemble flou* (ou *terme linguistique*) A de Θ est défini par une *fonction d'appartenance* $\mu_A : \Theta \rightarrow [0, 1]$ qui quantifie dans quelle mesure la variable θ appartient à A . La plupart du temps, un sous-ensemble flou A exprime une propriété partagée plus au moins par les éléments de Θ . Dans ce cas, pour $\theta \in \Theta$, $\mu_A(\theta)$ est le degré auquel θ satisfait cette propriété.

Exemple 14. Pour illustrer la notion de sous-ensembles flous ou termes linguistiques, nous pouvons prendre l'exemple suivant. Posons θ la variable "Température" et A l'ensemble flou "Tiède" (voir la figure 2.2). En dessous de $10^\circ C$, la température n'est pas considérée comme

tiède. Entre 10°C et 15°C, la température n'est pas encore complètement tiède. Entre 15°C et 25°C, la température est considérée complètement tiède.

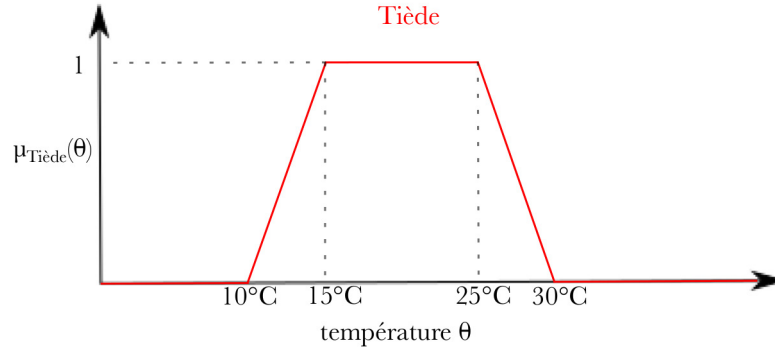


FIG. 2.2 Interprétation de l'ensemble flou "tiède"

Définition 11. Variables linguistiques Les variables définies sur un ensemble de référence Θ dont les valeurs sont des termes linguistiques définis sur Θ sont appelées *variables linguistiques*. Formellement, une variable linguistique V définie sur Θ est un triplet (V, Θ, T_V) , où $T_V = \{A_1, A_2, \dots\}$ est l'ensemble des termes caractérisant V sur Θ . Par exemple si V est la variable linguistique température définie sur $\Theta = \mathbb{R}$, nous pourrions la décrire par $T_V = \{\text{"Froid"}, \text{"Tiède"}, \text{"Chaud"}\}$.

Le traitement d'une information dans la théorie de la logique floue est effectué en deux étapes générales :

1. La *fuzzification* est le processus consistant à partitionner un espace en un ensemble de termes linguistiques. La méthode "Fuzzy grid partitionning" est la méthode la plus connue pour la fuzzification (Wang et Mendel, 1992).
2. La *défuzzification* est le processus contraire qui consiste à obtenir une sortie numérique à partir d'un sous-ensemble flou. Il existe plusieurs techniques de défuzzification. La plus connue est celle du *centre de gravité*.

La fuzzification permet de manipuler des informations exprimées dans le langage humain via des propositions floues et des règles floues :

Définition 12. Propositions floues A partir d'une variable linguistique V , nous pouvons exprimer des *propositions floues élémentaires* du type " V est A " où A est un terme linguistique. Un exemple de proposition floue élémentaire est "La température est tiède". A partir des propositions floues élémentaires, il est possible de construire des *propositions floues générales* comme des compositions de propositions floues élémentaires telle que :

$$\text{"}V \text{ est } A \text{ et } W \text{ est } B\text{"},$$

où V et W sont des variables linguistiques et A, B sont des termes linguistiques.

Définition 13. Règles floues La notion de proposition floue permet de manipuler des règles floues de type :

$$SI < \text{antécédent} > ALORS < \text{conséquent} >,$$

où les antécédents et les conséquents sont des propositions floues. Par exemple, "SI la température de la rivière est froide ALORS peu de personnes s'y baigne".

Les règles floues sont à la base de plusieurs modèles permettant l'approximation d'une fonction liant plusieurs entrées à plusieurs sorties. Le besoin d'approximation de fonction est nécessaire dans les situations suivantes : lorsque la structure mathématique de la fonction n'est pas précisément connue ; lorsque la fonction est complexe au point que trouver sa forme mathématique précise est impossible ou pratiquement irréalisable, etc. Les modèles à base de règles floues, appelés aussi les modèles FRBS (Fuzzy Rule-Based Systems), les plus connus sont le modèle de Mamdani et le modèle TSK (Takagi-Sugeno-Kang). Récemment ces modèles ont été adaptés pour des tâches de classification et de régression. Dans la sous-section 2.3.3, nous introduisons en détails certains modèles FRBCS (Fuzzy Rule-Based Classification Systems) dédiés à la tâche de classification.

2.2 Classification automatique supervisée

Nous nous intéressons à présent à un autre domaine de notre état de l'art, la classification automatique supervisée. Dans cette section, après une description des principes de la classification supervisée (sous-section 2.2.1), nous présentons les méthodes classiquement utilisées pour évaluer les performances des méthodes de classification (sous-section 2.2.2). Enfin nous présentons les principaux algorithmes de classification supervisée (sous-section 2.2.3), ainsi que plusieurs méthodes d'optimisation permettant d'améliorer leur efficacité et/ou leur performance (en temps et/ou en espace) (sous-section 2.2.4).

2.2.1 Principe de la classification supervisée

Considérons un objet représenté par un vecteur de caractéristiques (ou attributs) $\mathbf{x} = (x_1, x_2, \dots, x_d)$, $d > 0$ où chaque caractéristique $i \in \{1, \dots, d\}$ prend ses valeurs dans un espace \mathcal{X}_i . On note $\mathcal{X} = \prod_{i=1}^d \mathcal{X}_i$. Considérons également $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ un ensemble de classes (ou catégories). Le principe de la classification automatique supervisée consiste à associer à \mathbf{x} une classe, i.e., un élément de Θ , en prenant connaissance de L objets représentés par les mêmes caractéristique que \mathbf{x} et dont nous connaissons les classes associées. Nous représentons chaque objet j , $j \in \{1, \dots, L\}$ par une paire $(\mathbf{x}^j, \theta^j) \in \mathcal{X} \times \Theta$ que nous appelons *exemple*. Nous appelons, par abus de langage, *classifieur* l'application qui permet d'associer un objet représenté dans \mathcal{X} à une classe dans Θ . Si on note δ_Θ un classifieur qui associe à des objets dans \mathcal{X} des classes de Θ , i.e., $\delta_\Theta : \mathcal{X} \rightarrow \Theta$. $\delta_\Theta(\mathbf{x})$ correspond à la classe associée par le classifieur δ_Θ à \mathbf{x} .

Un classifieur se base sur un modèle pour représenter la règle de classification, ainsi le terme *apprentissage d'un classifieur* fait référence à l'identification des paramètres du modèle sous-jacent au classifieur à partir des exemples d'apprentissage. Il faut noter que deux classifieurs différents peuvent fournir deux classes différentes pour \mathbf{x} . Ainsi, il faut s'attacher à utiliser le classifieur le plus performant dans un contexte donné.

Exemple 15. *Par exemple considérons un problème de diagnostic de maladies dans $\Theta = \{\text{Covid}, \text{Grippe}, \text{Angine}\}$, à partir des caractéristiques $\mathcal{X} = (\text{âge}, \text{toux}, \text{fièvre}, \text{température})$. Le classifieur est appris sur un ensemble d'exemples d'apprentissage constitué des valeurs des caractéristiques et des maladies d'un ensemble de patients. Pour classifier un nouvel exemple (diagnostiquer un patient), nous appliquons le classifieur à un vecteur de caractéristiques du patient, par exemple $\mathbf{x} = (33 \text{ ans}, \text{oui}, \text{non}, 38^\circ\text{C})$. Le classifieur nous fournit ensuite sa sortie, par exemple $\delta_\Theta(\mathbf{x}) = \text{Angine}$.*

2.2.2 Mesures d'évaluation

L'évaluation d'un classifieur δ_Θ est classiquement menée sur une base d'exemples de test labellisés $(\mathbf{x}^t, \theta^t)_{1 \leq t \leq T}$, n'ayant pas servi pour l'apprentissage de δ_Θ . Idéalement ces exemples de test devraient être représentatifs des données et en assez grand nombre. Dans ce cas il suffit d'appliquer δ_Θ à l'ensemble des exemples de test puis de comparer les sorties $(\delta_\Theta(\mathbf{x}^t))_{1 \leq t \leq T}$ aux vraies classes (θ^t) des \mathbf{x}^t , $1 \leq t \leq T$. Afin de simplifier la présentation des mesures d'évaluation, nous commençons par traiter le cas de la classification binaire, i.e., $|\Theta| = 2$.

Cas binaire Considérons $\Theta = \{\text{Positif}, \text{Négatif}\}$. Une approche très pratique pour avoir un premier aperçu du comportement du classifieur est de représenter les résultats dans une matrice appelée matrice de confusion comme illustrée dans la TABLE 2.2 où la répartition par effectif des exemples de test est donnée. Dans cette table a désigne le nombre d'exemples de test dont la vraie classe est "positif" et pour lesquels le classifieur donne "positif" (exemples vrais positifs), b désigne le nombre d'exemples de classe "positif" ayant été classifiés comme "négatif" (exemples faux négatifs). De même, c désigne le nombre d'exemples de test de classe "négatif" classifiés dans la classe "positif" (exemples faux positif) et d désigne les exemples de test de la classe "négatif" attribués à la classe "négatif" par le classifieur.

		Sorties de δ_Θ		Total
		Positif	Négatif	
Classes des exemples de test	Positif	a (vrai positif)	b (faux négatif)	$a + b$
	Négatif	c (faux positif)	d (vrai négatif)	$c + d$
Total		$a + c$	$b + d$	T

TABLE 2.2 – Matrice de confusion

A partir de cette matrice, sont définis les scores suivants :

- Le *taux d'erreur* noté $\epsilon = \frac{c + b}{T}$.
- Le *taux de justesse* (*accuracy* en anglais) noté $acc = \frac{a + d}{T}$.
- Le *rappel* noté $rap = \frac{a}{a + b}$.
- La *précision* noté $prec = \frac{a}{a + c}$.

Les scores acc et ϵ sont très intuitifs et suffisent dans certaines applications. Cependant les scores de rappel et de précision permettent une distinction des erreurs faux positif et faux négatif (ce qui s'apparente aux erreurs de type I et II d'un test statistique). Concrètement, le rappel permet d'évaluer la capacité du classifieur à détecter les cas positifs : parmi les exemples positifs, combien ont effectivement été détectés comme tels. La valeur $rap = 1$ signifie que le classifieur a détecté tous les cas positifs. Cependant cette mesure n'est pas suffisante toute seule. En effet, si le classifieur prédit toujours "positif" quel que soit l'objet \mathbf{x} alors il obtiendra aussi un score de rappel maximal. Le score de précision permet de pénaliser les erreurs de type I (faux positif) : parmi les sorties "positif", lesquelles étaient justes. Suivant les applications et le sens de la classe "positif", il est possible de privilégier le rappel ou la précision. Il est aussi possible

d'intégrer les mesures de rappel et de précision dans une seule mesure d'évaluation comme le F-Score $_{\beta}$ qui traduit le compromis entre ces deux scores :

$$\text{F-Score}_{\beta} = \frac{(1 + \beta^2) \text{prec} \text{ rap}}{\beta^2 \text{prec} + \text{rap}},$$

avec $\beta \geq 0$ choisi suivant la préférence de l'évaluateur : plus β est choisi proche de 0 plus la précision prévaut, plus β est grand, plus c'est le rappel qui est crucial.

Cas multi-classes Les scores rap , prec , F-Score $_{\beta}$ initialement définis dans le cas binaire se généralisent au cas multi-classes, i.e., $|\Theta| > 2$. Ainsi, un score rap_i (resp. prec_i ou F-Score $_{\beta i}$) est défini pour chaque classe i en considérant le schéma "un contre tous". C'est-à-dire les exemples de classe θ_i sont considérés comme de classe "positif" et tous les autres de classe "négatif". Ensuite on définit le score global rap (resp. prec ou F-Score $_{\beta}$) en moyennant les scores rap_i (resp. prec_i ou F-Score $_{\beta i}$).

L'évaluation des sorties d'un classifieur intervient à deux niveaux importants de la classification qui nécessitent des précautions :

1. Lors du choix des paramètres du classifieur dans la phase d'apprentissage.
2. Lors de l'évaluation de la performance du classifieur

Dans le premier cas, il est nécessaire d'éviter une dépendance trop forte aux données d'apprentissage (*sur-apprentissage*). Dans le second cas, nous risquons de biaiser l'évaluation des performances en évaluant seulement sur un ensemble de tests qui n'est pas parfaitement représentatif. Pour répondre à ces problèmes, l'usage est d'évaluer les sorties du classifieur sur le modèle de la *validation croisée*.

Définition 14. Validation croisée La validation croisée est une méthode d'estimation de la performance d'un classifieur basée sur des techniques de ré-échantillonnage. Nous appelons *ensemble de validation*, l'ensemble sur lequel nous souhaitons estimer les performances du classifieur. Cet ensemble peut être un ensemble dédié à part (différent de l'ensemble d'apprentissage et de l'ensemble de test), ou bien l'ensemble d'apprentissage lorsque l'objectif est l'optimisation des paramètres lors de l'apprentissage, ou l'ensemble de tous les exemples disponibles pour évaluer/comparer les performances d'un classifieur. Un type très répandu de validation croisée est la *validation croisée à k blocs*³, *k-fold cross-validation* en anglais. Dans ce cas, l'ensemble de validation est divisé en k échantillons v_1, \dots, v_k d'exemples. Ensuite pour chaque bloc $j \in \{1 \dots, k\}$, le classifieur est entraîné avec les exemples de $\{v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_k\}$ et nous calculons les sorties du classifieur pour le bloc v_j . La Fig. 2.3 présente la répartition des exemples lors de l'application du classifieur à un bloc v_j . Une fois que le classifieur a prédit une classe pour tous les exemples de l'ensemble de validation, nous pouvons évaluer ses prédictions par la mesure d'évaluation qui nous intéresse.

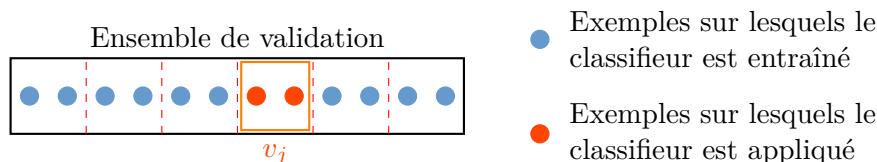


FIG. 2.3 Répartition des exemples d'un ensemble de validation lors du traitement du j -ième bloc.

3. Le cas particulier obtenu lorsque k est la taille de l'ensemble de validation moins un, est appelé *validation croisée d'un contre tous*, *leave-one-out cross-validation* en anglais.

En effectuant plusieurs validations (avec des délimitations de blocs différentes), nous obtenons des évaluations différentes des sorties. En agrégeant les évaluations obtenues (généralement par la moyenne), nous obtenons une estimation plus robuste de la performance du classifieur sur l'ensemble de validation.

2.2.3 Algorithmes de classification supervisée usuels

Dans cette section nous allons introduire brièvement une sélection de classifieurs usuels. Notons que, nous employons fréquemment la notation $p(\cdot|\mathbf{x})$ pour désigner la distribution de probabilité a posteriori discrète sur Θ étant donné un vecteur de caractéristiques \mathbf{x} .

Le classifieur NBC Il s'agit du *classifieur naïf bayésien* (Duda et al., 1973), appelé *Naïf bayes Classifier* (NBC) en anglais, il estime les probabilités que le nouvel exemple $\mathbf{x} = (x_1, \dots, x_d) \in \mathcal{X}$, $d > 0$ corresponde à chacune des classes de Θ . Puis il attribue à \mathbf{x} la classe ayant la probabilité maximale. Pour déterminer ces probabilités, le théorème de Bayes est employé, pour $\theta \in \Theta$:

$$p(\theta|\mathbf{x}) = \frac{p(\theta)p(\mathbf{x}|\theta)}{p(\mathbf{x})}.$$

Sans hypothèse, le calcul de $p(\mathbf{x}|\theta)$ est difficile en pratique lorsque le nombre de caractéristiques est grand. Pour simplifier la tâche, le modèle de NBC fait l'hypothèse d'indépendance des caractéristiques de \mathbf{x} conditionnellement à chaque classe. Ainsi on obtient :

$$p(\theta|\mathbf{x}) = \frac{p(\theta) \prod_{i=1}^d p(x_i|\theta)}{p(\mathbf{x})}.$$

Ensuite il ne reste plus qu'à estimer les $p(x_i|\theta)$ séparément et les probabilités a priori de chaque classe. Malgré l'hypothèse simpliste et rarement respectée en pratique de l'indépendance des caractéristiques conditionnées aux classes, ce classifieur est connu pour obtenir des résultats satisfaisants.

Le classifieur k -nn Il s'agit de l'algorithme des *k plus proches voisins*, *k nearest neighbour* (k -nn) en anglais (Cover et Hart, 1967). Pour traiter le nouvel exemple \mathbf{x} , le classifieur recherche les $k > 0$ plus proches voisins de \mathbf{x} dans la base d'apprentissage relativement à une métrique donnée. Ensuite, la classe majoritaire parmi ses k plus proches voisins est attribuée à \mathbf{x} . La Fig. 2.4 illustre la classification d'un exemple en deux dimensions pour un problème de classification binaire $\Theta = \{\theta_1, \theta_2\}$. Avec $k = 3$, la prédiction pour l'exemple \mathbf{x} est θ_1 (une majorité de 2 voisins sur les 3 sont θ_1). Avec $k = 10$ la prédiction pour \mathbf{x} serait θ_2 (une majorité de 7 voisins sur les 10 sont des θ_2). Le choix de k est déterminant, il est généralement optimisé par une validation croisée lors de l'apprentissage. La principale faiblesse de ce classifieur est que toutes les données d'apprentissage sont nécessaires pour traiter un nouvel exemple, ce qui est coûteux en temps et en espace.

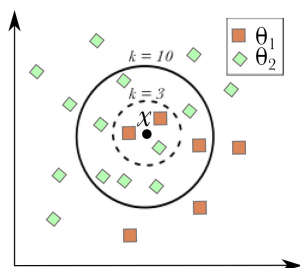


FIG. 2.4 Classification k -nn d'un exemple avec $k = 3$ et $k = 10$.

Le classifieur DT Il s'agit de classification par *arbre de décision*, *decision tree* (DT) en anglais, construit un modèle graphique arborescent constitué d'une racine, de nœuds et de feuilles. A chaque nœud de l'arbre, le classifieur sélectionne la variable et le seuil associé qui maximisent le pouvoir discriminant du test pour optimiser la segmentation ou le partitionnement sur les exemples d'apprentissage. Il existe plusieurs critères possibles pour évaluer la segmentation, comme l'entropie de Shannon (algorithme ID3 et C4.5) ou le critère de Gini (algorithme CART (Breiman et al., 1984)).

Pour la classification de \mathbf{x} , des tests sont effectués aux nœuds pour déterminer la branche à suivre jusqu'à tomber sur une feuille qui corresponde à la classe prédite. La Fig. 2.5 représente l'arbre obtenu avec le jeu de données *iris*⁴ pour lequel $\Theta = \{\text{Setosa}, \text{Virginica}, \text{Versicolor}\}$ et $\mathcal{X} = (\text{Longueur de la pétale}, \text{Largeur de la pétale}, \text{Longueur de la sépale}, \text{Largeur de la sépale})$. Ce classifieur à l'avantage d'être interprétable.

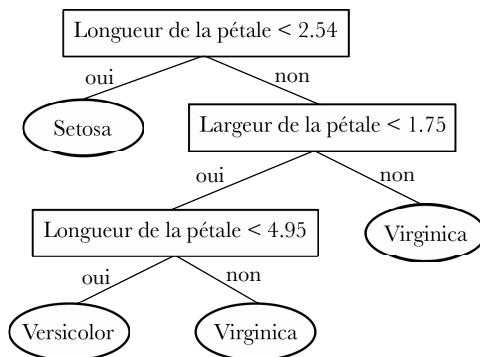


FIG. 2.5 Arbre CART entraîné sur le jeu de données *Iris*

Le classifieur RF Il s'agit du classifieur basé sur les *forêts aléatoires*, *random forest* (RF) en anglais (Ho, 1995). Cette méthode de classification consiste à apprendre plusieurs classifieurs DT sur des sous-échantillons de l'ensemble d'apprentissage. Les sorties de ces classifieurs sont ensuite agrégées pour obtenir la prédiction. Le classifieur RF obtient ainsi de meilleures performances que le classifieur DT mais perd l'avantage d'être facilement interprétable qu'avait le classifieur DT.

Le classifieur SVM Il s'agit des *machines à vecteur de support* (en anglais *support vector machine*, SVM) (Boser et al., 1992; Cortes et Vapnik, 1995). C'est à la base un classifieur

4. Jeux de données introduit dans (Fisher, 1936) très souvent utilisé comme illustration de problème de classification supervisée.

binaire dont l'objectif est de trouver un hyperplan maximisant la distance entre deux classes. Les points les plus proches des hyperplans sont appelés "support vector". Ce sont les seuls points qui entrent en jeu dans le problème de classification. Pour traiter le cas de données qui ne sont pas linéairement séparables, l'idée principale des SVM est de transformer l'espace des données en un espace de plus grande dimension pour lequel une séparation linéaire est possible (Mercer, 1909). Une illustration de cette transformation est montrée dans la Fig. 2.6.

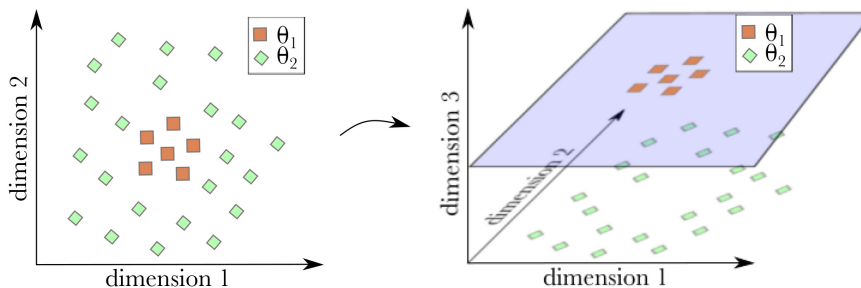


FIG. 2.6 Transformation de l'espace initial.

La généralisation au cas multi-classes est notamment rendue possible par une des options suivantes (qui sont applicables à tous les classifieurs binaires et ne sont pas spécifiques au SVM) :

- *un contre tous* : on construit n classifieurs binaires SVM pour séparer la classe θ_i contre la classe "non θ_i ". La classe θ_j attribuée à \mathbf{x} est celle qui correspond au classifieur ayant la marge la plus grande et prédisant θ_j .
- *un contre un* : $n(n - 1)/2$ classifieurs SVM sont entraînés pour séparer les classes deux à deux. La classe de \mathbf{x} est déterminée par vote majoritaire des prédictions de tous ces classifieurs.

Le classifieur ANN Il s'agit du classifieur basé sur les réseaux de neurones artificiels, artificial neural network (ANN) en anglais (Lettvin et al., 1959). Un neurone artificiel est un modèle simple qui traite des entrées en fonction de leurs poids, d'un biais, (valeur caractéristique au neurone) et d'une fonction d'activation. D'abord, les entrées sont pondérées par des poids puis sommées avec le biais. A la sortie du neurone, la somme traverse une fonction d'activation. La Fig. 2.7 illustre le fonctionnement d'un neurone avec $k > 0$ entrées e_1, \dots, e_k .

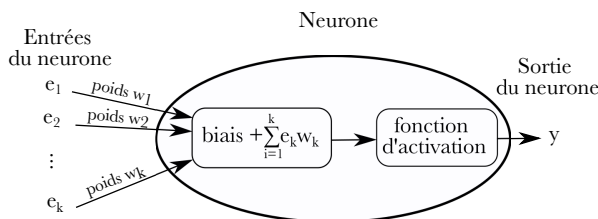


FIG. 2.7 Représentation d'un neurone

Un neurone artificiel prend tout son intérêt lorsqu'il est mis en interaction avec d'autres neurones (lorsque qu'il fait partie d'un réseau de neurones). Les réseaux de neurones sont généralement organisés en *couche d'entrée*, *couches cachées*, et *couches de sortie* (Krenker et al., 2011). Les caractéristiques d'un exemple sont données à la couche d'entrée et traversent tout le réseau jusqu'à la couche de sortie qui fournit la prédiction de l'exemple. L'apprentissage consiste

à chercher les poids et les valeurs de biais qui minimisent l'erreur de sortie lorsqu'on fait passer les exemples d'apprentissage dans le réseau de neurone.

Le classifieur RL Il s'agit du classifieur binaire basé sur la *régression logistique*, *logistic regression* (LR) en anglais (Berkson, 1944). Pour simplifier la présentation, considérons le cas d'un problème de classification binaire. L'idée de la régression logistique est d'écrire le problème de régression classique à l'aide d'une fonction $g : [0, 1] \rightarrow \mathbb{R}$ sous la forme

$$g(p(\theta = \theta_1)) = \beta_0 + \beta_1 x_1 + \dots + \beta_d x_d + \epsilon.$$

Les choix de fonctions g les plus répandus sont *probit*, *log-log* et *logit*. Le vecteur $\beta = (\beta_0, \beta_1, \dots, \beta_d)$ est déterminé sur les exemples de la base d'apprentissage. Pour le traitement d'un nouvel exemple \mathbf{x} , il suffit de calculer la probabilité $p(\theta = \theta_1) = g^{-1}(\beta_0 + \beta_1 \times x_1 + \dots + \beta_d \times x_d)$. Nous avons introduit ce classifieur dans le cas binaire pour simplifier l'explication mais des versions multinomiales de la régression logistique (Venables et Ripley, 2013) permettent de traiter le cas multi-classes.

2.2.4 Optimisation des classifieurs

Il existe de nombreuses méthodes permettant d'améliorer les performances et l'efficacité des classifieurs. Nous n'en présentons ici que trois. Ce sont celles que nous utilisons dans la suite du manuscrit. Il s'agit de méthodes de ré-équilibrage des classes, de gestion des exemples aberrants et de méthodes d'extraction de caractéristiques. Les deux premières permettent d'améliorer la qualité de la classification et la troisième permet de réduire la dimension des vecteurs de caractéristiques sans grande perte d'information.

2.2.4.1 Ré-équilibrage de classes

Il arrive que les effectifs d'exemples soient déséquilibrés suivant les classes. L'apprentissage du classifieur est alors déséquilibré et les exemples de petites classes sont plus souvent mal classés que les exemples de classes dominantes (Sun et al., 2009). De nombreuses solutions ont été proposées pour la gestion de l'apprentissage en cas de déséquilibre de classes (voir (Sun et al., 2009) pour un état de l'art). Les deux principales solutions proposées dans la littérature pour corriger ce biais sont la prise en compte de coûts spécifiques aux exemples d'apprentissage (Pazzani et al., 1994) et le ré-échantillonnage de la base de données. Dans le dernier cas, on distingue deux types d'approches qui peuvent être combinées : les méthodes de sous échantillonnage des classes majoritaires et les méthodes qui ajoutent des exemples synthétiques comme la méthode *synthetic minority over-sampling technique* (SMOTE) proposée dans (Chawla et al., 2002). Nous reviendrons sur la méthode SMOTE dans le chapitre 4.

2.2.4.2 Gestion des exemples aberrants

Certains exemples peuvent sembler inconsistants avec le reste de la base de données, ils sont qualifiés d'aberrants et peuvent nuire à la généralisation du classifieur. Les sources principales de ces exemples aberrants sont les erreurs de labellisation et les variabilités de mesures. L'identification de ces exemples est sujette à de nombreux travaux. Par exemple, dans (John, 1995), les exemples qualifiés de "confusing" sont détectés dans le cadre d'un arbre de décision C4.5, les nœuds de l'arbre sont élagués, seuls les plus informatifs sont conservés. Les exemples concernés par les nœuds supprimés sont ignorés. La gestion de ces exemples n'est pas forcément la suppression, le ré-étiquetage des exemples aberrants dans une classe plus appropriée est également envisagé dans certains travaux comme (Lallich et al., 2002). Dans cette référence, un

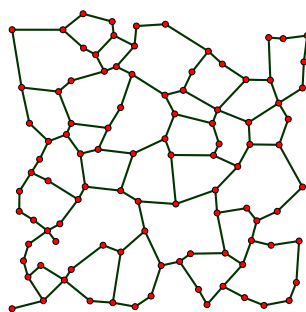


FIG. 2.8 Graphe RNG de 100 points aléatoires dans $[0, 1] \times [0, 1]$ ⁵

exemple d'apprentissage est considéré comme "suspect" si dans son voisinage (défini en fonction d'une distance $dist$), la proportion d'exemples de la même classe n'est pas significativement plus grande que dans le reste de l'ensemble de données. Le voisinage est défini par un graphe *Relative neighborhood graph* (RNG) où chaque noeud correspond à un exemple d'apprentissage et les arêtes représentent les proximités entre chaque exemple. Plus précisément, une arête relie les exemples \mathbf{x}^1 et \mathbf{x}^2 si $d(\mathbf{x}^1, \mathbf{x}^2) \leq \max(d(\mathbf{x}^1, \mathbf{x}^3), d(\mathbf{x}^2, \mathbf{x}^3))$ pour tout \mathbf{x}^3 différent de \mathbf{x}^1 et \mathbf{x}^2 . La Fig. 2.8 montre un exemple d'un tel graphe en deux dimensions.

Pour déterminer si un exemple \mathbf{x}^i labellisé θ de la base d'apprentissage est "suspect", les auteurs se basent sur le test d'hypothèse en considérant l'hypothèse H_0 suivante selon laquelle \mathbf{x}^i est un bon exemple :

La proportion d'exemples dans le voisinage de \mathbf{x}^i dont la classe est différente de θ est significativement inférieure à la proportion globale des exemples dont la classe est différente de θ dans toute la base d'apprentissage.

Les auteurs considèrent ensuite trois régions possibles pour la p -value associée à ce test : si cette dernière est inférieure à un niveau α_1 , alors l'exemple est considéré comme un "bon" exemple, si la p -value est supérieure à un niveau α_2 , l'exemple est considéré comme "mauvais". Enfin si la p -value est entre α_1 et α_2 l'exemple est considéré comme "suspect". Les exemples considérés comme "suspects" ou "mauvais" sont supprimés de la base d'apprentissage ou relabellisés. Supposons que l'exemple \mathbf{x}^i doive être relabellisé, la nouvelle classe assignée à \mathbf{x}^i est la classe θ_{maj} majoritaire des "bons" exemples de son voisinage.

2.2.4.3 Extraction de caractéristiques

Considérons ici que $\mathcal{X} \subseteq \mathbb{R}^d$. La réduction de dimension de \mathcal{X} , c'est-à-dire en travaillant avec un espace de caractéristiques réduit $\mathcal{X}' \subseteq \mathbb{R}^{d'}$ avec $d' < d$, facilite l'apprentissage et améliore les performances de la classification. Voici brièvement quelques exemples d'extractions usuelles :

- Analyse en Composantes Principales (PCA) (Hotelling, 1933) est une méthode d'exploration de données multivariées. L'objectif est de construire un petit nombre de nouvelles variables préservant au maximum la quantité de variation des données. Les nouvelles variables sont appelées *composantes principales* et forment une base orthonormée.
- Analyse en Composantes Indépendantes (ICA) (Hérault et Ans, 1984; Hérault et al., 1985) L'objectif de ICA est de trouver des composantes mutuellement indépendantes.
- Analyse Discriminante Linéaire (LDA) (Fisher, 1936) : cette extraction de caractéristiques est supervisée (basée sur un ensemble d'apprentissage). Pour un problème de classification

5. https://commons.wikimedia.org/wiki/File:Relative_neighborhood_graph.svg

à n classes, LDA réduit le nombre de caractéristiques à $n - 1$ en recherchant une combinaison linéaire des variables maximisant la variance intergroupe et minimisant la variance intragroupe.

- Analyse fonctionnelle en Composantes Principales (F-PCA) (Rao, 1958; Tucker, 1958) : F-PCA est une adaptation de la méthode PCA dans le cas de données fonctionnelles (la structure sous-jacente des données est supposée continue).

2.3 Gestion d'incertitudes en classification automatique supervisée

Dans la précédente section nous avons introduit la classification automatique supervisée dans le cas général. Dans cette section nous nous intéressons aux problèmes de classification en présence d'imprécision et/ou d'incertitude. Lors de la classification d'un exemple \mathbf{x} , il s'agit des sources d'imprécision et d'incertitude possibles qui nous empêchent de répondre avec précision et certitude à la question "quelle est la classe de \mathbf{x} ". Ces sources d'imprécisions et d'incertitudes sont variées (manque d'exemples disponibles, de caractéristiques discriminantes, de fiabilité des mesures ...), mais se manifestent principalement sous deux formes dans un problème de classification :

- Zone de *chevauchement* de classes dans l'espace des caractéristiques, où des exemples de classes différentes partagent des caractéristiques très proches.
- Zone de l'espace des caractéristiques avec peu d'exemples.

D'autres situations pour lesquelles il est difficile de répondre à la question de la classe d'un exemple existent. Par exemple, (Quost et al., 2020b) considère le cas d'exemple *non-conformes*, qui ne ressemblent à aucun exemple de la base d'apprentissage et propose des pistes de traitement évidentiel de ces exemples.

Remarque 2. Certains travaux considèrent l'hypothèse du monde ouvert où l'ensemble des classes possibles n'est pas représenté dans l'ensemble des exemples d'apprentissage (Quost et al., 2020b; Dubuisson et Masson, 1993). Ainsi un nouvel exemple pourrait appartenir à une classe inconnue. Dans cette thèse cette situation n'a pas été envisagée.

Les deux principales manifestations d'imprécision et/ou d'incertitude ont été interprétées dans la littérature. Pour (Hüllermeier et Brinker, 2008), les zones de chevauchement sont appelées *régions de conflit* et correspondent à de l'indifférence entre plusieurs classes, tandis que les zones avec peu d'exemples sont appelées *régions d'ignorance* et correspondent à des situations d'incomparabilité entre les classes. Pour (Quost et al., 2020b), les zones de chevauchement appelées *zones de mélange* correspondent à de l'incertitude aléatoire ou à de l'ambiguïté, et les zones avec peu d'exemples correspondent à de l'incertitude épistémique (manque de données). Afin d'illustrer les problèmes causés par la présence d'incertitude et d'imprécision dans un problème de classification, considérons le problème fictif de classification visant à discriminer les classes θ_1 et θ_2 sur la base de deux caractéristiques, présenté dans la Fig. 2.9. Dans cette figure, les rectangles représentent des régions qu'un classifieur pourrait construire pour caractériser une classe. Considérons la classification de trois nouveaux exemples \mathbf{x}^1 , \mathbf{x}^2 et \mathbf{x}^3 . Le nouvel exemple \mathbf{x}^2 se trouve dans une région clairement attribuable à la classe θ_2 , il y a donc peu de risque de commettre une erreur de classification avec cet exemple. En revanche, la classification est

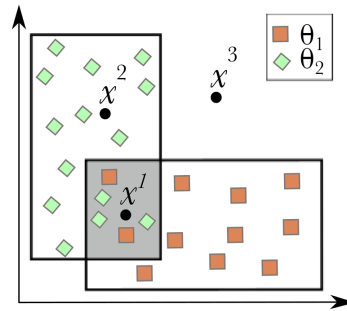


FIG. 2.9 Répartition des exemples d'apprentissage et trois nouveaux exemples x^1 , x^2 et x^3 , dans l'espace des deux caractéristiques. (Figure inspirée de (Hüllermeier et Brinker, 2008))

plus difficile pour l'exemple x^1 qui se trouve dans une zone de chevauchement des classes θ_1 et θ_2 , et pour l'exemple x^3 , qui se trouve dans une zone avec peu d'exemples. En effet, assigner des classes précises à x^1 et x^3 n'est pas justifiable : nous évitons les erreurs avec de la chance, et nous commettons des erreurs avec de la malchance. Dans la pratique, lorsque ces exemples difficiles à classer sont nombreux, nous commettrons des erreurs inévitables.

Ces erreurs inévitables constituent une limitation de la classification telle que présentée dans la section précédente. Dans la suite du manuscrit nous nous référons aux classifieurs assignant systématiquement une classe précise à un nouvel exemple comme des *classifieurs standard*.

Face à cette limitation, certains travaux ont proposé des classifieurs basés sur des théories de l'incertain, capables d'exprimer les doutes/hésitations du classifieur. Dans la section précédente, certains classifieurs étaient déjà construits dans une théorie de l'incertain, la théorie des probabilités, mais les informations probabilistes étaient perdues puisqu'une seule classe était finalement retenue. Les classifieurs que nous présentons dans cette section sont non seulement basés sur des cadres de l'incertain, mais fournissent l'information concernant l'incertitude. Cela permet, notamment, de combiner les sorties de plusieurs classifieurs, comme c'est souvent le cas avec l'apprentissage ensembliste ou le passage des résultats d'un classifieur binaire à un classifieur multi-classes. D'autre part, lorsque certaines erreurs sont plus graves que d'autres, nous pouvons prendre une décision par minimisation de risque en affectant des pertes plus ou moins importantes selon le type d'erreur grâce aux outils de la théorie de la décision. Ces informations constituent une connaissance plus riche et permettent au classifieur de s'abstenir de fournir une classe précise lorsque l'incertitude est trop importante.

Dans cette section, trois types de classifieurs fournissant ce type d'informations plus riches concernant la présence d'incertitude sont présentés. Il s'agit de classifieurs construits dans les cadres probabiliste, évidentiel et flou.

2.3.1 Classifieur probabiliste

Lorsqu'un classifieur δ_Θ est capable de fournir pour un nouvel exemple x , une distribution de probabilité a posteriori $p(\cdot|x) : \Theta \rightarrow [0, 1]$, nous l'appelons *classifieur probabiliste*. Nous conservons la notation δ_Θ comme pour les classifieurs standards car l'information fournie reste relative aux classes de Θ . Certains algorithmes de classification basés sur des modèles probabilistes comme le classifieur Naive Bayes (NBC) ou le classifieur par régression logistique (LR) produisent directement des distributions de probabilités. Pour les autres classifieurs, l'information fournie par le classifieur peut être transformée en une probabilité a posteriori. Cette transfor-

mation est connue sous le nom de *calibration* (Gebel, 2009). Les approches de calibration les plus largement acceptées utilisent la régression logistique (Platt et al., 1999) ou la régression isotonique (Zadrozny et Elkan, 2002).

Classifieur probabiliste et théorie de la décision Il est possible de retrouver un classifieur standard à partir d'un classifieur probabiliste. L'usage le plus fréquent est de se baser uniquement sur l'information $p(\cdot|\mathbf{x})$ pour associer une classe $\delta_\Theta(\mathbf{x})$ à un nouvel exemple \mathbf{x} en choisissant la classe ayant la probabilité maximale comme suit :

$$\delta_\Theta(\mathbf{x}) = \arg \max_{\theta \in \Theta} p(\theta|\mathbf{x}).$$

Pour une décision plus sophistiquée, nous pouvons faire appel aux outils de la théorie de la décision. Soit $\mathcal{A} = \{a_1, a_2, \dots, a_m\}$ un ensemble fini d'actions possibles. Dans le cas de la classification d'un objet \mathbf{x} , une action $a \in \mathcal{A}$ correspond à l'attribution d'une classe $\delta_\Theta(\mathbf{x}) \in \Theta$ à \mathbf{x} . Nous pouvons simplifier en posant $\mathcal{A} = \Theta$. Afin de jauger les actions de \mathcal{A} pour la décision, deux fonctions sont introduites :

- La *fonction de perte* $L : \mathcal{A} \times \Theta \rightarrow \mathbb{R}$ quantifie la perte $L(\delta_\Theta(\mathbf{x}), \theta^*)$ si on choisit l'action $a = \delta_\Theta(\mathbf{x}) \in \mathcal{A}$ alors que la vraie classe de \mathbf{x} est $\theta^* \in \Theta$. De façon analogue, nous pouvons utiliser une fonction de gain $G : \mathcal{A} \times \Theta \rightarrow \mathbb{R}$ quantifiant le gain à la place de la perte.
- La *fonction de risque* $r_{\delta_\Theta} : \mathcal{A} \rightarrow \mathbb{R}$ est définie comme l'espérance suivante :

$$r_{\delta_\Theta}(a) = E_{p(\cdot|\mathbf{x})}(L(a, \cdot)).$$

Lorsque la distribution de probabilité est discrète et finie, nous obtenons :

$$r_{\delta_\Theta}(a) = \sum_{\theta \in \Theta} L(a, \theta) p(\theta|\mathbf{x}) \tag{16}$$

Si nous choisissons d'utiliser une fonction de gain (au lieu de la fonction de perte) dans les définitions précédentes, nous parlons d'*espérance de gain* (au lieu de fonction de risque).

Le classifieur prend ensuite la décision de la classe à associer à l'exemple \mathbf{x} en choisissant la classe minimisant le risque :

$$\delta_\Theta(\mathbf{x}) = \arg \min_{a \in \mathcal{A}} r_{\delta_\Theta}(a).$$

De manière similaire, si nous utilisons une fonction de gain (au lieu d'une fonction de perte), nous choisirions la classe maximisant l'espérance de gain. Notons que la règle de décision basée uniquement sur $p(\cdot|\mathbf{x})$ peut être retrouvée avec la fonction de perte L définie telle que $L(\theta_j, \theta_i) = 1$, si $i \neq j$, et $L(\theta_j, \theta_j) = 0$.

2.3.2 Classifieur évidentiel

Lorsqu'un classifieur fournit à un nouvel exemple \mathbf{x} une fonction de masse $m(\cdot|\mathbf{x}) : 2^\Theta \rightarrow [0, 1]$, nous le désignons par *classifieur évidentiel*. Notons qu'un classifieur probabiliste tel que défini dans la précédente section est une catégorie de classifieur évidentiel car une probabilité est un type de fonction de masse (fonction de masse bayésienne).

Commençons par introduire l'un des classifieurs évidentiels les plus connus, le *k-nn évidentiel*.

Méthode évidentielle des k plus proches voisins Dans la version évidentielle du classifieur k -nn (Denoeux, 1995), une sélection d'exemples d'apprentissage constituée des voisins de \mathbf{x} est considérée comme source d'information sur la vraie classe de x . L'information provenant de chaque source (chaque voisin) est représentée par une fonction de masse tenant compte de la distance entre \mathbf{x} et chaque voisin. Plus un voisin est loin de \mathbf{x} , moins il nous informe sur la classe de \mathbf{x} . L'apport d'évidence d'un voisin de vecteur de caractéristique \mathbf{x}^v et de classe θ^v (où $\theta^v = \theta_i \in \Theta$, i l'indice de la classe du voisin dans l'ensemble Θ) est représenté sous la forme d'une fonction de masse m_v :

$$\forall A \subseteq \Theta, \quad m_v(A|\mathbf{x}) = \begin{cases} \alpha \phi_i(d(\mathbf{x}, \mathbf{x}^v)) & \text{si } A = \{\theta_i\}, \\ 1 - \alpha \phi_i(d(\mathbf{x}, \mathbf{x}^v)) & \text{si } A = \Theta, \\ 0 & \text{si } A \in 2^\Theta \setminus \{\Theta, \{\theta_i\}\}. \end{cases}$$

où $d(\mathbf{x}, \mathbf{x}^v)$ est la distance entre \mathbf{x} et son voisin \mathbf{x}^v , α un paramètre tel que $0 < \alpha < 1$ et ϕ_i une fonction décroissante vérifiant $\phi_i(0) = 1$ et $\lim_{z \rightarrow \infty} \phi_i(z) = 0$. Considérons maintenant k voisins (\mathbf{x}^v, θ^v) , $1 \leq v \leq k$ de \mathbf{x} . Les apports d'évidence provenant de chaque voisin sont ensuite combinés en utilisant la combinaison de Dempster, pour construire une seule fonction de masse $m(\cdot|\mathbf{x})$ distribuant les chances sur 2^Θ pour la vraie classe de \mathbf{x} :

$$m(\cdot|\mathbf{x}) = m_1(\cdot|\mathbf{x}) \oplus \dots \oplus m_k(\cdot|\mathbf{x}).$$

Une adaptation de ce classifieur en utilisant les réseaux de neurones a été proposée dans (Denoeux, 2000). Notons que ces versions évidentielles de k -nn et réseaux de neurones considèrent les éléments focaux seulement sur les singletons et l'ignorance totale. Ces dernières approches font partie de la famille de classifieurs évidentiels qualifiés de "distance-based" dans (Vanmoorenberghe et Denoeux, 2001) par opposition aux approches basées sur une vraisemblance. D'autres versions évidentielles de classifieurs basés ou non sur des distances ont également été proposées mais ne peuvent pas se généraliser à de nombreux classifieurs existants. Par exemple des versions évidentielles du classifieur par arbre de décision (Sutton-Charani et al., 2012, 2013) ou d'autres versions du k -nn (Jiao et al., 2013; Lian et al., 2015) ont été proposées.

L'autre famille de méthodes utilisées dans de nombreuses applications pour assigner de la masse est basée sur le principe de vraisemblance. Dans ce cas, les densités de probabilité conditionnées à chaque classe, $f(\mathbf{x}|\theta)$, sont supposées connues. Pour l'exemple observé \mathbf{x} , ces méthodes se basent donc sur la vraisemblance définie comme $L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta)$ pour $\theta \in \Theta$. Deux façons d'assigner de la masse à partir de cette vraisemblance existent :

La méthode consonante Cette méthode introduite dans (Shafer, 1976), définit une fonction de plausibilité sur Θ à partir de la vraisemblance pour un sous-ensemble A de Θ :

$$Pl(A|\mathbf{x}) = \frac{\max_{\theta \in A} L(\theta|\mathbf{x})}{\max_{\theta} L(\theta|\mathbf{x})}.$$

Cette méthode de construction de fonction de masse a posteriori est appelée consonante car la fonction de masse obtenue est consonante.

La méthode séparable Proposée dans (Appriou, 1998), cette méthode prend en compte séparément l'apport d'évidence en faveur de chaque classe θ_i pour la classe de \mathbf{x} . Cet apport s'exprime sous la forme d'une fonction de masse :

$$\forall A \subseteq \Theta, \quad m(A|\mathbf{x}) = \begin{cases} \alpha(1 - RL(\theta_i|\mathbf{x})) & \text{si } A = \overline{\{\theta_i\}}, \\ 1 - \alpha(1 - RL(\theta_i|\mathbf{x})) & \text{si } A = \Theta, \\ 0 & \text{si } A \in 2^\Theta \setminus \{\Theta, \overline{\{\theta_i\}}\}. \end{cases}$$

où $\overline{\{\theta_i\}}$ désigne le cardinal de $\{\theta_i\}$, R est un coefficient de normalisation et α est un coefficient d'affaiblissement associé à la classe θ . Les n fonctions de masse provenant de chaque classe sont ensuite combinées avec la combinaison de Dempster.

Classifieur évidentiel et théorie de la décision Nous pouvons regrouper les règles de décision basées sur les fonctions de croyance en deux catégories principales : celles qui induisent un ordre de préférence⁶ complet sur les classes et celles qui induisent un ordre de préférence partiel sur les classes (Denoeux, 2019). Nous nous intéressons pour le moment aux méthodes fournissant un ordre complet sur les classes pour lequel le choix final se porte naturellement sur la classe préférée dans l'ordre complet. Dans le cas d'un ordre partiel entre les classes, le choix final est généralement l'ensemble des classes non-dominées dans l'ordre partiel. Nous reviendrons sur ce dernier cas dans le chapitre suivant qui traitera principalement de sorties sous la forme d'ensembles de classes candidates.

Comme présenté dans (Denoeux, 2019) concernant les approches induisant un ordre complet, la majorité des critères de décision consistent à choisir l'action $a \in \mathcal{A}$ qui minimise la somme

$$\sum_{B \subseteq \Theta} m(B|\mathbf{x}) \mathcal{L}(a, B),$$

où la quantité $\mathcal{L}(a, B)$ est exprimée en fonction d'une fonction de perte $L : \mathcal{A} \times \Theta \rightarrow \mathbb{R}$. Cette quantité est différente suivant le critère de décision :

- Avec la décision basée sur la probabilité pignistique associée à m présentée dans la sous-section 2.1.3, nous avons :

$$\mathcal{L}(a, B) = \frac{1}{|B|} \sum_{\theta \in B} L(a, \theta).$$

Remarquons que nous retrouvons ici la décision basée sur la minimisation du risque r_{δ_Θ} (équation (16)) associé à la pignistique $betP_m$ (équation (5)) :

$$\begin{aligned} \sum_{B \subseteq \Theta} m(B|\mathbf{x}) \mathcal{L}(a, B) &= \sum_{B \subseteq \Theta} m(B|\mathbf{x}) \frac{1}{|B|} \sum_{\theta \in B} L(a, \theta) \\ &= \sum_{\theta \in \Theta} L(a, \theta) \sum_{B \subseteq \Theta, \theta \in B} \frac{m(B)}{|B|} \\ &= \sum_{\theta \in \Theta} L(a, \theta) betP_m(\theta) \\ &= r_{\delta_\Theta}(a) \end{aligned}$$

- Avec la décision que nous appelons *décision optimiste*, nous avons

$$\mathcal{L}(a, B) = \min_{\theta \in B} L(a, \theta).$$

Cette décision correspond à la minimisation de l'espérance définie par l'intégrale de Choquet associée à Bel (Gilboa et Schmeidler, 1994).

6. La préférence dont il question ici est à comprendre comme suit : pour un objet \mathbf{x} , on dit qu'on préfère la classe θ_1 à la classe θ_2 si le classifieur considère que \mathbf{x} a plus de chance d'être θ_1 que d'être θ_2 .

- De même, nous appelons *décision pessimiste*, la décision correspondant à la minimisation de l'espérance définie par l'intégrale de Choquet associée à Pl :

$$\mathcal{L}(a, B) = \max_{\theta \in B} L(a, \theta).$$

- Avec la généralisation du critère de Hurwicz (Jaffray, 1988) :

$$\mathcal{L}(a, B) = \alpha \min_{\theta \in B} L(a, \theta) + (1 - \alpha) \max_{\theta \in B} L(a, \theta),$$

où $\alpha \in [0, 1]$ est un degré de pessimisme.

- Avec la généralisation du critère OWA (Ordered weighted average) (Yager, 1988) :

$$\mathcal{L}(a, B) = \text{F-Score}_{|B|, \beta}(\{L(a, \theta) | \theta \in B\}),$$

où $\text{F-Score}_{|B|, \beta}$ est l'opérateur de maximum de l'entropie de OWA avec un degré de pessimisme β .

2.3.3 Classifieur flou

Nous appelons *classifieur flou*, tout classifieur δ_{Θ} qui pour un nouvel exemple \mathbf{x} est capable de déduire un vecteur d'appartenance à chacune des classes considérées, noté $(\mu_{\theta_1}(\mathbf{x}), \dots, \mu_{\theta_n}(\mathbf{x}))$. Notons que certains travaux, utilisent le terme "classifieur flou" de façon plus générale dès qu'un classifieur utilise les ensembles flous lors de son apprentissage ou de son application (Kuncheva, 2000).

Nous présentons, dans la suite, deux exemples de classifieurs flous. Le premier est le k -nn flou et le second est le type de classifieurs frbcs.

Le classifieur k -nn flou La méthode floue des k plus proches voisins est une adaptation basée sur les sous-ensembles flous du classifieur k -nn standard. Soit un exemple \mathbf{x} à classifier. Considérons que nous disposons d'une base d'exemples d'apprentissage sous la forme suivante : $(\mathbf{x}^l, (\mu_{\theta_1}(\mathbf{x}^l), \dots, \mu_{\theta_n}(\mathbf{x}^l)))_{1 \leq l \leq L}$, où $\mu_{\theta_i}(\mathbf{x}^l)$ désigne le degré d'appartenance de l'exemple d'apprentissage \mathbf{x}^l à la classe θ_i . En sortie, le k -nn flou fournit un vecteur de degrés d'appartenance $(\mu_{\theta_1}(\mathbf{x}), \dots, \mu_{\theta_n}(\mathbf{x}))$.

La première version du k -nn flou a été proposée dans (Keller et al., 1985). Soit $\theta \in \Theta$,

$$\mu_{\theta}(\mathbf{x}) = \frac{\sum_{t=1}^k \mu_{\theta}(\mathbf{x}^t) \frac{1}{\|\mathbf{x}^t - \mathbf{x}\|^{2/(m-1)}}}{\sum_{t=1}^k \frac{1}{\|\mathbf{x}^t - \mathbf{x}\|^{2/(m-1)}}},$$

où m est un paramètre déterminant la manière dont la distance avec les voisins est prise en compte dans la contribution pour le calcul du degré d'appartenance. Le k -nn flou donne une information plus riche que le k -nn standard en intégrant les degrés d'appartenance des voisins ainsi que leurs distances avec \mathbf{x} .

Notons que d'autres classifieurs standards ont des versions floues, par exemple, les arbres de décision (Yuan et Shaw, 1995) ou les réseaux de neurones (Buckley et Hayashi, 1994), etc.

Les classifieurs frbcs (Fuzzy Rule-Based Classification Systems) Il s'agit des classifieurs basés sur les règles floues. Ces classifieurs sont constitués de trois grandes étapes :

1. Défuzzification de chaque espace des variables d'entrées ;
2. Construction des règles floues en mettant en lien les termes déduits de la défuzzification et les classes ;
3. Construction de la règle de classification pour un nouvel exemple \mathbf{x} à partir des règles qui lui correspondent (voir (Riza et al., 2015) pour une description détaillée de ces classifieurs et leurs implémentions en R).

Finalement, le vecteur des degrés d'appartenance $(\mu_{\theta_1}(x), \dots, \mu_{\theta_n}(x))$ associé à \mathbf{x} dans le cas du classifieur frbcs est calculé de la façon suivante : $\mu_{\theta}(x) = 0$ si aucune des règles activées pour \mathbf{x} ne contient une conséquence θ . Sinon $\mu_{\theta}(x)$ est le degré maximal de correspondance des règles dont la conséquence est θ .

Lorsque les règles floues ne sont pas données par un expert, certains algorithmes permettent de les générer par apprentissage automatique (Nauck et al., 1997; Roubos et al., 2003; Ishibuchi et al., 1995).

2.4 Illustration

Dans cette section nous illustrons les performances de certains des algorithmes de classification présentés dans le chapitre. Les résultats sont donnés pour divers problèmes de classification dont les jeux de données sont présentés en Annexe B. Dans un second temps les sorties de classifieurs basés sur des théories de l'incertain sont illustrées sur un exemple de classification de pixels de fragments de plastiques.

2.4.1 Evaluation de classifieurs sur des jeux de données

Une sélection de classifieurs a été évaluée sur un ensemble de jeux de données. Les expérimentations ont été menées sur R (voir annexe B pour les détails de la machine et les principales routines R utilisées). Afin d'évaluer les classifieurs pour des situations les plus variées, nous avons choisi des jeux de données dont les caractéristiques (nombres d'attributs, de classes, d'exemples) sont très variables (voir annexe A). Les résultats des taux de justesse *acc* moyens obtenus sur 10 répétitions de validations croisées à 10 blocs sont présentés dans la TABLE 2.3. Les meilleurs scores (en gras) ne sont pas toujours obtenus avec les mêmes classifieurs suivant les jeux de données. De plus les performances de classification doivent être mises en perspective avec le contexte du problème. En effet dans certains cas une performance faible peut être satisfaisante si l'enjeu de la classification n'est pas important. Concentrons-nous sur la ligne du jeu de données "plastics" en jaune dans la TABLE 2.3. Le meilleur score, obtenu par le classifieur forêt aléatoire (RF), est 0.91. Ce score reste insuffisant pour assurer la possibilité de recycler les matériaux impliqués dans le problème de classification.

Un autre critère important dans le choix d'un classifieur est le temps de calcul. La table 2.4 présente les temps de calcul moyen en secondes sur les 10 répétitions des validations croisées et sur l'ensemble des jeux de données.

Nous remarquons que certains classifieurs ont une phase d'apprentissage assez longue (ici ANN) mais que la phase de test est beaucoup plus rapide. A l'inverse d'autres classifieurs comme les k -nns, qui n'ont pas de phase d'apprentissage (hormis une phase d'optimisation de paramètres pour le k -nn évidentiel), sont plus chronophages dans la phase d'application, notamment pour la recherche de voisins.

	NBC	DT	RF	SVM	ANN	RL	k -nn	k -nn évidentiel	k -nn flou
ecoli	0.82	0.77	0.85	0.76	0.84	0.84	0.85	0.82	0.84
forestTypes	0.95	0.93	0.95	0.93	0.93	0.90	0.92	0.94	0.93
glass	0.43	0.68	0.73	0.66	0.65	0.57	0.64	0.67	0.67
iris	0.96	0.95	0.95	0.95	0.95	0.94	0.94	0.96	0.89
optdigit	0.82	0.75	0.98	0.46	0.71	0.94	0.98	0.98	0.98
plastics	0.48	0.72	0.91	0.90	0.89	0.88	0.89	0.89	0.69
seeds	0.89	0.91	0.90	0.91	0.92	0.94	0.89	0.88	0.90
segmentation	0.76	0.85	0.90	0.14	0.44	0.83	0.78	0.68	0.80
vehicle	0.47	0.67	0.75	0.75	0.58	0.79	0.59	0.63	0.63
vowel	0.56	0.51	0.90	0.86	0.59	0.58	0.59	0.91	0.88
wallFollowing	0.53	0.98	0.99		0.79	0.70	0.83	0.85	0.87
wine	0.98	0.93	0.98	0.98	0.92	0.93	0.70	0.68	0.70
wineQuality	0.53	0.57	0.65	0.61	0.59	0.59	0.48	0.55	0.39
yeast	0.34	0.57	0.62	0.59	0.59	0.58	0.55	0.55	0.36

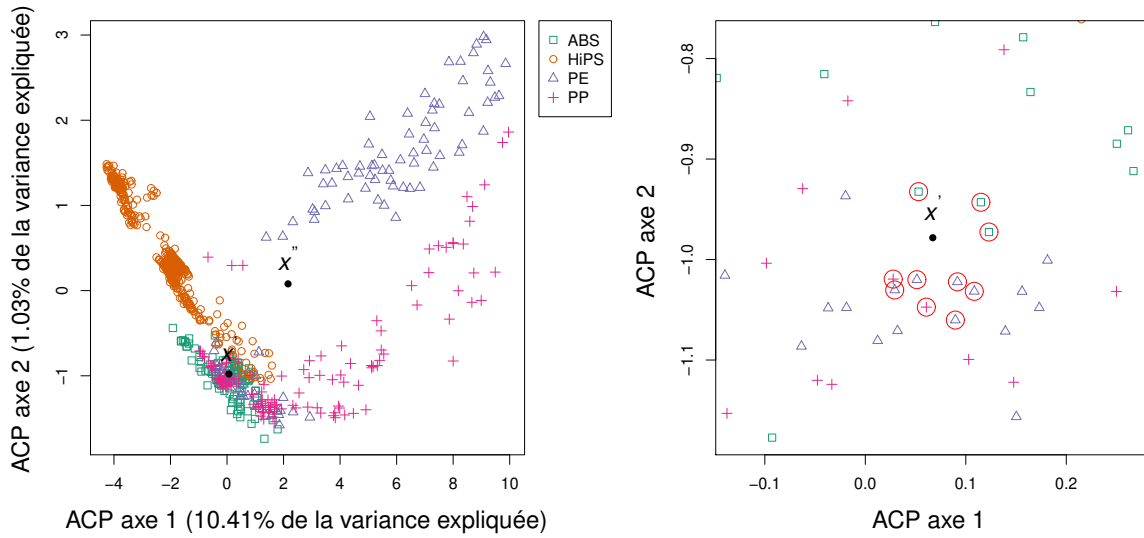
 TABLE 2.3 – acc moyens obtenus sur 10 répétitions de validations croisées à 10 blocs.

	Temps d'apprentissage	Temps d'application
NBC	0.04	4.27
DT	1.15	0.02
RF	14.67	0.44
SVM	20.15	2.37
ANN	85.81	0.03
RL	1.15	0.02
k -nn		9.30
k -nn évidentiel	7.49	8.15
k -nn flou		115.01

TABLE 2.4 – Temps moyens en secondes pour la classification d'un bloc de validation croisée (moyenne sur tous les jeux de données et les 10 répétitions des validations croisées)

2.4.2 Illustration de sorties probabilistes, évidentielles et floues

Dans cette sous-section, nous présentons des sorties standards, probabilistes, évidentielles et floues. Au total 717 spectres (appartenant à 280 fragments de plastiques différents) ont été utilisés pour l'apprentissage. Ces exemples sont illustrés en fonction des deux premières composantes de l'ACP Fig. 2.10a. Pour simplifier l'illustration, nous considérons les valeurs des deux composantes de l'ACP comme les seules caractéristiques dans le problème de classification. Nous nous intéressons aux prédictions de deux nouveaux exemples \mathbf{x}' et \mathbf{x}'' également représentés sur la figure. Le premier exemple \mathbf{x}' est situé dans une zone de chevauchement des classes alors que le second se trouve dans une zone avec peu d'exemples d'apprentissage. Pour mieux cerner l'exemple \mathbf{x}' dans la zone de chevauchement, la Fig 2.10b est un zoom sur cette zone. Les classifieurs utilisés sont tous des versions du classifieur k -nn avec $k = 10$. Les versions standard, évidentielle et floue ($m = 2$) sont celles présentées dans la section précédente. La version probabiliste estime ses probabilités par les pourcentages de vote de chaque voisin. Les résultats sont donnés dans les tables 2.5 et 2.6.



(a) Extrait de la base d'apprentissage (base de données plastics) en fonction des deux premières composantes de l'ACP.

(b) Zoom sur l'exemple \mathbf{x}' . Ses 10 plus proches sont entourés en rouge.

k -nn standard	k -nn probabiliste	k -nn évidentiel	k -nn flou
PE	$p(ABS \mathbf{x}') = 0.3$ $p(PE \mathbf{x}') = 0.5$ $p(PP \mathbf{x}') = 0.2$	$m(\{ABS\} \mathbf{x}') = 0.002$ $m(\{PE\} \mathbf{x}') = 0.998$	$\mu_{ABS}(\mathbf{x}') = 0.339$ $\mu_{PE}(\mathbf{x}') = 0.491$ $\mu_{PP}(\mathbf{x}') = 0.169$

TABLE 2.5 – Prédictions obtenues pour l'exemple de test \mathbf{x}' (zone de chevauchement des classes)

k -nn standard	k -nn probabiliste	k -nn évidentiel	k -nn flou
PE	$p(HIPS \mathbf{x}'') = 0.3$ $p(PE \mathbf{x}'') = 0.4$ $p(PP \mathbf{x}'') = 0.3$	$m(\{HIPS\} \mathbf{x}'') = 0.001$ $m(\{PE\} \mathbf{x}'') = 0.316$ $m(\Theta \mathbf{x}'') = 0.683$	$\mu_{HIPS}(\mathbf{x}'') = 0.218$ $\mu_{PE}(\mathbf{x}'') = 0.578$ $\mu_{PP}(\mathbf{x}'') = 0.204$

TABLE 2.6 – Prédictions obtenues pour l'exemple de test \mathbf{x}'' (zone avec peu d'exemples)

Les quatre classifieurs nous fournissent des informations concernant les classes de \mathbf{x}' et \mathbf{x}'' . Dans tous les cas c'est la classe PE qui obtient les plus grands degrés de certitude. Cependant les sorties s'interprètent différemment. Par exemple, pour \mathbf{x}' , le classifieur standard prédit la classe PE sans plus de détails. Le classifieur probabiliste estime que la probabilité que \mathbf{x}' soit de la classe PE est 0.5. Similairement le classifieur évidentiel donne une masse de 0.998 pour que PE soit la classe de \mathbf{x}' . Le classifieur flou lui estime que le degré d'appartenance de \mathbf{x}' à la classe PE est 0.491.

Notons que la masse assignée par le classifieur évidentiel à PE pour \mathbf{x}' est importante (0.998). En effet, l'exemple \mathbf{x}' se trouve dans une zone de chevauchement, et est donc très proche de ses voisins. Ainsi, les apports d'évidence de chaque voisin sont représentés par des fonctions de masse avec une forte masse sur le singleton correspondant à leur classe. Ensuite ces

fonctions de masse sont combinées avec la combinaison de Dempster qui a tendance à resserrer la masse sur l'intersection (voir l'exemple 10 de ce chapitre pour un cas similaire). En revanche pour l'exemple \mathbf{x}'' , les distances avec les voisins amoindrissent les masses sur les classes des voisins et favorisent l'ignorance (finalement $m(\Theta|\mathbf{x}'') = 0.683$).

2.5 Conclusion du chapitre

Afin de répondre à notre problématique de tri de plastiques sous contraintes à partir de données incertaines, nous nous sommes intéressés aux cadres existants permettant de manipuler des données incertaines. Ces cadres ont été introduits avec la présentation de cinq théories de l'incertain : la théorie des probabilités, les probabilités imprécises, les fonctions de croyances, les possibilités et la logique floue. Chacune de ces théories propose des représentations de l'information, probabilité, ensemble crédal, fonction de masse, possibilité, degré d'appartenance qui sont capables de tenir compte de différents types d'incertitudes.

Nous avons également introduit la classification supervisée et présenté les classifieurs standards usuels. Dans l'illustration, nous avons constaté que les classifieurs étaient plus ou moins performants en fonction du problème de classification considéré. Outre les performances en terme de qualité des sorties, nous avons noté l'importance de la question de la complexité et des besoins de stockage dans le choix d'un classifieur.

Enfin nous avons identifié la limite des classifieurs standards lorsque les données sont entachées d'imperfections : ils associent des classes précises même lorsque l'information est insuffisante pour décider. Cette attitude entraîne des erreurs inévitables et peut empêcher d'atteindre les niveaux d'exigence requis pour certaines applications. En particulier les performances des classifieurs standards ne sont pas suffisantes dans le problème du tri de plastiques. Au vu de ce constat, nous avons introduit des classifieurs alternatifs qui remettent en question la forme de sortie classique en proposant des sorties exprimant les doutes/hésitations du classifieur. Nous avons présenté trois types de ces classifieurs alternatifs s'appuyant chacun sur une théorie de l'incertain différente : les classifieurs probabilités, les classifieurs évidentiels et les classifieurs flous. Les informations fournies par ces classifieurs peuvent être exploitées à de nombreuses fins telles que la combinaison de plusieurs sorties ou la prise en compte de fonctions de pertes différentes en fonction du type d'erreur. Pour l'application au tri des plastiques, nous proposons d'exploiter ce type d'informations plus riches pour améliorer la fusion des informations provenant des différents pixels constituant un fragment. D'autre part, notons que notre objectif final n'est pas seulement la reconnaissance des plastiques, mais leur tri. Ainsi, nous proposons de tirer profit de ces informations en les intégrant à un processus sensible aux imperfections des données afin d'optimiser le tri sous contraintes.

Dans le prochain chapitre nous nous concentrons sur ces classifieurs capables de fournir des informations plus riches reflétant l'imperfection des données, que nous nommons *classifieurs prudents*. En particulier, nous proposons un état de l'art sur une famille de classifieurs prudents dont la forme de sortie est compatible avec le dispositif de tri. Ces classifieurs que nous appelons *classifieurs imprécis* consistent à fournir un ensemble de classes candidates le plus petit possible en fonction de l'information disponible. Cette forme de sortie se traduirait par la décision d'envoyer un fragment dans un bac de tri, moins sélectif, dédié à plusieurs plastiques, d'une même famille par exemple.

Sommaire

3.1	Principe	61
3.1.1	Définition et les attentes de la classification imprécise	61
3.1.2	Illustration de l'intérêt de la classification imprécise	65
3.1.3	Distinction avec la classification multi-labels	68
3.1.4	Classification imprécise versus la classification hiérarchique	69
3.1.5	Classification imprécise et option de rejet	70
3.2	Méthodes de classification imprécise	71
3.2.1	Classification imprécise basée sur la théorie des probabilités	71
3.2.2	Classification imprécise basée sur les probabilités imprécises	73
3.2.3	Classification imprécise basée sur les fonctions de croyance	74
3.2.4	Classification imprécise basée sur la logique floue	75
3.2.5	Classification imprécise basée sur la théorie des possibilités	75
3.2.6	Autres méthodes de classification imprécise sans hypothèse de modèle	75
3.3	Évaluation d'un classifieur imprécis	75
3.3.1	Justesse affaiblie	77
3.3.2	Mesure d'évaluation F_β	78
3.3.3	Mesures d'évaluation g_{65} et g_{80}	80
3.3.4	Lien avec l'évaluation de classifieurs hiérarchiques	81
3.4	Positionnement	85
3.5	Conclusion du chapitre	87

Dans le chapitre précédent, nous avons présenté la classification standard qui consiste à fournir une classe précise pour un nouvel exemple à classifier. Nous avons également vu que les performances des classifieurs standards usuels ne suffisent pas toujours à atteindre les exigences selon l'application. En effet, la présence d'imperfections trop importantes rend parfois difficile le choix entre deux classes aux caractéristiques voisines ce qui complique la tâche de classification et en augmente les erreurs. Ces erreurs de classification peuvent avoir de lourdes conséquences économiques, environnementales, etc. Pour pallier ces problèmes, des méthodes de classification reposant sur des théories de l'incertain autres que les probabilités ont été introduites. Elles rendent possible en particulier de ne pas devoir trancher entre deux classes précises si l'information disponible sur un élément ne le permet pas de façon fiable, mais de ranger cet élément dans

une surclasse conformément moins informative. Ces méthodes offrent ainsi des représentations plus riches et adaptées à l'information imparfaite dont on dispose pour opérer la classification : la possibilité de choisir une classe non nécessairement précise permet d'améliorer les performances de la classification dès lors que le choix de ranger l'objet imparfaitement observé dans une classe non élémentaire, plus abstraite, fait sens. Cependant, lorsque ces méthodes sont employées pour une tâche de classification standard, où la sortie est nécessairement une classe élémentaire précise, elles n'expriment alors malheureusement pas leur potentiel à gérer l'abstraction dans le processus de la classification et se trouvent alors évaluées selon des modalités qui ne correspondent pas aux objectifs fixés à leur conception. En d'autres termes ces méthodes de classification sont contraintes à parier sur une classe précise, au risque de commettre une erreur.

Dans ce chapitre nous nous intéressons à l'exploitation de la représentation des incertitudes et des imprécisions sur les informations disponibles concernant un objet par des outils capables de fournir des sorties exprimant le doute/l'hésitation conformément aux informations imparfaites caractérisant l'objet. Nous faisons référence à ce type de sorties par l'expression *sorties prudentes* et nous appelons *méthode de classification prudente* ou *classifieur prudent* le type de classifieur associé. De nombreuses méthodes de classification de la littérature satisfont notre définition de la classification prudente. Nous en avons déjà rencontré dans le chapitre précédent. En effet, les classifieurs probabilistes (Duda et al., 1973), évidentiels (Denoëux, 1995, 2000) et flous (Keller et al., 1985; Riza et al., 2015) sont capables d'exprimer divers types d'incertitudes traduisant le doute et/ou l'hésitation du classifieur. Ainsi les sorties sous forme de distributions de probabilités (probability mass function (pmf) en anglais), de fonctions de masses (bba en anglais) ou de vecteurs d'appartenance à chacune des classes sont des catégories de sorties prudentes. D'autres formes de sorties prudentes ont été proposées, par exemple, certaines approches sont capables d'inférer pour un objet, des relations de préférences entre les classes qui ne sont pas nécessairement totales (ce qui permet d'exprimer de l'incomparabilité entre certaines classes) ou strictes (indifférence entre certaines classes) (Fürnkranz et Hüllermeier, 2010). Par exemple, (Hüllermeier et Brinker, 2008) propose un classifieur basé sur des techniques de décomposition par paires, capable de fournir pour un objet à classifier, une structure de relations floues sur l'ensemble des classes. Une sortie prudente peut également être exprimée sous la forme d'un ensemble de classes candidates. Ce type de sorties prudentes, est possible avec la catégorie de classification prudente appelée *classification imprécise* et qui fait l'objet de ce chapitre. Remarquons que certaines catégories de sorties prudentes peuvent se déduire d'autres catégories de sorties prudentes. Par exemple, une distributions de probabilités pmf peut être obtenue à partir d'une fonction de masses bba en passant par la probabilité pignistique. Par ailleurs, nous verrons que les sorties prudentes sous forme d'ordres partiels peuvent induire des sorties prudentes sous forme d'ensemble de classes candidates. Notons que les différentes formes de sortie prudentes ne sont pas toujours exploitables dans n'importe quel problème applicatif.

Dans l'application pour le tri de plastiques, la classification prudente permettra d'exploiter des informations incertaines et imprécises sur la nature des fragments de plastiques afin de diminuer l'impact sur les erreurs de tri. Le choix du type de sortie prudente est contraint par le processus de tri, i.e., la sortie prudente doit être intégrable à ce processus. Une façon simple d'autoriser des sorties prudentes est de considérer des bacs de tri dédiés non pas seulement à des plastiques particuliers, mais à des familles de plastiques. Ainsi, lorsque nous disposons d'une information imprécise sur la nature d'un fragment, par exemple, le fragment est soit un *ABS* ou un *HIPS*, nous ne serions pas obligés d'effectuer un choix risqué entre le bac dédié à l'*ABS* et celui dédié à l'*HIPS*. De façon plus prudente, lorsque l'information sur le fragment est trop imparfaite, le fragment pourrait être envoyé dans un bac dédié à l'ensemble des deux plastiques $\{ABS, HIPS\}$, tous deux appartenant à la famille des styréniques. Notre stratégie

de tri prudente sera basée sur ce type de sorties prudentes, i.e., sous forme de sous-ensembles de classes, avec pour objectif de maximiser le remplissage des bacs les mieux valorisables sur le marché du recyclage tout en respectant les seuils d'impuretés maximaux propres à chaque bac. Avant de nous intéresser à ce problème d'optimisation global, nous nous concentrons sur les méthodes de classification capables de renvoyer un ensemble de classes, sorties exploitables par la stratégie de tri. La *classification imprécise* est pour nous une étape intermédiaire dans la construction de notre approche proposée dans cette thèse. Nous la considérons comme la démarche la plus naïve pour résoudre le problème général.

Depuis les méthodes de classification avec *option de rejet* (Chow, 1957, 1970) basées sur les probabilités, permettant aux classifieurs de s'abstenir de renvoyer une classe précise lorsque le doute/l'hésitation est trop important, plusieurs techniques de classification ont été mises en place en s'appuyant sur les théories modernes de l'incertain afin de permettre des abstentions partielles (autorisant des sorties sous forme d'ensembles de classes candidates). L'idée est de représenter les types d'imperfections dans des cadres adaptés pour mieux sélectionner les classes candidates. Lors du choix d'un classifieur imprécis, le décideur devrait considérer plusieurs critères dont celui de la performance. La performance d'un classifieur imprécis ne peut pas être évaluée comme celle de la classification standard puisque certaines prédictions ne sont plus directement comparables avec les vraies classes. De plus l'appréciation de la prudence est subjective et dépend de l'application. Ce chapitre propose une introduction à la classification imprécise, un état de l'art des techniques existantes de classification imprécise et enfin un positionnement introduisant notre approche de classification imprécise.

L'organisation du chapitre est la suivante. Dans la section 3.1, le principe de la classification imprécise est présenté. La section 3.2 est consacrée aux principaux classifieurs imprécis existants. Des mesures d'évaluation de classifieurs imprécis sont présentées dans la section 3.3. Enfin, dans la section 3.4, nous donnons le positionnement de notre approche de classification imprécise (présentée au chapitre suivant).

3.1 Principe

La classification imprécise est une catégorie de classification prudente capable d'exprimer les doutes/hésitations du classifieur en fournissant des sorties sous la forme d'un ensemble de classes pour les exemples *difficiles*, i.e., les exemples pour lesquels les informations disponibles sont fortement affectées par des imperfections et rendent l'attribution à une classe précise trop incertaine.

Etant donné que la notion de classification imprécise, à l'instar de la classification prudente de façon plus générale, est une notion récente, nous proposons dans cette section de lui donner une définition précise pour notre travail et de la distinguer des autres notions qui lui sont proches comme la classification hiérarchique ou avec lesquelles il peut y avoir confusion comme la classification multi-labels. Enfin, nous verrons en quoi la classification imprécise est une généralisation de l'*option de rejet*.

3.1.1 Définition et les attentes de la classification imprécise

Un classifieur standard est pertinent dans tous les cas où il prédit juste. Permettre à un classifieur de fournir, pour un objet à classifier, un ensemble de classes dont seulement une est la bonne peut conduire à des situations où le classifieur prédit juste mais sans être pertinent. Le compromis entre la pertinence et la prudence est au coeur de l'introduction de la classification imprécise. Le principe consistant à améliorer la prudence d'une prédiction au détriment de sa pertinence (privilégier la fiabilité de l'information plutôt que la précision de son contenu) a été

initialement introduit dans les travaux de R. A. Fisher (Fisher, 1935) et J. Neyman (Neyman, 1937) en inférence statistique avec le concept d'intervalle de confiance, ou région de confiance en statistique bayésienne, en associant un seuil de confiance à l'ensemble fourni comme estimation d'un paramètre par exemple.

La classification imprécise se construit en suivant les mêmes techniques que celles d'un classifieur standard. Dans le cadre de la classification standard (chapitre 2), un classifieur δ_Θ associe à un nouvel exemple \mathbf{x} dont nous ignorons la classe réelle notée θ_x^* , une classe précise $\delta_\Theta(\mathbf{x})$ dans l'ensemble des classes possibles $\Theta = \{\theta_1, \dots, \theta_n\}$. Ainsi, l'information précise que l'on espère du classifieur δ_Θ concernant la vraie classe inconnue θ_x^* de \mathbf{x} , est

$$" \theta_x^* = \delta_\Theta(\mathbf{x}) "$$

En présence d'imprécisions et/ou d'incertitudes importantes sur les caractéristiques des exemples, le risque que le classifieur se trompe, c'est à dire qu'en réalité $\theta_x^* \neq \delta_\Theta(\mathbf{x})$, peut devenir inacceptable. Plus les données sont entachées d'imperfections, plus les chances que le classifieur se trompe augmentent. Si certaines applications peuvent tolérer une certaine propension à l'erreur, d'autres très sensibles ne peuvent pas se permettre de prendre des risques trop conséquents.

Exemple 1. Diagnostic médical Prenons l'exemple d'un diagnostic médical. Considérons un patient présentant les symptômes "toux", "fièvre", et "maux de tête" (le vecteur de caractéristiques est $\mathbf{x} = (\text{toux}, \text{fièvre}, \text{maux de tête})$). Supposons que nous devions effectuer le diagnostic dans l'ensemble des maladies possibles $\Theta = \{\text{Covid}, \text{Grippe}, \text{Angine}\}$. Pour ce patient les trois maladies sont envisageables (toutes compatibles avec les symptômes). Cependant avec un classifieur standard δ_Θ nous aboutirons à un diagnostic précis sans réel fondement, par exemple $\delta_\Theta(\mathbf{x}) = \text{Angine}$, qui pourrait avoir des conséquences graves sur le suivi du patient si il ne s'agissait pas du bon diagnostic.

La classification imprécise aussi appelée *classification indéterminée* (Yang et al., 2016) ou encore *class-selective* (Ha, 1997) ou *set-valued* (Mortier et al., 2019) en anglais, vise à diminuer les erreurs que commettraient des classifieurs standards dans des contextes comme celui de l'exemple 1. Pour ce faire, la classification imprécise autorise des sorties sous la forme d'ensembles de classes candidates au lieu de parier sur l'une des classes. Nous donnons la définition formelle suivante qui est commune à toutes les références citées précédemment.

Définition 1. Soit \mathcal{X} l'espace de représentation des objets à classifier. Un *classifieur imprécis*, noté δ_{2^Θ} , est un classifieur qui associe, pour un objet à classifier $\mathbf{x} \in \mathcal{X}$ dont on ignore la classe notée θ_x^* , un sous-ensemble de $2^\Theta \setminus \emptyset$ noté $\delta_{2^\Theta}(\mathbf{x})$. L'interprétation d'une sortie $\delta_{2^\Theta}(\mathbf{x})$ est la suivante

$$" \theta_x^* \in \delta_{2^\Theta}(\mathbf{x}) ",$$

ce qui signifie que la vraie classe θ_x^* de \mathbf{x} , par essence une classe unique, est l'une des classes du sous-ensemble $\delta_{2^\Theta}(\mathbf{x})$, faute d'information supplémentaire, il n'est pas possible de dire laquelle précisément. Afin d'alléger la lecture nous employons l'expression *sortie imprécise* pour désigner une sortie $\delta_{2^\Theta}(\mathbf{x})$ d'un classifieur imprécis (même lorsque cette sortie est un singleton).

Plusieurs remarques sont à souligner à partir de cette définition :

- Le niveau maximal d'imprécision est atteint lorsque $\delta_{2^\Theta}(\mathbf{x}) = \Theta$. Cette sortie correspond à l'ignorance totale sur la vraie classe de \mathbf{x} , θ_x^* . Nous faisons référence à cette sortie par l'expression *prudence maximale*.
- Il est possible de restreindre l'ensemble d'arrivée car certains sous-ensemble de Θ peuvent ne pas être d'intérêt pour la problématique étudiée. Dans ce cas, la complexité de la décision est réduite puisque nous n'avons plus besoin de parcourir les $2^n - 1$ sorties possibles.

- La définition laisse la liberté quant à la façon d'apprendre $\delta_{2\Theta}$. En effet, il est possible d'apprendre $\delta_{2\Theta}$ à partir d'exemples dont les classes sont uniquement des éléments de Θ comme il est possible d'envisager des exemples dont les classes imprécisément connues sont dans 2^Θ .

La classification imprécise concerne en premier lieu les domaines sensibles pour lesquels les erreurs de classification ont de graves conséquences. Par exemple, comme le montre l'exemple 1, un mauvais diagnostic médical pourrait engendrer des soins inadaptés et avoir des répercussions importantes sur la santé des patients. L'usage de la classification imprécise est d'autant plus profitable lorsque le problème de classification est difficile en raison de données disponibles imparfaites. Nous donnons dans la suite quatre exemples d'applications/illustrations de la classification imprécise :

Exemple 2. Diagnostic médical Reprenons l'exemple 1, concernant le diagnostic d'un patient dont les caractéristiques sont données par le vecteur $\mathbf{x} = (\text{toux}, \text{fièvre}, \text{maux de tête})$. Dans ce cas, un classifieur imprécis peut fournir l'ensemble $\delta_{2\Theta}(\mathbf{x}) = \{\text{Covid}, \text{Grippe}, \text{Angine}\}$ des maladies compatibles avec les symptômes observés. Le classifieur imprécis ne fournira pas de diagnostic précis, mais évitera de commettre une erreur sur la base des seules informations disponibles.

Exemple 3. Détection d'obstacle Un autre exemple de classification impliquant de graves conséquences a été donné dans la thèse (Yang, 2016) avec un problème de reconnaissance d'obstacles par des véhicules intelligents : les trois types de situations envisagées sont la présence d'un humain ou d'une bicyclette (qui doivent être évités) ou l'absence d'obstacle. Nous avons $\Theta = \{\text{humain}, \text{bicyclette}, \text{absence d'obstacle}\}$. Dans cet exemple, les erreurs peuvent avoir des conséquences graves comme un accident si la présence d'un humain ou d'une bicyclette n'est pas détectée. Par exemple, en cas d'hésitation du classifieur entre "bicyclette" et "absence d'obstacle", le classifieur est capable d'éviter une erreur en renvoyant la sortie $\delta_{2\Theta}(\mathbf{x}) = \{\text{bicyclette}, \text{absence d'obstacle}\}$. Cette information plus prudente permettrait au véhicule intelligent d'éviter la zone par précaution.

Exemple 4. Diagnostic de démences Une application de la classification imprécise dans le cas du diagnostic de la démence a été proposée dans (Zaffalon et al., 2003). Dans cet article, l'objectif est de détecter des cas de démence dans $\Theta = \{\text{maladie d'Alzheimer}, \text{démence vasculaire}, \text{Démence à corps de Lewy}\}$, pour lesquels il n'existe pas de solutions médicales lorsque la maladie est avancée, mais dont certaines manifestations pourraient être retardées voire évitées si la démence était détectée précocement. Pour ce diagnostic, les symptômes disponibles sont des données cognitives pour lesquelles il n'est pas toujours possible d'avoir l'information sur tous les attributs. En s'abstenant de donner des diagnostics précis lorsque les informations étaient insuffisantes, les auteurs ont obtenu un taux de justesse très élevé sur les sorties précises (diagnostic précis).

Exemple 5. Tri de plastiques Dans notre étude qui concerne le tri de plastiques dans $\Theta = \{\text{ABS}, \text{HIPS}, \text{PE}, \text{PP}\}$ à des fins de recyclage, les données disponibles sont affectées par de l'incertitude et de l'imprécision (qualité des mesures, vieillissement des plastiques, etc.). De plus, les erreurs de tri correspondent à des impuretés dans les bacs de tri : si la proportion d'impuretés dans un bac excède un certain seuil, le recyclage est compromis. Par exemple, d'après (Perrin et al., 2016), la tolérance maximale d'impuretés de type "plastique ABS" dans un mélange "plastique HIPS" est de 4%. Au delà de ce seuil, les propriétés mécaniques de la matière recyclée seraient trop dégradées. Dans le cas d'hésitation entre ABS et HIPS le classifieur imprécis doit renvoyer la sortie $\delta_\Theta(\mathbf{x}) = \{\text{ABS}, \text{HIPS}\}$ ce qui permet d'envoyer ce

fragment dans le bac dédié aux styreniques (plastiques ABS ou HIPS) et évite une possible impureté dans le bac de l'ABS ou celui de l'HIPS.

La classification imprécise constitue un moyen de diminuer les erreurs mais elle n'est pas forcément applicable dans tous les contextes. Notamment une sortie imprécise n'est pas toujours exploitable si le contexte applicatif impose des contraintes au décideur.

Exemple 6. Manufacture Prenons l'exemple d'une usine de manufacture dans laquelle un système est prévu pour détecter automatiquement les pièces défectueuses. Nous avons $\Theta = \{\text{défectueuse}, \text{non défectueuse}\}$. Dans le cas où l'usine ne dispose pas d'autre moyen que ce système pour évaluer les pièces, il n'y a pas d'intérêt à prédire la sortie imprécise Θ . La classification imprécise retrouverait de l'intérêt seulement si un traitement spécifique au cas de pièces douteuses était envisagé comme la vérification par du personnel.

Les sorties imprécises doivent donc être exploitables dans le problème applicatif. Les sorties imprécises peuvent être utiles dans les situations suivantes :

- Comme l'a souligné (Yang et al., 2016), la classification imprécise peut également constituer un moyen de pré-sélectionner un sous-ensemble de classes, réduisant ainsi la taille du problème de classification qui peut par la suite être traité par des moyens plus conséquents en termes de calculs, d'identifications expertes ou analyses supplémentaires (Antonucci et al., 2012).
- Plus généralement, la classification imprécise est utile comme étape intermédiaire dans un processus de décision afin de garder une information prudente sur les objets classifiés.

Le cas du tri de plastiques rentre dans les deux dernières situations. En effet, les sorties fournies par un classifieur imprécis peuvent être directement associées au choix d'un bac dédié à plusieurs plastiques. Le contenu d'un tel bac, par exemple dédié aux *styreniques* (plastiques ABS ou HIPS), pourrait être traité en effectuant une seconde passe de tri avec plus de moyens techniques, soit être valorisé pour des besoins de recyclage moins exigeants. De plus les informations prudentes concernant les fragments triés dynamiquement peuvent être intégrées au processus de décision pour les nouveaux fragments à trier afin de respecter des contraintes sur les impuretés dans les bacs de tri. Nous traitons ce dernier aspect dans le chapitre 5.

Pour garder son intérêt, la classification imprécise doit adapter son niveau de prudence au niveau de l'imperfection des données. Un classifieur imprécis doit réserver sa prudence pour les exemples difficiles (avec imperfections) en fournissant des ensembles de classes. Pour les exemples faciles (sans imperfections), la prudence n'est pas pertinente. En effet dans le cas extrême, fournir systématiquement une sortie de *prudence maximale* n'apporte aucune information utile. Dans notre domaine d'application, cela reviendrait à classer systématiquement des plastiques dans des bacs moins sélectifs et donc de valeur marchande moindre. Dans l'exemple d'un diagnostic médical, cela correspondrait à un médecin qui dirait systématiquement à tous ses patients (même pour les cas évidents) : "je ne sais pas ce dont vous souffrez". Ainsi, nous voulons insister sur le fait qu'en pratique la classification imprécise ne se résume pas à la seule possibilité de fournir des ensembles de classes candidates pour réduire les risques d'erreurs car la prudence a un coût dans n'importe quelle situation. Par conséquent, nous affinons la définition 1 de la classification imprécise en insistant sur ce qui est attendu de celle-ci.

Définition 2. Soit \mathcal{X} l'espace de représentation des objets à classifier. Un *classifieur imprécis*, noté $\delta_{2\Theta}$, est un classifieur qui associe, pour un objet à classifier $\mathbf{x} \in \mathcal{X}$ dont on ignore la classe notée $\theta_{\mathbf{x}}^*$, un sous-ensemble de $2^\Theta \setminus \emptyset$ tenant compte des contraintes suivantes associées à deux ambitions antagonistes :

- **La prudence** : La sortie imprécise $\delta_{2\Theta}(\mathbf{x})$ doit contenir la vraie classe $\theta_{\mathbf{x}}^*$ afin de minimiser les erreurs.
- **La pertinence** : Le cardinal de la sortie imprécise $\delta_{2\Theta}(\mathbf{x})$ doit être le plus petit possible de sorte que l'information " $\theta_{\mathbf{x}}^* \in \delta_{2\Theta}(\mathbf{x})$ " fasse encore sens pour la décision à prendre.

A la lumière de ce qui précède, nous pouvons formuler l'objectif de la classification imprécise d'un exemple \mathbf{x} , comme l'attribution d'un sous-ensemble $\delta_{2\Theta}(\mathbf{x})$ dont tous les éléments sont candidats à la véritable classe $\theta_{\mathbf{x}}^*$ et le manque d'information empêche de renvoyer une sortie plus pertinente (un sous-ensemble $\delta_{2\Theta}(\mathbf{x})$ plus petit).

Notons que les deux contraintes de la prudence et de la pertinence sont antagonistes car plus l'ensemble $\delta_{2\Theta}(\mathbf{x})$ est grand, plus il a de chances de contenir la vraie classe $\theta_{\mathbf{x}}^*$ (donc plus il est prudent de le choisir) mais en revanche moins l'information $\theta_{\mathbf{x}}^* \in \delta_{2\Theta}(\mathbf{x})$ est pertinente. Dans la pratique, un classifieur imprécis doit donc effectuer un compromis entre la prudence et la pertinence dont la subjectivité est fonction des enjeux de l'application.

Dans le cas du tri de plastiques, par exemple, plus l'algorithme pour le choix du bac sera prudent, plus il aura tendance à envoyer des fragments dans des bacs dédiés à des catégories non spécifiques de plastiques. Cela permettra de réduire les erreurs de tri (impuretés) dans les bacs dédiés à des plastiques spécifiques. En revanche, moins cet algorithme sera prudent, plus les bacs dédiés aux plastiques spécifiques se rempliront ce qui permettra de maximiser la quantité de matière valorisée, stratégie la plus rentable bien sûr à condition de ne pas commettre trop d'erreurs et dépasser la tolérance d'impuretés dans un bac.

Dans cet exemple, l'algorithme doit trouver un compromis entre les exigences de pureté sur les matières triées et la quantité de ces matières valorisables.

Nous verrons plus loin (section 3.3) plusieurs façons de modéliser le compromis qu'induit le respect des deux contraintes de la classification imprécise.

La section sous-suivante présente une illustration de l'intérêt de la classification imprécise lorsque de l'imprécision apparaît dans les données.

3.1.2 Illustration de l'intérêt de la classification imprécise

Afin d'illustrer les apports de la classification imprécise par comparaison avec la classification standard, considérons le classifieur standard du *naïf bayésien* (NBC) et le classifieur imprécis NONDET (qui est présenté dans la section 3.2.1.1) basé ici sur une version probabiliste de NBC. Considérons le problème de classification multi-classes standard pour lequel un nouvel exemple doit être attribué à l'une des trois classes $\theta_1, \theta_2, \theta_3$ sur la base de deux caractéristiques. Les trois classes sont simulées par des distributions gaussiennes bivariées de paramètres respectifs $\mu_1 = (0, 0)$, $\mu_2 = (0, 1)$, $\mu_3 = (1/2, \sqrt{3}/2)$ et $\sigma_1 = \sigma_2 = \sigma_3 = 0.01$. Nous allons considérer deux configurations dans cet exemple :

1. Le cas d'un problème de classification facile pour lequel les données ne sont pas entachées d'imperfections (voir Fig. 3.1a). Ce serait le cas par exemple, si les plastiques que nous triions étaient des plastiques non vieillis par le temps, dont les caractéristiques d'origine laissent peu de place à l'ambiguïté. Notons que les paramètres $\mu_1, \mu_2, \mu_3, \sigma_1, \sigma_2$ et σ_3 ont été choisis pour faire en sorte qu'il n'y ait pas de chevauchements de classes.
2. Le cas d'un problème de classification plus difficile pour lequel les données sont imparfaites. L'imperfection des données (imprécision et l'incertitude) est simulée par l'ajout d'un bruit gaussien ϵ de variance 0.1 (voir Fig. 3.1b). Avec ces imperfections les classes se chevauchent nettement ce qui rend la tâche de classification plus difficile.

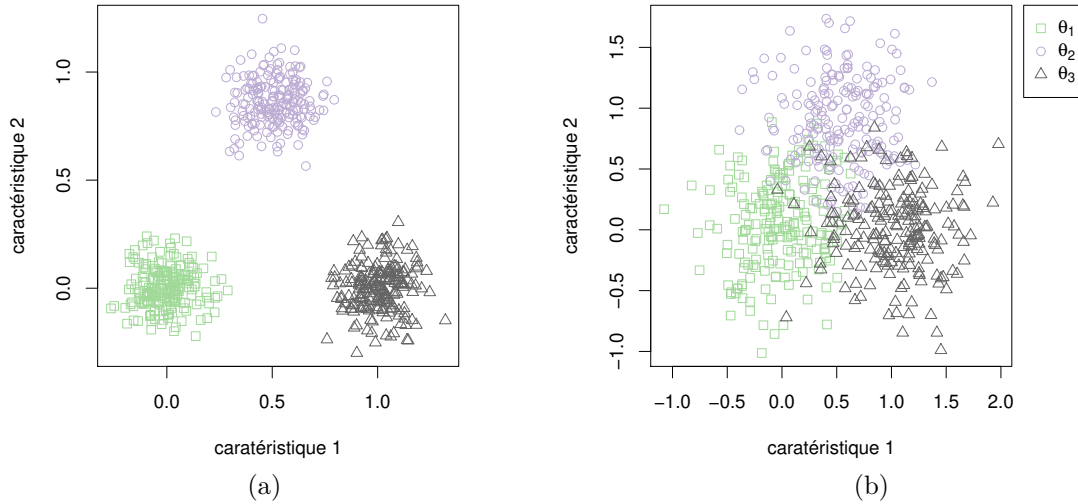


FIG. 3.1 Bases d'apprentissage pour les configurations facile (a) et difficile (b)

La Fig. 3.1 montre des bases d'apprentissage ainsi simulées avec chacune 600 exemples (200 exemples dans chaque classe). Afin d'illustrer le comportement d'un classifieur standard nous le testons sur 600 autres données pareillement simulées. La Fig. 3.2 montre les classifications obtenues pour les nouveaux exemples dans les deux configurations. Dans le cas facile, le taux de justesse est de 1. Dans le cas difficile, il est de 0.87 (le taux d'erreur est donc 0.13). Avec cette configuration, les erreurs sont inévitables pour le classifieur standard. En effet, lors du traitement d'un exemple situé dans une zone de chevauchement, il y a plusieurs classes possibles pour l'exemple. Cependant le classifieur est obligé de parier sur une seule classe. Cette prise de risque peut aboutir à une bonne prédiction ou à une erreur, mais dans tous les cas la sortie du classifieur standard n'est pas justifiée.

Voyons maintenant ce que donne le classifieur imprécis sur ces mêmes exemples de test. La Fig 3.3 donne les résultats de classification pour les deux configurations. Insistons ici sur le fait que pour des problèmes de classification faciles, faire appel à un classifieur imprécis n'est pas nécessaire. Du point de vue d'un classifieur imprécis, nous devons tout de même nous assurer que ses sorties sont pertinentes dans le cas facile. Le taux de justesse ne s'applique pas trivialement ici car une sortie sous la forme d'un ensemble n'est pas directement comparable à la vraie classe (l'évaluation d'un classifieur imprécis est traitée dans la section 3.3). Néanmoins nous obtenons 100% de classifications exactes (la sortie imprécise est un singleton qui correspond à la vraie classe) pour le cas facile et 0.817 pour le cas difficile. Dans le cas difficile, nous obtenons des sorties imprécises dans 10.5% des cas constituées de plus d'une classe pour les exemples situés dans les zones de chevauchements ce qui intuitivement semble légitime. Ces sorties sous forme de sous-ensembles traduisent l'attitude prudente du classifieur imprécis, là où le classifieur standard est obligé de *parier*.

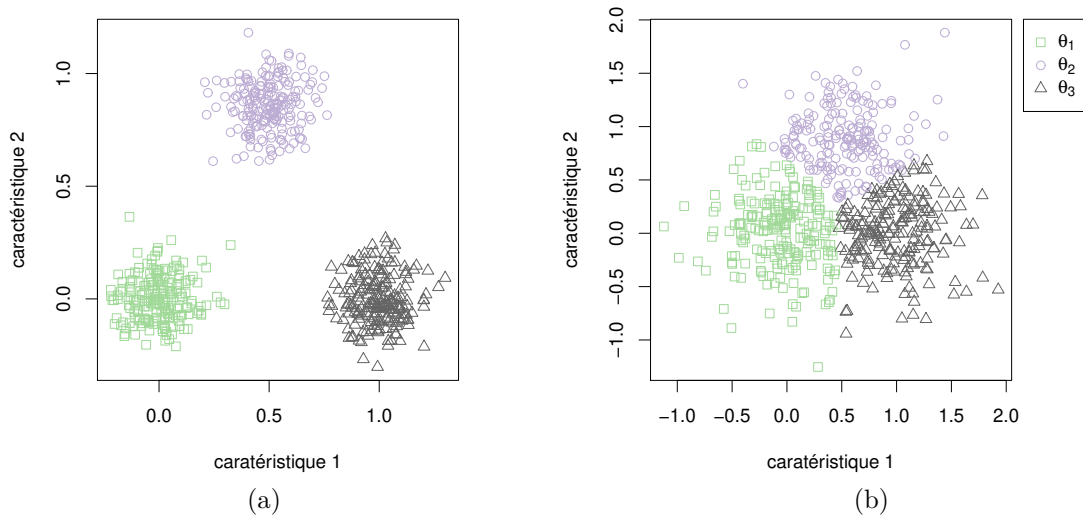


FIG. 3.2 Classification **standard** pour les configurations facile (a) et difficile (b)

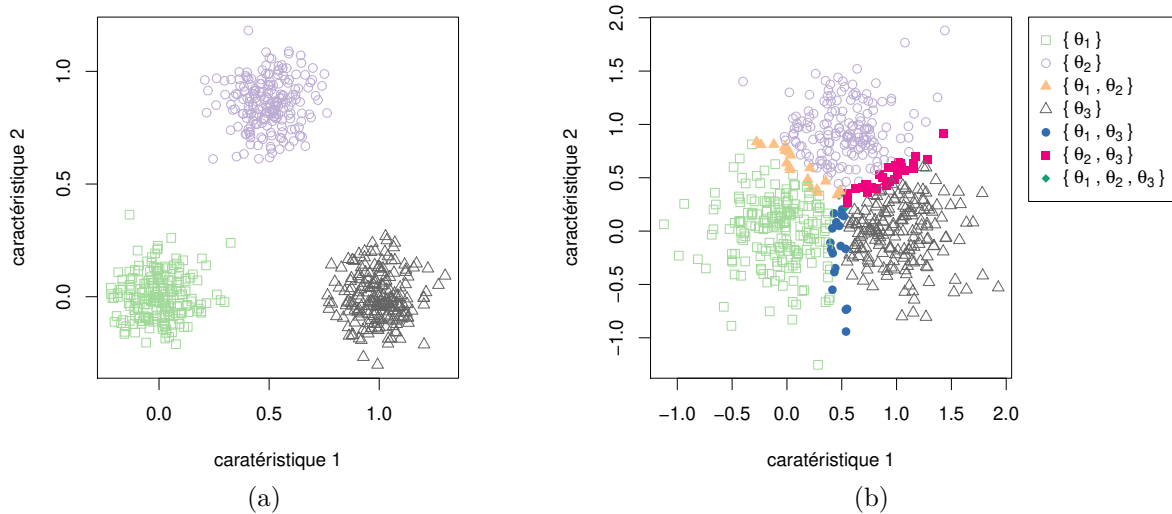


FIG. 3.3 Classification **imprécise** pour les configurations facile (a) et difficile (b)

Afin de comparer les deux classifieurs, considérons la matrice ci-dessous qui montre la répartition (normalisée sur l'intervalle $[0, 1]$) des exemples de tests suivant les résultats des deux classifieurs. Pour le classifieur standard, les résultats ont à chaque fois deux issues possibles : pour un exemple \mathbf{x} de vraie classe $\theta_{\mathbf{x}}^*$, nous comptons "exactitude" si $\delta_{\Theta}(\mathbf{x}) = \theta_{\mathbf{x}}^*$ et "erreur" si $\delta_{\Theta}(\mathbf{x}) \neq \theta_{\mathbf{x}}^*$. Pour le classifieur imprécis, nous avons plus de résultats possibles en fonction de l'inclusion de la vraie classe $\theta_{\mathbf{x}}^*$ et de la taille du sous-ensemble $\delta_{2\Theta}(\mathbf{x})$. Pour simplifier nous répertorions ici seulement trois types de résultats pour le classifieur imprécis : "exactitude" si $\delta_{2\Theta}(\mathbf{x}) = \theta_{\mathbf{x}}^*$, "imprécision" si $\theta_{\mathbf{x}}^* \in \delta_{2\Theta}(\mathbf{x})$ et "erreur" si $\theta_{\mathbf{x}}^* \notin \delta_{2\Theta}(\mathbf{x})$.

		Classifieur imprécis				
		exactitude ($\delta_{2\Theta}(\mathbf{x}) = \theta_{\mathbf{x}}^*$)	imprécision ($\theta_{\mathbf{x}}^* \in \delta_{2\Theta}(\mathbf{x})$)	erreur ($\theta_{\mathbf{x}}^* \notin \delta_{2\Theta}(\mathbf{x})$)		
Classifieur standard	exactitude ($\delta_{\Theta}(\mathbf{x}) = \theta_{\mathbf{x}}^*$)	0.817	0.05	0	0.867	Total classifieur standard
	erreur ($\delta_{\Theta}(\mathbf{x}) \neq \theta_{\mathbf{x}}^*$)	0	0.055	0.075	0.13	
		0.817	0.105	0.075	Total classifieur imprécis	

Nous constatons que les erreurs sont fortement réduites avec le classifieur imprécis : le taux d'erreurs passe de 0.13 à 0.075 soit une baisse d'environ 42% (42% des erreurs de δ_{Θ} deviennent des imprécisions pour $\delta_{2\Theta}$). En contrepartie, le taux d'exactitude baisse également : il passe de 0.867 à 0.817 c'est à dire une baisse d'environ 6%. Ce type de contrepartie n'est pas complètement évitable étant donné que le classifieur standard a parfois de *la chance en pariant*. Un classifieur imprécis a pour objectif de limiter cette contrepartie.

3.1.3 Distinction avec la classification multi-labels

Commençons par différencier les notions de *label* et de *classe*. Dans cette thèse nous avons jusqu'à présent considéré que chaque objet à classifier est un représentant d'une seule catégorie de Θ , appelée *classe*. Lorsque qu'un objet peut être représentant de plusieurs catégories, nous parlons de *labels*.

La classification multi-labels est une généralisation de la classification standard multi-classes pour laquelle les exemples peuvent posséder plusieurs labels simultanément. Pour un exemple $x \in \mathcal{X}$, un classifieur multi-labels attribue un vecteur (\dots, i, \dots) où $i = 1$ si x est attribué au i -ème label et $i = 0$ sinon.

Exemple 7. *Considérons un problème multi-labels consistant à assigner des labels de l'ensemble {témoignage, courrier, facture, justificatif de domicile, carte d'identité} à des documents. Dans le cas, où l'exemple x à classifier est une facture d'un fournisseur d'électricité, le classifieur multi-labels doit fournir la sortie $(0, 1, 1, 1, 0)$ car le document x est à la fois un courrier, une facture et un justificatif de domicile.*

La classification multi-labels permet donc également de fournir plusieurs labels mais l'interprétation de la sortie est différente de la classification imprécise. Avec un classifieur multi-labels nous obtenons une information précise concernant les vrais labels de x , alors que dans le cas d'un classifieur imprécis, seulement l'un des éléments de l'ensemble est la bonne classe mais on ne sait pas lequel.

Des versions imprécises du problème multi-labels ont récemment été proposées (Pillai et al., 2013; Nguyen et Hüllermeier, 2019) où l'ignorance concernant un label peut être exprimée. Pour un exemple à classifier x , ce type de classifieur imprécis multi-labels fournit un vecteur (\dots, i, \dots) où $i = 1$ si le classifieur estime que x est un exemple appartenant au i -ème label, $i = 0$ sinon et $i = \{0, 1\}$ s'il souhaite exprimer son ignorance. Reprenons le problème de classification multi-labels de l'exemple 7. Si un classifieur imprécis multi-labels fournissait une sortie $(\{0, 1\}, 1, 0, 0, 0)$ pour un exemple x , cela signifierait que le classifieur estime que x est un courrier mais ne sait pas s'il peut s'agir également d'un témoignage ou pas.

3.1.4 Classification imprécise versus la classification hiérarchique

La classification imprécise pourrait plus justement être rapprochée de la classification hiérarchique. D'après (Silla et Freitas, 2011), la classification hiérarchique se distingue de la classification multi-classes standard en considérant des classes faisant partie d'une structure hiérarchique pré-établie muni d'un ordre partiel traduisant une relation d'inclusion. Les classes correspondent alors à des nœuds dans une structure régie par des liens "IS-A" dits *parents-enfants* où la classe *enfant* est incluse dans la classe *parents* (e.g., la classe *enfant* Souris est incluse dans la classe *parent* Rongeurs). Les nœuds n'ayant pas d'enfants sont appelés *feuilles* de la structure hiérarchique. La classification imprécise peut être rapprochée de la classification hiérarchique en considérant que la prédiction d'un nœud *parent* dans la hiérarchie correspond à la prédiction imprécise d'un ensemble contenant toutes les feuilles *enfant* rattachées à ce nœud. Nous constatons deux différences entre ces deux types de classification.

Premièrement, dans le cas de la classification hiérarchique, les classes sont organisées dans une structure hiérarchique pré-établie alors que pour la classification imprécise les sorties peuvent être dans n'importe quel sous ensemble de Θ , même s'il n'a pas de sémantique particulière dans la réalité. Par exemple, l'ensemble $\{Souris, Mulot, Belette\}$ ne définit pas une famille, un genre ou une variété d'animaux qui fasse sens dans la classification codifiée du règne animal, mais elle reste pour autant une sortie possible pour un classifieur imprécis. Il serait en effet simple de restreindre l'ensemble des sorties imprécises aux seuls ensembles existant dans une hiérarchie pré-établie. De ce point de vue, un classifieur imprécis pourrait alors être employé pour faire de la classification hiérarchique.

Deuxièmement, toujours d'après (Silla et Freitas, 2011), le cas d'exemples partiellement étiquetés, i.e., dont la classe n'est pas forcément une feuille de la hiérarchie, a été considéré dans plusieurs travaux de classification hiérarchique. Cette situation se produit lorsque des exemples d'une base de données n'ont pas pu être précisément assignés à une feuille par manque de connaissance. Notons que nous ne sommes pas dans le cas d'une représentation conceptuelle telle que considérée dans (Harispe et al., 2015). Ainsi les exemples partiellement étiquetés ne sont pas des instanciations dans la nature de concepts mais bien des exemples ayant été attribués à un nœud dans la hiérarchie par manque de connaissance, la réalité de l'exemple est toujours une feuille dans notre cas. Nous verrons dans la sous-section 3.3.4 que l'existence de ces exemples partiellement étiquetés motive la notion d'*erreur de spécialisation* qui consiste à prédire un nœud plus spécifique que le nœud connu. A notre connaissance, cette notion n'est pas présente en classification imprécise qui considère des bases de données précisément étiquetés pour évaluer les performances d'un classifieur imprécis. La présence d'exemples partiellement étiquetés n'est pas une différence de fond entre la classification hiérarchique et la classification imprécise car il est envisageable de construire un classifieur que l'on entraînerait et/ou testerait sur des exemples partiellement étiquetés. Cependant nous souhaitons insister sur cette notion d'erreur de spécialisation qui pourrait avoir une autre interprétation en classification imprécise : l'erreur consistant à prédire un ensemble de classes candidates plus petit que ne le permet l'information disponible. Nous reviendrons sur cette notion dans la section 3.4.

Remarque 1. La classification hiérarchique est par construction un problème multi-labels car prédire un nœud équivaut à prédire le nœud et tous ses ancêtres (liens "IS-A"). Cependant la classification hiérarchique multi-labels désigne les problèmes de classification hiérarchique pour lesquels plusieurs nœuds distincts (plusieurs chemins de liens "IS-A") peuvent être simultanément attribués à un exemple.

3.1.5 Classification imprécise et option de rejet

La classification imprécise a été introduite par l'idée d'abstention apparue dans les années 70 (Chow, 1957, 1970) avec la notion d'*option de rejet* dans le cadre probabiliste. Cette option de rejet consiste à permettre une abstention totale pour les exemples difficiles. Cela correspond à la sortie Θ (prudence maximale) dans notre définition de la classification imprécise. Nous retrouvons l'option de rejet dans de nombreux problèmes de classification. Par exemple, (Loeffel et al., 2015) a proposé un classifieur avec l'option de rejet pour résoudre des problèmes de classification en ligne en situation de dérive conceptuelle, c'est à dire lorsque la relation entre les observations et la variable de la classe que l'on cherche à prédire peut changer au cours du temps.

Nous pouvons comprendre la classification imprécise comme une généralisation de l'option de rejet : l'abstention n'est plus nécessairement totale, elle peut être partielle (sous la forme d'un ensemble réduit de classes candidates).

La Fig. 3.4b illustre la notion d'option de rejet comme intermédiaire entre la classification standard et la classification imprécise. Sur cette figure, nous considérons un exemple fictif de problème de classification à 3 classes $\Theta = \{\theta_1, \theta_2, \theta_3\}$ sur la base de deux caractéristiques ($\mathcal{X} = \mathbb{R}^2$). Les figures 3.4a, 3.4b et 3.4c illustrent différentes partitions de l'ensemble des caractéristiques. Chaque partition est associée à la sortie qu'obtiendrait un exemple repéré dans l'espace de description. Dans la figure 3.4a, nous représentons un exemple de partition que nous pourrions obtenir avec un classifieur standard. Nous avons une frontière entre les différentes classes. En présence d'imperfections dans les données la définition précise de ces frontières est illusoire. Nous nous attendons donc à avoir davantage d'erreurs près des frontières. La figure 3.4b montre une modification de la partition en ajoutant une nouvelle option de sortie, Θ (option de rejet) pour la zone dans l'espace de caractéristiques proche des frontières. Cette option permet d'éviter les erreurs près des frontières puisqu'elle permet de conclure qu'on ne sait rien pour tout point difficile, mais elle peut être affinée. En effet les exemples proches de la frontière entre les classes θ_1 et θ_2 mais qui sont loin de la classe θ_3 , ont peu de chances d'appartenir à la classe θ_3 . Des options de sorties prudentes mais plus précises que Θ tout entier comme $\{\theta_1, \theta_2\}$ sont alors possibles avec la classification imprécise (illustrée dans la Fig. 3.4c).

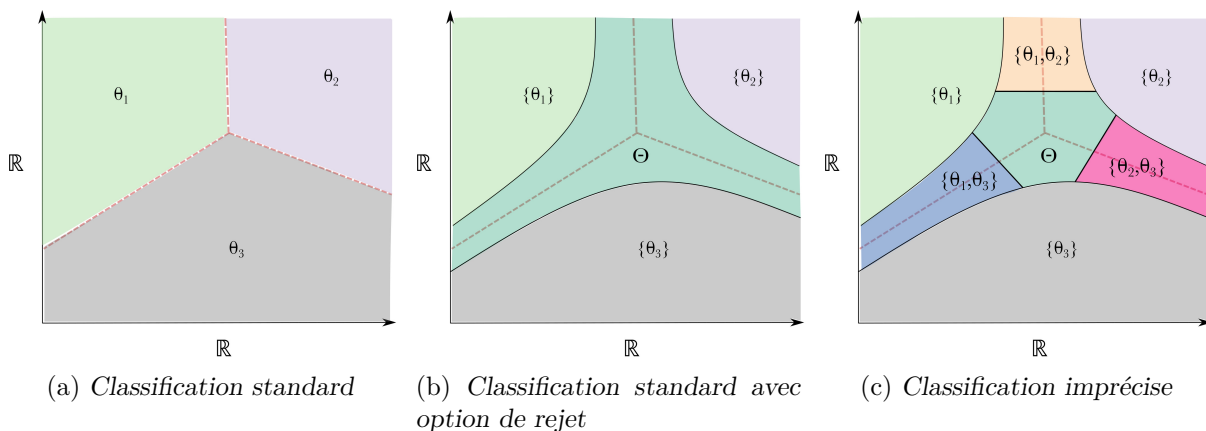


FIG. 3.4 Illustration de partitions pour un problème de classification (figure inspirée de (Ha, 1997))

3.2 Méthodes de classification imprécise

Dans la section précédente, nous avons donné la définition d'un classifieur imprécis en précisant son utilité et sa différence avec d'autres techniques de classification mais nous n'avons pas montré comment se construit un tel classifieur. Dans cette section, nous présentons une sélection de techniques de construction de classifieur imprécis qui illustre les principales tendances des approches de classification imprécises proposées dans la littérature. Etant donné que les erreurs de classification standard peuvent être causées par différents types d'imperfections des données et que les théories de l'incertain (chapitre 2) sont capables de représenter des types d'imperfections différentes, nous choisissons d'organiser la section selon les théories de l'incertain exploitées par la classification imprécise.

3.2.1 Classification imprécise basée sur la théorie des probabilités

Les méthodes de classification imprécises basées sur la théorie des probabilités exploitent la représentation de l'incertitude du cadre probabiliste pour fournir des sorties imprécises. Parmi les méthodes de classification imprécise basées sur la théorie des probabilités, nous distinguons deux catégories d'approches. La première catégorie est basée sur la maximisation d'espérance de gain et la seconde appelée *classification conforme* est basée sur la théorie des tests statistiques.

3.2.1.1 Méthodes basée sur l'espérance de gain

Afin de fournir à un exemple \mathbf{x} une sortie imprécise dans 2^Θ , les méthodes basées sur l'espérance de gain utilisent la probabilité a posteriori $p(\cdot|\mathbf{x})$ pour représenter l'incertitude à propos de la vraie classe de \mathbf{x} . Cette probabilité a posteriori peut être calculée par n'importe quel classifieur probabiliste. Les premiers travaux permettant une abstention totale (option de rejet) proposaient la modification suivante de l'espérance optimale (chapitre 2 sous-section 2.3.1) (Chow, 1957, 1970) :

$$\delta_{2^\Theta}(\mathbf{x}) = \begin{cases} \Theta & \text{(Prudence maximale)} & \text{si } \max_{j=1,\dots,n} p(\theta_j|\mathbf{x}) < t \\ \arg \max_{j=1,\dots,n} p(\theta_j|\mathbf{x}) & \text{(Espérance optimale)} & \text{sinon} \end{cases},$$

pour un seuil de rejet $t \in [0, 1]$. En d'autres termes, si l'incertitude est trop importante, de telle sorte que la plus grande probabilité a posteriori est inférieure à un seuil t , alors le classifieur s'abstient de donner une classe précise (prudence maximale Θ) pour cet exemple. Cette notion de rejet a été étendue au cas où le nombre de classes est mal connu dans (Dubuisson et Masson, 1993) avec une option de rejet basée sur la distance. Dans cet article, les exemples très distants des exemples d'apprentissage sont considérés comme des exemples appartenant potentiellement à une nouvelle classe (qui n'était pas représentée dans l'apprentissage) et sont rejetés pour cette raison.

Comme nous l'avons noté dans la section précédente, l'abstention partielle sous la forme d'ensemble de classes (classification imprécise) est plus récente. Parmi les premières règles de décision connues permettant cette abstention partielle, (Ha, 1997) évoquait les règles dites "*top-s*" et "*risque constant*" :

- La décision "*top-s*" consiste à fournir les s classes ayant obtenu les plus grandes probabilités.
- La décision "*risque constant*" consiste à fournir l'ensemble minimal de classes dont la probabilité cumulée dépasse un seuil.

Avec la décision "top- s ", la taille s de l'ensemble fourni est invariante quel que soit l'exemple \mathbf{x} . Cependant, il est préférable d'adapter la taille de l'ensemble fourni au niveau d'incertitude de l'exemple. En effet, pour un exemple pour lequel on dispose d'une information très incertaine, une sortie plus prudente sous forme d'un grand ensemble de classes est plus judicieuse. En revanche pour un exemple pour lequel on dispose d'une information de bonne qualité, un excès de prudence diminue la pertinence de la sortie inutilement. Concernant la décision "risque constant", la sensibilité au seuil peut être difficile à contrôler. Plus généralement, les deux règles de décision ne permettent pas de prendre en compte les éventuels pertes/gains associés aux classes.

Récemment, des règles de décision ont été employées pour fournir des sorties prudentes. Ces approches sont des généralisations des règles de décision basée sur la maximisation d'espérance de gain (ou minimisation de risque). Ainsi l'ensemble des actions qui était classiquement $\mathcal{A} = \Theta$ devient $\mathcal{A} = 2^\Theta \setminus \emptyset$ (ou un sous-ensemble de $2^\Theta \setminus \emptyset$). Le risque associé à une action $A \in \mathcal{A}$ (sortie imprécise), i.e., action de choisir A comme ensemble des classes candidates pour un objet \mathbf{x} , est défini comme :

$$\mathbb{E}_p[A|\mathbf{x}] = \sum_{\theta \in \Theta} L(\theta, A)p(\theta|\mathbf{x})$$

où $L(\theta, A)$ correspond à la perte encourue lorsque qu'à un exemple dont la vraie classe est θ est associée la sortie imprécise A du classifieur. Plusieurs fonctions de gain/perte ont été proposées comme la mesure F_β (Coz et al., 2009) (qui sera présentée dans la sous-section 3.3.2) ou les fonctions de gain proposées dans (Zaffalon et al., 2012) (sous-section 3.3.3). Ces approches consistent ensuite à chercher l'ensemble minimisant le risque (ou maximisant l'espérance de gain) :

$$\delta_{2^\Theta}(\mathbf{x}) = \arg \min_{A \subseteq \Theta} \mathbb{E}[A|\mathbf{x}].$$

La recherche de l'optimum de manière naive est exponentielle $\mathcal{O}(2^n)$. Cette complexité ne permet pas de traiter les cas où le nombre de classes n est grand. Des algorithmes efficaces de recherche de l'optimum ont été proposés dans le cas où la fonction perte est $1 - F_\beta$ (voir sous-section 3.3.2) (Coz et al., 2009) et dans le cas d'une famille de fonctions de perte plus générale (Mortier et al., 2019). Dans cette dernière référence, l'algorithme proposé (Set-Valued Bayes-Optimal Prediction) garantit une solution exacte du problème avec une complexité $\mathcal{O}(n \log(n))$.

Nous pouvons retenir, dans cette catégorie d'approches, l'approche de classification imprécise proposée dans (Coz et al., 2009) appelée "nondeterministic classification" en anglais (et que nous notons dans cette thèse par NONDET), qui est basée sur la minimisation de risque avec la fonction de perte $1 - F_\beta$.

Ce type d'approche a l'avantage d'être générique, i.e., la probabilité a posteriori $p(\cdot|\mathbf{x})$ peut être fournie par n'importe quel classifieur probabiliste. De plus les pertes sont prises en compte ce qui n'était pas le cas avec les règles basées sur des seuils.

3.2.1.2 Classification conforme

Une autre famille de classifieurs imprécis appelés *classifieurs conformes* (Shafer et Vovk, 2008) généralise le concept d'intervalle de confiance visant à améliorer la justesse d'une sortie au détriment de sa pertinence. Les classifieurs conformes ont été originellement introduits pour la configuration de classification en ligne, c'est à dire que chaque exemple \mathbf{x} est traité, puis sa vraie classe est révélée et enfin \mathbf{x} est ajouté à la base de données. En fonction d'une mesure de non-conformité¹, et pour une probabilité d'erreur ϵ , un classifieur conforme garantit que

1. Une mesure de non-conformité mesure à quel point un exemple semble inhabituel comparé aux exemples précédents.

l'ensemble A des classes candidates contient la vraie classe θ_x^* avec une probabilité de $1 - \epsilon$. Pour le traitement d'un exemple arrivant \mathbf{x} , les anciens exemples sont utilisés pour calculer des scores de non-conformité concernant chaque classe candidate $\theta \in \Theta$. Ces scores de non-conformité sont ensuite utilisés pour rejeter ou non chaque classe. L'ensemble des classes n'ayant pas été rejetées constitue la sortie imprécise. Plus précisément chaque classe candidate θ fait l'objet d'un test d'hypothèse où l'hypothèse $\theta_x^* = \theta$ est rejetée si le score de non-conformité est compris parmi les $\epsilon\%$ plus grands scores obtenus pour tous les anciens exemples. Récemment, des approches de classification conforme hors ligne ont été proposées (Papadopoulos, 2008).

3.2.2 Classification imprécise basée sur les probabilités imprécises

Alors que la théorie des probabilités ne considère que l'incertitude, d'après le chapitre 2, le cadre des probabilités imprécises permet de modéliser l'incertitude et l'imprécision. Nombre des méthodes de classification imprécise ont été proposées dans le cadre des probabilités imprécises (voir la TABLE 3.6 pour quelques exemples de travaux). Ces approches consistent à fournir l'ensemble des classes non-dominées selon un ordre partiel qu'elles induisent sur l'ensemble des classes à partir d'un ensemble crédal \mathcal{P} . Une introduction à l'une de ces approches, le classifieur naïf crédal, "Naive Credal Classifier" (NCC) en anglais (Zaffalon, 2001), est donnée en annexe section D. Retenons que la complexité de la classification d'un exemple $\mathbf{x} = (x_1, \dots, x_d)$ avec cette méthode est $\mathcal{O}(dn^2)$ avec $n = |\Theta|$.

Dans la suite nous exposons la procédure générale pour construire un classifieur imprécis dans le cadre des probabilités imprécises.

1) Inférence d'un ensemble crédal \mathcal{P} Dans ces travaux, l'information disponible concernant l'exemple à traiter \mathbf{x} est donc un ensemble crédal \mathcal{P} de distribution de probabilités précises sur les classes de Θ . La difficulté de la procédure est d'inférer cet ensemble crédal. Plusieurs approches ont été considérées pour fournir \mathcal{P} . L'une des façons est d'utiliser le modèle de Dirichlet imprécis, comme cela a été fait dans (Zaffalon, 2001) (voir annexe section D). Une autre façon de construire un ensemble crédal est de se baser sur des contraintes d'intervalles (De Campos et al., 1994) (voir (Zaffalon, 1999) pour une version du classifieur NCC basé sur cette méthode).

2) Déduction d'un ordre partiel sur les classes Dans l'approche NCC de (Zaffalon, 2001) introduite en annexe D, l'ordre partiel, noté $\succ_{\mathcal{ID}}$ est inféré en utilisant le critère de *dominance crédale* qui consiste à préférer la classe θ_i si la probabilité inférieure $\underline{P}(\theta_i|\mathbf{x})$ associée à θ_i est supérieure à la probabilité supérieure $\overline{P}(\theta_j|\mathbf{x})$ associée à θ_j où $\underline{P}(\theta_i|\mathbf{x}) = \min_{P \in \mathcal{P}} P(\theta_i|\mathbf{x})$ et $\overline{P}(\theta_j|\mathbf{x}) = \max_{P \in \mathcal{P}} P(\theta_j|\mathbf{x})$. Rappelons que $P(\theta_i|\mathbf{x})$ se calcule, ici, en utilisant l'hypothèse naïve de l'indépendance conditionnelle par rapport aux classes. Nous avons :

$$\theta_i \succ_{\mathcal{ID}} \theta_j \iff \max_{P \in \mathcal{P}} P(\theta_j|\mathbf{x}) < \min_{P \in \mathcal{P}} P(\theta_i|\mathbf{x}).$$

D'autres critères existent comme le *critère de maximalité* (Walley et Peter, 1991), dont l'ordre partiel induit est noté $\succ_{\mathcal{M}}$, qui consiste à préférer θ_i si la probabilité associée à la classe θ_i est supérieure à la probabilité associée à la classe θ_j quelle que soit la probabilité considérée dans l'ensemble crédal :

$$\theta_i \succ_{\mathcal{M}} \theta_j \iff \forall P \in \mathcal{P}, \quad P(\theta_j|\mathbf{x}) < P(\theta_i|\mathbf{x}).$$

Notons que le critère de maximalité est moins conservatif que le critère de dominance crédale².

3) Sélection des classes candidates Une fois que nous avons un ordre partiel sur l'ensemble des classes, le classifieur imprécis n'a plus qu'à renvoyer le sous ensemble des classes non dominées :

$$\delta_{2\Theta}(\mathbf{x}) = \{\theta \in \Theta \mid \nexists \theta' \in \Theta \text{ telle que } \theta' \succ \theta\}.$$

Sur le même principe, (Nguyen et al., 2018) propose un classifieur imprécis basé sur la déduction d'un pré-ordre à partir d'un ensemble de classifieurs probabilistes.

3.2.3 Classification imprécise basée sur les fonctions de croyance

Peu de méthodes évidentielles ont été directement proposées avec comme objectif premier de fournir un ensemble de classes candidates pour classifier un objet. On note tout de même l'article (Liu et al., 2014) qui considère le k -nn proposé dans (Zouhal et Denoeux, 1998) pour fournir une fonction de masse qui sera utilisée comme entrée pour une méthode de seuillage qui fournit en sortie un ensemble de classes candidates.

Néanmoins la théorie de la décision adaptée au cas évidentiel (Dempster, 2008; Shafer, 1981; Denoeux, 2019) permet de construire simplement un classifieur imprécis à partir d'un classifieur évidentiel. En effet, supposons qu'un classifieur évidentiel nous fournisse pour l'exemple \mathbf{x} une fonction de masse $m(\cdot|\mathbf{x})$ définie sur 2^Θ . Nous avons vu dans le chapitre 2 sous-section 2.3.2, que de nombreuses stratégies de choix de la classe de \mathbf{x} sont proposées dans la littérature. Parmi elles, certaines induisent un ordre de préférence complet sur les classes et d'autres un ordre de préférence partiel sur les classes (Denoeux, 2019).

Le cas de la décision à partir d'un ordre complet, induit par le critère du risque ou celui de l'espérance de gain a été présenté dans le chapitre précédent pour un ensemble d'actions \mathcal{A} . Afin de choisir un sous-ensemble, nous fixons $\mathcal{A} = 2^\Theta \setminus \emptyset$ et nous nous basons sur une fonction de perte L définie sur $\Theta \times 2^\Theta$. Ainsi nous pouvons utiliser les critères de décision évoqués dans le chapitre précédent (décision basée sur la probabilité pignistique, décision optimiste, décision pessimiste, généralisation du critère de Hurwicz, etc.). Par exemple, la stratégie optimiste consiste à minimiser \underline{r} et la stratégie pessimiste à minimiser \bar{r} , définis comme suit :

$$\underline{r}(A|\mathbf{x}) = \sum_{B \subseteq \Theta} m(B|\mathbf{x}) \min_{\theta \in B} L(A, \theta), \quad \bar{r}(A|\mathbf{x}) = \sum_{B \subseteq \Theta} m(B|\mathbf{x}) \max_{\theta \in B} L(A, \theta).$$

Ensuite un classifieur imprécis peut être construit en fonction du critère de décision. Par exemple avec la décision pessimiste, nous avons :

$$\delta_{2\Theta}(\mathbf{x}) = \arg \min_{A \subseteq \Theta} \bar{r}(A|\mathbf{x})$$

Il est donc possible de construire avec des outils existant un classifieur imprécis dans le cadre des fonctions de croyance en utilisant un classifieur évidentiel de l'état de l'art comme le k -nn évidentiel puis décider de la sortie imprécise par une minimisation de risque ou maximisation d'espérance de gain.

Il est également possible de construire un classifieur imprécis basé sur les fonctions de croyance en induisant un ordre partiel de préférence sur les classes. Par exemple, avec le critère de la *dominance d'intervalle*, les auteurs de (Quost et al., 2020a) proposent l'ordre partiel suivant :

$$\theta_i \underset{ID}{\succ} \theta_j \iff Bel(\{\theta_i\}|\mathbf{x}) > Pl(\{\theta_j\}|\mathbf{x}).$$

2. $\theta_i \underset{ID}{\succ} \theta_j \Rightarrow \theta_i \underset{\mathcal{M}}{\succ} \theta_j$ mais $\theta_i \underset{\mathcal{M}}{\succ} \theta_j \not\Rightarrow \theta_i \underset{ID}{\succ} \theta_j$.

Avec ce critère de dominance d'intervalle, nous pouvons sélectionner l'ensemble des classes non dominées (une classe θ_i est non dominée si $\forall j \neq i, \theta_j \not\prec_{\mathcal{D}} \theta_i$, c'est à dire si $Bel(\{\theta_j\}|\mathbf{x}) \leq Pl(\{\theta_i\}|\mathbf{x})$). Ainsi la sortie imprécise serait :

$$\delta_{2\Theta}(\mathbf{x}) = \{\theta_i | Pl(\{\theta_i\}|\mathbf{x}) \geq Bel(\{\theta_j\}|\mathbf{x}), \forall j \neq i\}.$$

Dans la pratique, la condition $Bel(\{\theta_j\}|\mathbf{x}) > Pl(\{\theta_i\}|\mathbf{x})$ est difficilement vérifiée. Nous pouvons nous attendre à ce que de nombreuses classes ne soient pas dominées. Ce critère aura donc tendance à favoriser les sorties imprécises très prudentes et peu pertinentes.

3.2.4 Classification imprécise basée sur la logique floue

Comme pour les fonctions de croyance, peu de méthodes basées sur la logique floue ont été directement proposées pour la classification imprécise. Nous notons tout de même l'approche proposée dans (Hüllermeier et Brinker, 2008) qui interprète des scores d'un classifieur binaire dans un schéma "un contre un", comme une relation de préférence floue et en déduit un ordre partiel qui peut être utilisé pour fournir un ensemble de classes candidates.

Nous pouvons également citer (Ishibuchi et Nii, 2000) qui propose plusieurs méthodes de classification imprécise basées sur des seuils à partir d'une version floue du classifieur basé sur les réseaux de neurones artificiels.

Notons aussi le travail de (Jiao et al., 2015) qui propose une classification basée sur des règles floues fournissant une fonction de masse en sortie. Cette sortie évidentielle pourrait être exploitée selon les règles de décision évoquées dans la section 3.2.3.

3.2.5 Classificaion imprécise basée sur la théorie des possibilités

Dans (Nguyen et al., 2018), la particularité réside dans le fait de distinguer l'incertitude épistémique de l'incertain aléatoire. Sur la base d'un ensemble de probabilités, les auteurs définissent quatre critères modélisant les deux types d'incertitudes pour chaque paire de classes et impliquant des distributions de possibilités. Une comparaison entre les deux classes sur la base de ces critères est établie. L'ensemble des classes fourni correspond aux classes non-dominées.

3.2.6 Autres méthodes de classification imprécise sans hypothèse de modèle

Certaines méthodes d'apprentissage automatique ne reposent pas directement sur des théories de l'incertain présentées dans les sections précédentes. Par exemple une approche présentée dans (Zhang et al., 2017) propose une famille de classifieurs imprécis basée sur des marges, i.e., impliquant une distance à une frontière de décision du classifieur.

3.3 Évaluation d'un classifieur imprécis

Le problème de l'évaluation d'un classifieur imprécis $\delta_{2\Theta}$, consiste à mesurer la satisfaction procurée par $\delta_{2\Theta}$ quand il fournit une prédiction $\delta_{2\Theta}(\mathbf{x})$ pour \mathbf{x} sachant que la réalité est la classe θ_x^* . Nous appelons *récompense* la valeur de la satisfaction mesurée.

Dans le cas classique de la classification supervisée, l'évaluation d'un classifieur standard δ_{Θ} pour une entrée \mathbf{x} dont la classe θ_x^* est connue, consiste à vérifier si la sortie $\delta_{\Theta}(\mathbf{x})$ est égale à θ_x^* (voir chapitre 2, sous-section 2.2.2). La question de l'évaluation d'une sortie imprécise est similaire au cas de la classification standard lorsque la sortie imprécise est un singleton (exactitude ou erreur). Cependant, lorsqu'une sortie imprécise est un ensemble constitué de plusieurs classes, l'évaluation de la sortie imprécise est tributaire de l'aversion au risque ou

autrement dit à l'appréciation de la prudence. Cette évaluation n'est donc pas absolue, mais dépend du contexte applicatif. Par conséquent, l'évaluation d'un classifieur classique ne pourrait être appliquée. La question de l'évaluation de sorties imprécises a fait l'objet de plusieurs travaux (Zaffalon, 1999; Abellan et Masegosa, 2012; Yang et al., 2016).

Exemple 8. *A travers cet exemple, nous voulons mettre en exergue les enjeux soulevés lors de l'évaluation d'un classifieur imprécis $\delta_{2\Theta}$. Considérons un problème de classification à trois classes correspondant au diagnostic de patients pour lesquels les maladies Covid-19, Grippe ou Angine sont plausibles. Nous avons $\Theta = \{Covid, Grippe, Angine\}$. Pour un vecteur de caractéristiques \mathbf{x} concernant un patient, le classifieur imprécis nous renseigne sur la maladie du patient en fournissant la sortie imprécise $\delta_{2\Theta}(\mathbf{x})$ correspondant au sous-ensemble des maladies candidates pour le patient. Supposons que le patient soit véritablement atteint de la covid-19 ($\theta_{\mathbf{x}}^* = Covid$). Dans ce cas, la meilleure sortie possible $\delta_{2\Theta}(\mathbf{x}) = \{Covid\}$ doit obtenir la récompense maximale égale à 1. A l'inverse toutes les sorties qui ne contiennent pas la vraie maladie du patient ne devraient pas être récompensées (récompenses égales à 0). Il reste la question des sorties imprécises sous forme de sous-ensembles comprenant $\theta_{\mathbf{x}}^*$. Par exemple : quelle récompense donner à la sortie imprécise $\delta_{2\Theta}(\mathbf{x}) = \{Covid, Grippe\}$? Cette sortie a l'inconvénient de fournir une information imprécise mais reste intéressante car l'hésitation entre Covid-19 et Grippe peut être levée par un examen plus poussé. La Table 3.1 résume le problème de l'évaluation d'une sortie fournie $\delta_{2\Theta}(\mathbf{x})$ lorsque la vraie maladie du patient est $\theta_{\mathbf{x}}^*$ (les points d'interrogation représentent les récompenses qui restent à déterminer). Pour chaque mesure d'évaluation présentée dans les prochaines sous-sections, nous verrons comment la prudence est récompensée dans cet exemple.*

		Ensemble fourni : $\delta_{2\Theta}(\mathbf{x})$						Θ
		$\{Covid\}$	$\{Grippe\}$	$\{Angine\}$	$\{Covid, Grippe\}$	$\{Covid, Angine\}$	$\{Grippe, Angine\}$	
Vraie classe : $\theta_{\mathbf{x}}^*$	<i>Covid</i>	1	0	0	?	?	0	?
	<i>Grippe</i>	0	1	0	?	0	?	?
	<i>Angine</i>	0	0	1	0	?	?	?

TABLE 3.1 – Récompenses de sorties imprécises

Afin de simplifier notre exposé nous introduisons dès maintenant une généralisation triviale de la justesse (chapitre 2) au cas de sorties imprécises. Cette mesure intervient dans la construction de la plupart des mesures d'évaluation des classifieurs imprécis pour récompenser la prudence. Soit un objet à classifier \mathbf{x} dont la vraie classe est $\theta_{\mathbf{x}}^*$. La fonction qui quantifie, de façon binaire, la justesse d'un classifieur $\delta_{2\Theta}$ quand il classe \mathbf{x} est notée, acc , et définie par :

$$acc(\theta_{\mathbf{x}}^*, \delta_{2\Theta}(\mathbf{x})) = \begin{cases} 1 & \text{si } \theta_{\mathbf{x}}^* \in \delta_{2\Theta}(\mathbf{x}), \\ 0 & \text{sinon.} \end{cases}$$

Cette généralisation de la justesse n'est pas suffisante pour évaluer une sortie imprécise car elle ne récompense pas la pertinence. En effet, un classifieur trop prudent renvoyant systématiquement la sortie Θ (prudence maximale) obtiendrait toujours la récompense maximale avec la justesse acc alors que ce classifieur ne serait aucunement pertinent.

Dans les sous-sections suivantes, nous présentons trois types de mesures d'évaluation de sorties imprécises : la *justesse* affaiblie qui modifie la mesure de la justesse en affaiblissant les récompenses pour les sorties moins pertinentes ; la mesure F_{β} construite comme un compromis entre la justesse et la justesse affaiblie, et les mesures g_{65} , g_{80} construites comme des fonctions de

la justesse affaiblie permettant d’augmenter la récompense des sorties imprécises plus prudentes.

Ces différentes mesures d’évaluation ont chacune leur façon de quantifier à quel degré les deux contraintes imposées à la classification imprécise, i.e., la prudence et la pertinence, doivent être respectées. Les mesures présentées s’expriment toutes sous la forme d’une fonction $u : \Theta \times 2^\Theta \rightarrow \mathbb{R}$ où $u(\theta_x^*, \delta_{2^\Theta}(\mathbf{x}))$ quantifie la satisfaction procurée par la sortie $\delta_{2^\Theta}(\mathbf{x})$ affectée à \mathbf{x} de vraie classe θ_x^* . La généralisation à l’évaluation d’un classifieur imprécis sur un ensemble de tests $(\mathbf{x}^t, \theta^t)_{1 \leq t \leq T}$, notée $U((\mathbf{x}^t, \theta^t)_{1 \leq t \leq T}, \delta_{2^\Theta})$, est obtenue en prenant la moyenne de toutes les satisfactions obtenues pour chaque exemple de test :

$$U((\mathbf{x}^t, \theta^t)_{1 \leq t \leq T}, \delta_{2^\Theta}) = \frac{1}{T} \sum_{t=1}^T u(\theta^t, \delta_{2^\Theta}(\mathbf{x}^t)).$$

Nous faisons également un lien avec l’évaluation d’une classification hiérarchique qui doit satisfaire d’autres contraintes, qui selon nous, pourraient être profitables pour la classification imprécise.

3.3.1 Justesse affaiblie

Une première façon d’évaluer les sorties imprécises a été empruntée à la classification multi-labels (Tsoumakas et Vlahavas, 2007). Cette mesure d’évaluation appelée *discounted accuracy* en anglais et traduite en français par *justesse affaiblie* dans (Yang, 2016) est une modification de la fonction *acc* permettant de moduler la satisfaction en fonction de l’ensemble des classes retourné. La justesse affaiblie notée $acc_{discounted}$ est définie comme suit :

$$acc_{discounted}(\theta_x^*, \delta_{2^\Theta}(\mathbf{x})) = \begin{cases} \frac{1}{|\delta_{2^\Theta}(\mathbf{x})|} & \text{si } \theta_x^* \in \delta_{2^\Theta}(\mathbf{x}), \\ 0 & \text{sinon.} \end{cases}$$

Ainsi, la justesse affaiblie récompense la pertinence (les sorties sont pénalisées en fonction de leur taille $|\delta_{2^\Theta}(\mathbf{x})|$). Considérant cette mesure d’évaluation dans l’exemple 8, nous pouvons donner les récompenses manquantes de la table 3.1. Nous obtenons la TABLE 3.2.

		Ensemble fourni : $\delta_{2^\Theta}(\mathbf{x})$						Θ
		$\{Covid\}$	$\{Grippe\}$	$\{Angine\}$	$\{Covid, Grippe\}$	$\{Covid, Angine\}$	$\{Grippe, Angine\}$	
Vraie classe : θ_x^*	<i>Covid</i>	1	0	0	1/2	1/2	0	1/3
	<i>Grippe</i>	0	1	0	1/2	0	1/2	1/3
	<i>Angine</i>	0	0	1	0	1/2	1/2	1/3

TABLE 3.2 – $acc_{discounted}$

Bien que raisonnable dans certain cas, une réserve importante concernant l’utilisation de cette mesure d’évaluation a été présentée dans (Zaffalon et al., 2012). Reprenant l’argument avancé dans la dernière référence pour illustrer cette limite avec notre exemple, considérons deux classifieurs ayant les résultats extrêmes suivants :

- Un classifieur aléatoire, i.e., qui renvoie au hasard (de façon uniforme) une classe précise,
- Un classifieur trop prudent au sens qu’il renvoie toujours l’ensemble total Θ .

Dans notre exemple, le classifieur aléatoire correspondrait à un médecin imaginaire donnant des diagnostics précis parmi les maladies *Covid-19*, *Grippe* ou *Angine* au hasard (donc sans aucun fondement). Le classifieur trop prudent correspondrait à un médecin imaginaire trop prudent

qui dirait systématiquement : "*je ne sais pas ce dont vous souffrez*".

Aucun de ces deux classifieurs ne connaît la vraie classe mais ces classifieurs ont des manières opposées de gérer leur ignorance. Comment évaluer de tels classifieurs? En moyenne, la réponse du classifieur aléatoire est exacte $1/3$ du temps sur une base d'exemples équilibrée; la récompense attendue avec la justesse affaiblie serait donc de $1/3$. D'autre part, la récompense du classifieur trop prudent recevrait systématiquement la récompense $1/3$. En choisissant la justesse affaiblie, les deux classifieurs reçoivent donc la même récompense $1/3$. Pourtant le comportement du classifieur trop prudent semble préférable parce qu'il exprime son incapacité à fournir une information précise, ce dont n'est pas capable le classifieur aléatoire. Suivant un raisonnement similaire nous pouvons déduire que la justesse affaiblie récompense un classifieur imprécis exactement comme s'il s'agissait d'un classifieur standard dont la sortie serait choisie aléatoirement dans le sous-ensemble fourni par le classifieur imprécis. Finalement, la justesse affaiblie ne valorise pas la prudence (elle la récompense au même niveau que le hasard).

3.3.2 Mesure d'évaluation F_β

Afin d'évaluer les performances d'un classifieur imprécis en terme de prudence et de pertinence, d'autres mesures d'évaluation ont été introduites telle que la mesure F_β (Coz et al., 2009). Cette mesure d'évaluation initialement développée pour des problème de recherche d'information (RI) permet aussi d'évaluer des sorties sous forme de sous-ensembles de classes. La mesure F_β est construite comme un compromis, paramétré par $\beta \geq 0$, entre la justesse acc (capturant la prudence) et la justesse affaiblie $acc_{discounted}$ (capturant la pertinence). Elle est définie comme suit :

$$\begin{aligned} F_\beta(\theta_{\mathbf{x}}^*, \delta_{2\Theta}(\mathbf{x})) &= \frac{(1 + \beta^2) acc(\theta_{\mathbf{x}}^*, \delta_{2\Theta}(\mathbf{x})) \cdot acc_{discounted}(\theta_{\mathbf{x}}^*, \delta_{2\Theta}(\mathbf{x}))}{\beta^2 acc_{discounted}(\theta_{\mathbf{x}}^*, \delta_{2\Theta}(\mathbf{x})) + acc(\theta_{\mathbf{x}}^*, \delta_{2\Theta}(\mathbf{x}))} \\ &= \frac{1 + \beta^2}{\beta^2 + |\delta_{2\Theta}(\mathbf{x})|} \mathbb{1}_{\delta_{2\Theta}(\mathbf{x})}(\theta_{\mathbf{x}}^*) \end{aligned}$$

Le paramètre β est introduit pour contrôler le niveau de pertinence requis :

- Avec des valeurs faibles de β , la pertinence de la sortie est plus importante dans l'évaluation. En effet plus β est petit plus la taille du sous-ensemble $|\delta_{2\Theta}|$ pénalise la récompense. Le critère de la pertinence est maximal lorsque $\beta = 0$. Dans ce cas nous retrouvons la justesse affaiblie : $acc_{discounted}(\theta_{\mathbf{x}}^*, \delta_{2\Theta}(\mathbf{x})) = \mathbb{1}_{\delta_{2\Theta}(\mathbf{x})}(\theta_{\mathbf{x}}^*)/|\delta_{2\Theta}(\mathbf{x})|$.
- A l'inverse plus β est grand, moins la pertinence prime dans l'évaluation. Lorsque $\beta \rightarrow \infty$, nous retrouvons la justesse définie pour la classification imprécise :

$$F_\beta(\theta_{\mathbf{x}}^*, \delta_{2\Theta}(\mathbf{x})) \xrightarrow{\beta \rightarrow \infty} \mathbb{1}_{\delta_{2\Theta}(\mathbf{x})}(\theta_{\mathbf{x}}^*) = acc(\theta_{\mathbf{x}}^*, \delta_{2\Theta}(\mathbf{x})).$$

Ainsi pour les grandes valeurs de β , seule la prudence de la sortie compte.

Remarque 2. La mesure d'évaluation F_β présentée ici n'a pas de lien avec le F-Score $_\beta$ (chapitre 2) même si les notations et les constructions de ces deux mesures peuvent prêter à confusion. Pour rappel, le F-Score $_\beta$ permet dans l'évaluation d'un classifieur standard, de faire un compromis entre les mesures de rappel et de précision. La mesure F_β définie ici a une construction analogue mais pour faire un compromis entre la justesse et la justesse affaiblie. Les deux mesures justesse et justesse affaiblie sont deux généralisations différentes de la justesse dans la classification standard (il n'y a donc pas de lien avec les mesures de précision et rappel de la classification standard).

		Ensemble fourni : $\delta_{2\Theta}(\mathbf{x})$						
		{Covid}	{Grippe}	{Angine}	{Covid, Grippe}	{Covid, Angine}	{Grippe, Angine}	Θ
Vraie classe : $\theta_{\mathbf{x}}^*$	Covid	1	0	0	$\frac{1 + \beta^2}{2 + \beta^2}$	$\frac{1 + \beta^2}{2 + \beta^2}$	0	$\frac{1 + \beta^2}{3 + \beta^2}$
	Grippe	0	1	0	$\frac{1 + \beta^2}{2 + \beta^2}$	0	$\frac{1 + \beta^2}{2 + \beta^2}$	$\frac{1 + \beta^2}{3 + \beta^2}$
	Angine	0	0	1	0	$\frac{1 + \beta^2}{2 + \beta^2}$	$\frac{1 + \beta^2}{2 + \beta^2}$	$\frac{1 + \beta^2}{3 + \beta^2}$

TABLE 3.3 – F_{β} .

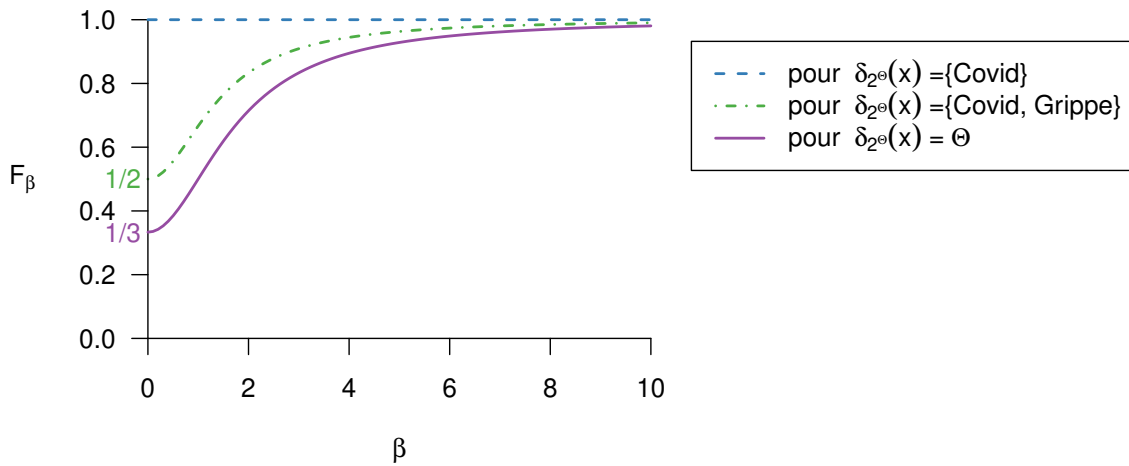


FIG. 3.5 Récompenses fournies par F_{β} en fonction de β .

Lorsque nous utilisons la mesure F_{β} pour évaluer les sorties possibles pour l'exemple 8, nous obtenons la TABLE 3.3.

Exemple 9. Afin d'illustrer les récompenses fournies par F_{β} considérons l'exemple d'un patient \mathbf{x} dont la vraie maladie est la covid-19. La ligne surlignée en jaune dans la table 3.3 donne les récompenses pour chaque sortie imprécise possible. Concentrons nous sur trois cas de sorties qui suffisent à illustrer la manière dont F_{β} donne ses récompenses :

- cas où $\delta_{2\Theta}(\mathbf{x}) = \{\text{Covid}\}$ (exactitude),
- cas où $\delta_{2\Theta}(\mathbf{x}) = \{\text{Covid, Grippe}\}$,
- cas où $\delta_{2\Theta}(\mathbf{x}) = \Theta$ (prudence totale).

La Figure 3.5 montre les récompenses données dans ces trois cas en fonction de β . Nous observons que pour $\beta = 0$, nous retrouvons bien les récompenses de la justesse affaiblie (données dans la Table 3.2) qui récompensent seulement la pertinence. Lorsque β augmente, les sorties prudentes ($\{\text{Covid, Grippe}\}$ et Θ) sont de mieux en mieux récompensées. Pour les grandes valeurs de β ces récompenses se rapprochent de la récompense donnée pour l'exactitude (égale à 1) et c'est la pertinence qui n'est plus récompensée.

La mesure F_{β} est donc bien capable d'évaluer des sorties imprécises en fonction du compromis (déterminé par β) entre la pertinence et la prudence.

3.3.3 Mesures d'évaluation g_{65} et g_{80}

Les auteurs de (Zaffalon et al., 2012) ont proposé une autre façon d'évaluer les sorties imprécises en modifiant la justesse affaiblie. Cette modification consiste à augmenter les récompenses données aux sorties imprécises prudentes. Pour ce faire la récompense, notée $acc_{gen}(\theta_x^*, \delta_{2\Theta}(\mathbf{x}))$, donnée à une sortie $\delta_{2\Theta}(\mathbf{x})$ lorsque la vraie classe est θ_x^* est construite comme une fonction de la récompense donnée par la justesse affaiblie :

$$acc_{gen}(\theta_x^*, \delta_{2\Theta}(\mathbf{x})) = g(acc_{discounted}(\theta_x^*, \delta_{2\Theta}(\mathbf{x}))) = g\left(\frac{\mathbb{1}_{\delta_{2\Theta}(\mathbf{x})}(\theta_x^*)}{|\delta_{2\Theta}(\mathbf{x})|}\right),$$

où g est une fonction définie sur $[0, 1]$ telle que :

- $g(0) = 0$, pour toutes les sorties qui ne contiennent pas la vraie classe θ_x^* , la récompense reste ainsi la même que pour la justesse affaiblie, c'est à dire 0.
- $g(1) = 1$, ce qui permet de récompenser de façon maximale l'exactitude comme le fait déjà la justesse affaiblie.
- $\forall z \in [0, 1]$, $g(z) \geq z$, la récompense avec g est au moins égale à celle donnée par la justesse affaiblie.

La dernière condition permet de récompenser la prudence. Reprenant l'argument présenté dans la sous-section 3.3.1 pour l'exemple du problème de diagnostic entre les maladies *Covid-19*, *Grippe* ou *Angine*, la justesse affaiblie donne la même récompense de $1/3$ au classifieur trop prudent qu'au classifieur aléatoire. Le fait de donner une récompense $g(1/3) \geq 1/3$ à la sortie du classifieur trop prudent permet de valoriser ce classifieur. L'argument étant généralisable à toutes les tailles de sorties imprécises, la contrainte est appliquée à toute les valeurs z de justesse affaiblie.

Les auteurs proposent de choisir comme fonction g un polynôme d'ordre 2 noté g_m tel que :

$$g_m(z) = az^2 + bz,$$

où m caractérise la valeur de la récompense donnée à un sous-ensemble de cardinal 2 incluant θ_x^* , c'est à dire un sous-ensemble qui aurait pour récompense 0.5 avec la justesse affaiblie. Plus précisément, la valeur de m correspond à $g_m(0.5) = m/100$. En particulier les auteurs estiment que $g(0.5)$ pourrait raisonnablement se situer entre 0.65 et 0.8. Ainsi deux mesures alternatives pour évaluer des sorties imprécises peuvent être obtenues par les fonctions :

$$\begin{aligned} g_{65}(z) &= -0.6z^2 + 1.6z, \\ g_{80}(z) &= -1.2z^2 + 2.2z. \end{aligned}$$

Remarquons que $g_{50}(z) = z$ correspond à la justesse affaiblie. La Fig. 3.6 représente les fonctions g_{50} , g_{65} et g_{80} . Les marques aux abscisses $\{0, 1/3, 1/2, 1\}$ correspondent aux valeurs de justesse affaiblie possibles dans notre exemple à 3 classes. Nous observons que g_{80} s'écarte plus de la justesse affaiblie (g_{50}) que g_{65} ce qui traduit une plus grande valorisation de la prudence. La TABLE 3.4 montre la matrice des évaluations obtenue avec g_{65} .

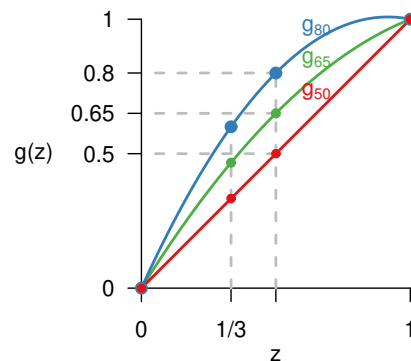


FIG. 3.6 Représentation des fonctions g_{50} , g_{65} et g_{80} inspirée des auteurs (Zaffalon et al., 2012).

		Ensemble fourni : $\delta_{2\Theta}(\mathbf{x})$						Θ
		$\{Covid\}$	$\{Grippe\}$	$\{Angine\}$	$\{Covid, Grippe\}$	$\{Covid, Angine\}$	$\{Grippe, Angine\}$	
Vraie classe : $\theta_{\mathbf{x}}^*$	<i>Covid</i>	1	0	0	0.65	0.65	0	0.467
	<i>Grippe</i>	0	1	0	0.65	0	0.65	0.467
	<i>Angine</i>	0	0	1	0	0.65	0.65	0.467

TABLE 3.4 – Récompenses de la mesure d’évaluation obtenues avec g_{65}

Exemple 10. Dans cet exemple, nous allons vérifier que la mesure définie par g_{65} valorise effectivement la sortie imprécise prudente $\{Covid, Grippe\}$ par rapport à une décision aléatoire qui consiste à donner le diagnostic Covid ou le diagnostic Grippe au hasard. L’objectif ici est de vérifier qu’à connaissance égale, le médecin prudent est valorisé par rapport au médecin donnant un diagnostic au hasard avec cette mesure. Considérons un ensemble de patients pour lesquels les symptômes connus ne permettent pas d’en savoir plus sur le fait qu’ils peuvent être chacun atteint de la Covid ou de la Grippe. Sur cet échantillon de patients, le médecin donnant une réponse au hasard se trompe une fois sur deux si les maladies sont réparties uniformément dans l’échantillon de patients. En moyenne l’attitude de ce médecin est récompensée par $1/2$. Le médecin prudent lui reçoit systématiquement la récompense 0.65 (voir TABLE 3.4, $g_{65}(acc_{discounted}(Covid, \{Covid, Grippe\})) = g_{65}(acc_{discounted}(Grippe, \{Covid, Grippe\})) = 0.65$). De fait, ce classifieur est récompensé par $0.65 > 1/2$ sur cet échantillon de patients. Le médecin prudent est donc bien valorisé par rapport au médecin donnant une réponse au hasard.

3.3.4 Lien avec l’évaluation de classifieurs hiérarchiques

Étant donné les liens entre la classification hiérarchique et la classification imprécise, nous retrouvons dans l’évaluation d’un classifieur hiérarchique certains aspects de l’évaluation d’un classifieur imprécis. Avec la classification hiérarchique (sous-section 3.1.4), la sortie à évaluer est un nœud dans une hiérarchie pré-établie. Afin d’être évaluée, cette sortie doit être comparée à la vraie classe qui est elle aussi un nœud de la hiérarchie. Dans le cas d’exemples partiellement

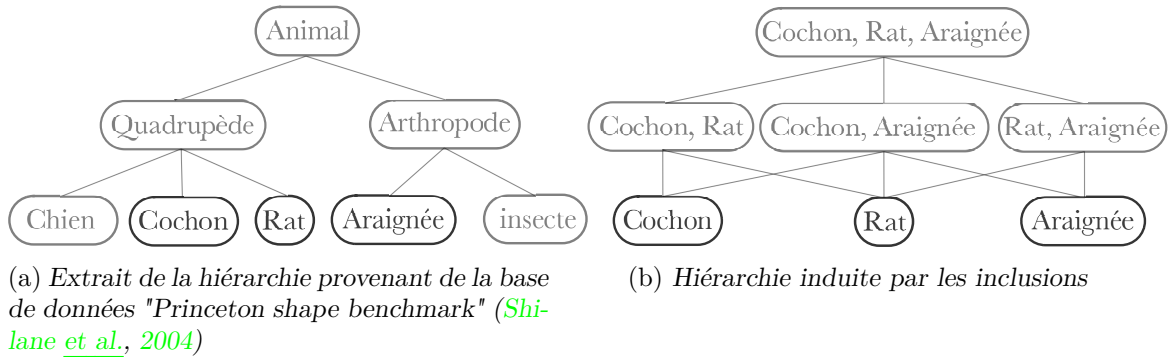


FIG. 3.7 Deux structures de hiérarchie

étiquetés, la vraie classe n'est pas forcément une feuille et peut être n'importe quel nœud de la hiérarchie. L'évaluation d'un classifieur hiérarchique doit tenir compte de plusieurs aspects :

- l'erreur de généralisation (erreur qui consiste à prédire un nœud ancêtre du vrai nœud)
- l'erreur de spécialisation (erreur qui consiste à prédire un nœud descendant du vrai nœud).
- la proximité entre le nœud prédit et le vrai nœud dans la hiérarchie.

L'une des mesures les plus répandues pour évaluer un classifieur hiérarchique est une mesure basée sur un compromis entre le contrôle de la généralisation et de l'erreur de spécialisation. Cette mesure appelée F-mesure hiérarchique, notée hF , est basée sur la précision hiérarchique hP et le rappel hiérarchique hR définis pour le nœud prédit et le vrai nœud comme (Silla et Freitas, 2011) :

$$hP = \frac{|Pred \cap True|}{|Pred|}, \quad (25)$$

$$hR = \frac{|Pred \cap True|}{|True|}, \quad (26)$$

où $Pred$ désigne l'ensemble constitué du nœud prédit et de tous ses ancêtres dans la hiérarchie et $True$ est l'ensemble composé du nœud réel le plus spécifique ainsi que de tous ses ancêtres.

Exemple 11. Considérons la hiérarchie présentée dans la Fig. 3.7a. Par exemple, si la vérité est "Rat" et que la sortie est "Quadrupède", $True$ est l'ensemble composé des nœuds "Rat", "Quadrupède" et "Animal" et $Pred$ est l'ensemble composé des nœuds "Quadrupède" et "Animal". Ainsi $hP = 1$ et $hR = 2/3$.

La TABLE 3.5 présente les valeurs de hP et hR obtenues pour une partie des sorties et vérités possibles. En reprenant une illustration analogue présentée dans (Silla et Freitas, 2011), nous constatons que pour une sortie fixée, prenons "Cochon", plus l'erreur de spécialisation (sortie plus précise que la réalité) est importante plus hP baisse alors que hR reste constant (cellules jaunes dans le tableau). A l'inverse, pour une vérité fixée, prenons "Rat" (voir cellules bleues), plus l'erreur de généralisation (sortie plus imprécise que la réalité) est grande, plus hR diminue alors que hP reste constant.

Finalement hP sanctionne l'erreur de spécialisation et hR sanctionne l'erreur de généralisation. De plus dans leur construction hP et hR prennent en compte la structure de la hiérarchie.

Exemple 12. Considérons maintenant la structure représentée sur la Fig. 3.7b qui est induite par les relations d'inclusion entre les ensembles. Considérons à nouveau l'exemple 11 où la

		nœud prédit					
		Cochon	Rat	Araignée	Quadrupède	Arthropode	Animal
vrai nœud	Cochon	$hP = 1$ $hR = 1$	$hP = 2/3$ $hR = 2/3$	$hP = 2/3$ $hR = 2/3$	$hP = 1$ $hR = 2/3$	$hP = 1/2$ $hR = 1/3$	$hP = 1$ $hR = 1/3$
	Rat	$hP = 2/3$ $hR = 2/3$	$hP = 1$ $hR = 1$	$hP = 2/3$ $hR = 2/3$	$hP = 1$ $hR = 2/3$	$hP = 1/2$ $hR = 1/3$	$hP = 1$ $hR = 1/3$
	Araignée	$hP = 2/3$ $hR = 2/3$	$hP = 2/3$ $hR = 2/3$	$hP = 1$ $hR = 1$	$hP = 1/2$ $hR = 1/3$	$hP = 1$ $hR = 2/3$	$hP = 1$ $hR = 1/3$
	Quadrupède	$hP = 2/3$ $hR = 1$	$hP = 2/3$ $hR = 1$	$hP = 1/3$ $hR = 1/2$	$hP = 1$ $hR = 1$	$hP = 1/2$ $hR = 1/2$	$hP = 1$ $hR = 1/2$
	Arthropode	$hP = 1/3$ $hR = 1/2$	$hP = 1/3$ $hR = 1/2$	$hP = 2/3$ $hR = 1$	$hP = 1/2$ $hR = 1/2$	$hP = 1$ $hR = 1$	$hP = 1$ $hR = 1/2$
	Animal	$hP = 1/3$ $hR = 1$	$hP = 1/3$ $hR = 1$	$hP = 1/3$ $hR = 1$	$hP = 1/2$ $hR = 1$	$hP = 1/2$ $hR = 1$	$hP = 1$ $hR = 1$

TABLE 3.5 – Evaluations hP et hR obtenues considérant la hiérarchie de la Fig. 3.7a.

vérité est "Rat" et la sortie est "Quadrupède" (qui correspond à l'ensemble {"Cochon", "Rat"} dans la hiérarchie considérée), alors True est l'ensemble des nœuds "Rat", {"Cochon", "Rat"}, {"Rat", "Araignée"}, {"Cochon", "Rat", "Araignée"} et Pred est l'ensemble des nœuds {"Cochon", "Rat"} et {"Cochon", "Rat", "Araignée"}. Donc $hP = 1$ et $hR = 1/2$ (nous avons $hP = 1$ et $hR = 2/3$ en considérant la hiérarchie de la Fig. 3.7a).

La mesure hF est un compromis entre la pénalisation de l'erreur de spécialisation et de l'erreur de généralisation. Elle est définie comme la moyenne harmonique de hP et hR (Silla et Freitas, 2011) :

$$hF = \frac{2 hP hR}{hP + hR}.$$

Remarque 3. La mesure hF pourrait être rapprochée de la mesure F_β (définie sous-section 3.3.2) avec $\beta = 1$, lorsque la hiérarchie considérée est induite par les relations d'inclusions entre des ensembles de feuilles et que la vraie classe est une feuille, c'est à dire lorsque nous sommes dans les conditions de la classification imprécise considérées dans cette thèse. La Fig. 3.7b représente un exemple d'une telle hiérarchie. Dans ce cas nous pouvons considérer le problème de classification hiérarchique comme un problème de classification imprécise d'un objet x tel que :

- L'ensemble des classes Θ est l'ensemble des feuilles.
- La sortie imprécise $\delta_{2\Theta}(x)$ est l'ensemble correspondant au nœud prédit.
- La vraie classe θ_x^* est la vraie feuille.

Posons temporairement $A = \delta_{2\Theta}(x)$ pour simplifier les explications. Nous pouvons ré-écrire les mesures hP et hR dans le cadre de la classification imprécise (en fonction de Θ , A et θ_x^*). En effet, $Pred$ est l'ensemble des nœuds constitué du nœud prédit et de tous ses ancêtres donc dans notre cas,

$$\begin{aligned} Pred &= \{B \subseteq \Theta, B \supseteq A\}, \\ &= \{A \cup C \mid C \subseteq \Theta, C \subseteq A^c\}. \end{aligned}$$

Nous en déduisons que $|Pred| = 2^{|A^c|} = 2^{|\Theta|-|A|}$. Par ailleurs, $True$ est l'ensemble composé du nœud réel le plus spécifique et de tous ses ancêtres. Dans notre cas nous obtenons

$$\begin{aligned} True &= \{B \subseteq \Theta, B \ni \theta_x^*\}, \\ &= \{\{\theta_x^*\} \cup C \mid C \subseteq \Theta, C \subseteq \{\theta_x^*\}^c\}. \end{aligned}$$

Nous en déduisons $|True| = 2^{|\{\theta_x^*\}^c|} = 2^{|\Theta|-|\{\theta_x^*\}|}$. De même,

$$\begin{aligned} Pred \cap True &= \{B \subseteq \Theta, B \supseteq A \cup \{\theta_x^*\}\}. \\ &= \{A \cup \{\theta_x^*\} \cup C \mid C \subseteq \Theta, C \subseteq (\{\theta_x^*\} \cup A)^c\}. \end{aligned}$$

d'où $|Pred \cap True| = 2^{(|\{\theta_x^*\} \cup A|^c)} = 2^{|\Theta|-|\{\theta_x^*\} \cup A|}$.

En remplaçant dans les définitions (25)-(26) nous obtenons $hP = 2^{|A|-|\{\theta_x^*\} \cup A|}$ et $hR = 2^{|\theta_x^*|-|\{\theta_x^*\} \cup A|}$. Reprenons la notation $\delta_{2\Theta}(\mathbf{x})$ au lieu de A . Finalement puisque nous considérons θ_x^* comme un singleton nous avons :

$$hP = \begin{cases} 1 & \text{si } \theta_x^* \in \delta_{2\Theta}(\mathbf{x}), \\ 1/2 & \text{sinon.} \end{cases}$$

et

$$hR = \begin{cases} 2^{1-|\delta_{2\Theta}(\mathbf{x})|} & \text{si } \theta_x^* \in A, \\ 2^{-|\delta_{2\Theta}(\mathbf{x})|} & \text{sinon,} \end{cases}$$

Dans le cas général de la classification hiérarchique, hP contrôle l'erreur de spécialisation mais dans le cas où le vrai nœud est une feuille, cette erreur ne peut plus être commise. Nous constatons que hP est ici réduite à une forme de récompense de la prudence (récompense maximale 1 dès que la vraie classe θ_x^* appartient à l'ensemble prédit $\delta_{2\Theta}(\mathbf{x})$). En ce sens, hP pourrait dans ce cas être rapproché de la justesse ($acc(\theta_x^*, \delta_{2\Theta}(\mathbf{x})) = \mathbb{1}_{\delta_{2\Theta}(\mathbf{x})}(\theta_x^*)$). La différence entre acc et hP étant que hP donne une récompense non nulle même égale à $1/2$, lorsque $\theta_x^* \notin A$. Cette différence n'a pas de réelle importance puisque le cas $\theta_x^* \notin A$ est toujours pénalisé par rapport au cas $\theta_x^* \in A$. D'autre part, hR (qui pénalise l'erreur de généralisation) pourrait être rapproché de la justesse affaiblie qui récompense la pertinence ($acc_{discounted}(\theta_x^*, \delta_{2\Theta}(\mathbf{x})) = \mathbb{1}_{\delta_{2\Theta}(\mathbf{x})}(\theta_x^*)/|\delta_{2\Theta}(\mathbf{x})|$). En effet plus le cardinal de l'ensemble prédit $\delta_{2\Theta}(\mathbf{x})$ est grand, moins hR récompense $\delta_{2\Theta}(\mathbf{x})$. La différence notable avec $acc_{discounted}$ est que hR donne des récompenses non-nulles lorsque $\theta_x^* \notin \delta_{2\Theta}(\mathbf{x})$. Cette fois cette différence est en contradiction avec les attentes de la classification imprécise, car nous pouvons obtenir des valeurs de hR plus élevées pour certaines erreurs que pour certaines prédictions certes imprécises mais justes.

Exemple 13. Par exemple, prenons $\Theta = \{Cochon, Rat, Araignée\}$, et supposons que la vraie classe est $\{Cochon\}$. Alors la récompense donnée par hR pour la prédiction $\{Cochon, Rat, Araignée\}$ est $2^{1-3} = 0.25$ alors que la prédiction fausse mais précise $\{Rat\}$ est mieux récompensée par $2^{-1} = 0.5$.

Dans cette sous-section, nous avons montré les différences entre les exigences de la classification hiérarchique et celles de la classification imprécise. La notion d'erreur de spécialisation développée pour l'évaluation de classifieur hiérarchique, pourrait être utilisée dans le cas de la classification imprécise pour pénaliser les sorties plus précises que ne le permet l'information disponible. Nous détaillons cette idée dans la prochaine section.

Remarque 4. Dans cette section nous avons présenté des mesures d'évaluation basées sur des récompenses unitaires³. Certaines mesures d'évaluation (Abellan et Masegosa, 2012; Yang

3. Une récompense unitaire consiste à donner récompense maximale égale à 1 en cas d'exactitude et récompense minimale égale à 0 en cas d'erreur.

[et al., 2016](#)) sont capables de prendre en compte des récompenses qui ne sont pas unitaires. Les récompenses non unitaires sont appropriées lorsque certaines erreurs ont des conséquences plus graves que d'autres.

Exemple 14. *Reprenant l'exemple 1, l'erreur de diagnostic consistant à diagnostiquer "angine" alors que le patient est atteint de la "covid-19" a des conséquences importantes (contaminations, mauvais suivi, ...). A l'inverse, l'erreur de diagnostic consistant à prédire "covid-19" au lieu d'"angine" a des conséquences moins importantes (mise en quarantaine inutile, ...). La première erreur devrait donc être plus pénalisée que la seconde.*

Dans le problème industriel qui motive cette thèse, nous sommes également dans cette situation : certaines erreurs de tri de plastiques engendrent des impuretés dans les bacs triés plus ou moins néfastes selon le type de plastique. Cependant, nous délaissions cet aspect du problème pour le moment et nous concentrons sur des mesures basées sur des récompenses unitaires. En effet comme nous l'avons mentionné dans l'introduction de ce chapitre, la classification imprécise est dans cette thèse une étape intermédiaire. L'intégration des différentes conséquences des erreurs de tri est traitée plus tard lors de la résolution du problème de tri en ligne (chapitre 5).

3.4 Positionnement

De nombreuses approches de classification imprécise ont été proposées ces dernières décennies. Ces approches ont été proposées dans plusieurs théories de l'incertain et sont généralement basées sur une des trois familles de règles de décision que nous appelons :

- **Les règles de dominance** : lorsque la sortie imprécise est l'ensemble des classes non-dominées d'après un ordre partiel inféré.
- **Les règles de minimisation du risque** : lorsque la sortie imprécise résulte d'une minimisation de risque (de manière similaire d'une maximisation d'espérance de gain).
- **Les règles de seuil** : lorsque la sortie imprécise est déterminée par le choix d'un seuil.

La TABLE 3.6 présente des références d'articles proposant des approches de classification imprécise basées sur des théories de l'incertain classées par familles de règles de décision.

Historiquement, les théories les plus utilisées dans la littérature pour faire de la classification imprécise sont les probabilités et les probabilités imprécises. Cependant certains travaux ne reposant pas sur ces deux théories ont été proposés. Notamment la théorie des fonctions de croyance offre un cadre propice à la classification imprécise car contrairement à la théorie des probabilités ou la théorie des probabilités imprécises une fonction de masse quantifie directement la croyance pour qu'un ensemble de classes candidates soit pertinent pour contenir la vraie classe. Dans le cas de la théorie des probabilités ou la théorie des probabilités imprécises, la chance pour qu'un ensemble de classes candidates soit pertinent pour contenir la vraie classe est déduite de la chance de chacune des classes.

Dans un cadre applicatif, le choix du classifieur imprécis peut notamment reposer sur les critères suivants :

- **La prise en compte de l'information** : Suivant les théories de l'incertain sous-jacentes, les types d'imperfections des données prises en compte peuvent ne pas correspondre aux imperfections des données réelles.

		Règles de décision		
		seuil	risque	dominance
Avec Modèle	Probabilités	(Gupta, 1965)	(Ha, 1997) (Coz et al., 2009) (Mortier et al., 2019)	
	Probabilités Imprecises			(Zaffalon, 1999) (Abellán et Moral, 2005) (Corani et Zaffalon, 2008) (Corani et Zaffalon, 2009) (Moral et al., 2020)
	Possibilités			(Nguyen et al., 2018)
	Fonctions de croyance	(Liu et al., 2014)	(Jacquin et al., 2019)	
	Logique floue	(Ishibuchi et Nii, 2000)		(Hüllermeier et Brinker, 2008)
Sans Modèle		(Zhang et al., 2017)		

TABLE 3.6 – Exemples d’articles proposant des approches de classification imprécise (répartis en fonction du cadre théorique et du type de décision utilisés)

- **La complexité** : Certaines applications nécessitent des prédictions rapides ou sont contraintes par l’espace de stockage disponible. Dans ces cas la complexité devrait être minimisée.
- **La flexibilité** : Le compromis entre les exigences de prudence et les exigences de pertinence peuvent varier suivant l’application. De ce point de vue une approche de classification imprécise capable de s’adapter au compromis visé serait intéressante.
- **La généralité** : Une méthode générique s’adapte à un grand nombre d’algorithmes existants, facilite leur utilisation. De plus, toute évolution positive d’un classifieur existant peut être bénéfique au classifieur imprécis s’il est générique.
- **La performance** : Enfin, la performance d’un classifieur imprécis est également un critère important. Celle-ci est subjective car elle dépend du compromis souhaité par le décideur entre pertinence et prudence. Le choix de la mesure d’évaluation (section 3.3) doit donc être en accord avec le compromis attendu.

Donnons une analyse des deux catégories de classifieurs imprécis les plus présentes dans la littérature vis-à-vis de ces critères.

La première catégorie est constituée des approches de classification imprécise basées sur la théorie des probabilités qui utilisent une règle de décision de type **minimisation du risque**. Avec ce type d’approches, l’information prise en compte qui provient uniquement de probabilités a posteriori sur les classes de Θ ne permet pas de représenter l’imprécision (chapitre 2) mais la complexité de ces classifieurs imprécis est quasi-linéaire ($\mathcal{O}(n \log n)$) en fonction du nombre de classes n . Ces classifieurs sont également flexibles car il est possible d’adapter la fonction perte ou de gain aux exigences de l’application. Enfin, ces classifieurs imprécis ont l’avantage d’être génériques puisque la probabilité a posteriori qu’ils utilisent peut être fournie par n’importe quel classifieur probabiliste.

La seconde catégorie de classifieurs imprécis très présente dans la littérature est constituée des approches basées sur les probabilités imprécises et qui prennent leur décision avec des règles de type **dominance**. Contrairement aux approches basées sur les probabilités, ces classifieurs

imprécis prennent en compte l'incertitude et l'imprécision. Cependant la complexité de ces approches est plus importante que celle des approches probabilistes (quadratique en fonction de n pour le classifieur NCC par exemple). D'autre part, ces approches peuvent être flexibles en choisissant une fonction perte ou de gain appropriée. Finalement ces approches sont peu génériques (nécessité d'inférer un ensemble crédal).

Dans cette thèse nous considérons les avantages de construire un classifieur imprécis basé sur les fonctions de croyances. En effet en assignant de la masse à un sous-ensemble de classes, une fonction de masse fournit directement une information concernant une sortie imprécise possible. Cette information est directement liée à l'objectif de la classification imprécise. Pour rappel cet objectif consiste à fournir le sous-ensemble A contenant la vraie classe θ_x^* , le plus petit possible étant donnée l'information disponible. D'autre part une fonction de masse assigne à un sous-ensemble A , une masse $m(A)$ qui peut être interprétée comme la probabilité qu'on ne sache rien de plus que " $\theta_x^* \in A$ ". Pour mieux illustrer le lien entre une fonction de masse et l'objectif de la classification imprécise, considérons un exemple tel que $m(A) = 1$. Cette fonction de masse indique que nous ne savons rien de plus que " $\theta_x^* \in A$ " avec une masse de 1. Cette information suffit pour répondre à l'objectif de la classification imprécise : la sortie imprécise doit être le sous-ensemble A .

De plus, une notion rencontrée dans l'état de l'art nous paraît intéressante pour mesurer la satisfaction d'une sortie imprécise. Il s'agit de la notion d'*erreur de spécialisation*. Dans le contexte de la classification hiérarchique, cette notion concerne les nœuds prédits plus spécifiques que le vrai nœud (le plus spécifique connu). Dans une approche de classification imprécise, cette erreur pourrait s'appliquer aux prédictions plus précises que l'information ne le permet. Dans le cadre des fonctions de croyances, lorsque la décision est basée sur une minimisation de risque, la fonction perte L associée au risque est classiquement définie pour une réalité précise $\theta \in \Theta$ et une sortie $A \subseteq \Theta$. Cependant nous pourrions considérer une fonction perte plus générale qui quantifierait la perte encourue si un ensemble A était prédit alors que l'information disponible permettait seulement de savoir que la vraie classe appartenait à un ensemble B plus grand. Nous pourrions définir une nouvelle règle de décision basée sur une minimisation de risque en tenant compte de cette erreur.

3.5 Conclusion du chapitre

Dans notre étude qui concerne le tri de plastiques à des fins de recyclage, nous nous intéressons à intégrer l'imperfection des données dans leur représentation afin de proposer une procédure de tri prudente. La classification imprécise (catégorie de classification prudente) comme étape intermédiaire (ou étape naïve) dans cette thèse nous intéresse pour les raisons qui suivent. La première raison est que le tri des plastiques est un problème très sensible aux erreurs de tri (les impuretés ont des conséquences importantes). Ensuite, nous sommes en présence de forte imperfection des données (incertitude et imprécision). Enfin les sorties imprécises peuvent être exploitées dans l'application en ajoutant des bacs de tri supplémentaires dédiés à des familles de plastiques par exemple.

Dans ce chapitre, nous avons introduit le principe et les attentes de la classification imprécise. Nous avons ensuite proposé un état de l'art des approches de classification imprécise existantes ainsi que les principales mesures d'évaluation adaptées aux sorties imprécises.

Parmi les principales approches de classification imprécise, les approches probabilistes ont l'avantage d'être génériques mais ne prennent pas compte de l'imprécision potentielle existante dans les données. À l'inverse les probabilités imprécises sont peu génériques, mais prennent en compte l'imprécision.

Les fonctions de croyance semblent être un cadre de l'incertain alternatif intéressant. En effet, en quantifiant directement la masse de probabilité des ensembles, et non pas la probabilité rapportée par les éléments qu'il contient, nous devrions améliorer la prise en compte des imperfections des données et donc optimiser les sorties en fonction de l'information disponible.

Nous nous intéressons à la possibilité de construire un classifieur imprécis basé sur les fonctions de croyance, également pour des raisons liées au contexte applicatif de cette thèse. Premièrement les données disponibles pour effectuer le tri des plastiques sont entachées d'incertitude et d'imprécision. Ce type d'imperfection des données est pris en compte dans le cadre des fonctions de croyance. Ensuite, les fonctions de croyance offrent un cadre pour la fusion d'information qui est nécessaire pour le traitement d'un fragment de plastique constitué d'un ensemble de pixels. Enfin, dans l'approche globale de cette thèse, nous devons respecter des contraintes sur des proportions d'impuretés dans les bacs de tri. Nous gardons une information sous la forme d'une fonction de croyance pour les deux raisons évoquées précédemment. Dans la procédure de tri, les informations sous la forme de fonctions de masse permettent de suivre l'évolution de notre connaissance sur la composition des bacs de tri et ainsi de respecter les contraintes sur les proportions d'impuretés.

Bien que les quatre types de plastiques considérés dans cette thèse soient organisés en deux familles (*polyoléfines* et *styreniques*), nous n'exploitons pas de structure hiérarchique dans cette thèse. En effet, puisque la structure hiérarchique est très petite, les avantages de son exploitation seraient très limités. Nous nous distinguons donc des classifieurs hiérarchiques. Notons tout de même que la notion d'erreur de spécialisation de la classification hiérarchique peut avoir son utilité dans le processus de décision d'un classifieur imprécis. En effet un classifieur imprécis doit éviter de fournir un ensemble plus précis que les informations disponibles ne le permettent.

Dans le prochain chapitre ces dernières considérations sont développées et une approche de classification imprécise générique basée sur les fonctions de croyance est présentée.

 Proposition d'un classifieur imprécis évidentiel

Sommaire

4.1 Proposition d'une approche de classification imprécise évidentielle	90
4.1.1 Construction générique d'une fonction de masse via un ré-étiquetage imprécis	91
4.1.2 Choix d'une sortie imprécise à partir d'une fonction de masse et d'un critère de décision	97
4.2 Discussion	99
4.3 Illustration de l'approche proposée ECLAIR	100
4.4 Performances d'ECLAIR par rapport à l'état de l'art	106
4.5 Conclusion du chapitre	109

Dans le chapitre précédent, nous avons montré l'intérêt de la classification imprécise quand les données sont imprécises et/ou incertaines et détaillé les raisons d'effectuer une classification imprécise pour le problème de tri de plastiques considéré dans cette thèse. Plusieurs classifieurs imprécis ont été proposés dans la littérature basés sur la théorie des probabilités, la théorie des probabilité imprécises, des fonctions de croyances, de la logique floue, etc. Nous avons identifié deux courants importants de classifieurs imprécis : les classifieurs imprécis probabilistes basés sur une décision par minimisation (resp. maximisation) de risque (resp. espérance de gain) ([Ha, 1997](#); [Coz et al., 2009](#); [Mortier et al., 2019](#)), et les classifieurs imprécis proposés dans le cadre des probabilités imprécises, basés sur des décisions par règle de dominance ([Zaffalon, 1999](#); [Abellán et Moral, 2005](#); [Corani et Zaffalon, 2008, 2009](#); [Moral et al., 2020](#)). Nous avons constaté une limite pour chacun de ces types de classifieurs imprécis. Concernant les classifieurs imprécis basés sur la théorie des probabilités, l'imprécision des données n'est pas prise en compte. D'autre part, nous constatons que les classifieurs imprécis basés sur les probabilités imprécises sont peu génériques, dans le sens où, il est difficile d'adapter un classifieur standard existant dans ce cadre. Nous avons conclu que le cadre des fonctions de croyance est propice à la classification imprécise, puisque les masses sont directement assignées à des sous-ensembles (candidats de sorties imprécises). De plus, d'un point de vu applicatif, les fonctions de croyance nous offrent des outils pour combiner les informations provenant de chacun des pixels d'un fragment de plastique.

Dans ce chapitre, nous présentons notre approche de classification imprécise basée sur la théorie des fonctions de croyance. Le classifieur imprécis proposé a été conçu de façon à bénéficier

des avantages du cadre évidentiel et de la panoplie des classifieurs probabilistes standards. Pour ce faire, nous avons proposé un moyen de construire une version évidentielle pour n'importe quel classifieur probabiliste existant.

Cette proposition est basée sur une phase d'explicitation de l'imprécision des données via un ré-étiquetage imprécis et une phase de décision où nous avons proposé un critère de décision pour prédire un ensemble à partir d'une fonction de masse a posteriori. Ce critère est inspiré de l'erreur de spécialisation de la classification hiérarchique et vise à empêcher une prédiction plus précise que l'observation/information disponible le permet. Nous nous référons, dans la suite, à l'approche proposée de classifieur évidentiel et imprécis par le nom de ECLAIR (Evidential CLAssification of Incomplete data via Imprecise Relabelling).

Ce chapitre est organisé de la façon suivante. La section 4.1 est dédiée à la présentation de l'approche proposée ECLAIR. Puis dans la section 4.2, nous discutons des travaux proches de notre proposition. Dans la section 4.3, une illustration de l'approche sur un problème de classification simulé est donnée. Enfin, dans la section 4.4, nous présentons les résultats de l'approche de classification imprécise proposée sur des jeux de données provenant du site *UCI Machine Learning Repository* et sur un jeu de données provenant de l'entreprise Pellenc ST, et nous comparons les résultats obtenus avec ceux de deux classifieurs imprécis de l'état de l'art représentant les deux catégories de classifieurs introduits précédemment.

4.1 Proposition d'une approche de classification imprécise évidentielle

L'approche de classification imprécise évidentielle proposée est destinée à résoudre les problèmes de classification supervisée où les données sont incertaines et/ou imprécises. Nous considérons les zones de chevauchement et les zones avec peu d'exemples comme deux causes possibles d'imprécision et/ou d'incertitude concernant la classe d'un exemple à classifier (chapitre 2, section 2.3).

L'approche est basée sur trois idées principales : 1) l'explicitation de l'imprécision dans la base de données par le ré-étiquetage ; 2) la construction d'une fonction de masse, pour un nouvel exemple, à partir d'une base d'apprentissage partiellement étiquetée¹ ; 3) la construction d'un critère de décision pour la sortie imprécise.

L'approche se construit de la façon suivante. Considérons une base d'apprentissage de L exemples $(\mathbf{x}^l, \theta^l)_{1 \leq l \leq L}$ où pour tout $l, 1 \leq l \leq L$, $\mathbf{x}^l \in \mathcal{X}$ et $\theta^l \in \Theta = \{\theta_1, \dots, \theta_n\}$. Dans un premier temps, chaque objet \mathbf{x}^l est ré-étiqueté par le plus petit sous-ensemble de Θ contenant θ^l et exprimant l'imprécision de ses caractéristiques, i.e., chevauchement ou isolement dans la base d'apprentissage. Nous notons par $\mathbb{A} \subseteq 2^\Theta$ l'ensemble des nouvelles étiquettes.

Dans un second temps, une fonction de masse est construite en utilisant des outils classiques d'apprentissage automatique. Notre approche consiste à utiliser la faculté d'un classifieur probabiliste δ à assigner une masse de probabilité (classiquement sur Θ), pour quantifier une fonction de masse (probabilité sur 2^Θ). Nous modifions les étiquettes des exemples utilisés pour l'apprentissage du classifieur probabiliste δ de telle sorte que lorsqu'on applique δ à un nouvel exemple \mathbf{x} , nous obtenons une fonction de masse $m(\cdot|\mathbf{x})$ définie sur 2^Θ au lieu d'une probabilité $p(\cdot|\mathbf{x})$ définie sur Θ . Plus précisément, nous manipulons l'espace de discernement considéré par δ : au lieu de considérer des classes précises de Θ , nous entraînons δ sur des sous-ensembles de Θ . Les relations d'inclusion pouvant exister entre les éléments de \mathbb{A} sont ignorées par δ . Donc lorsque δ est appliqué à x , la masse de chaque sous-ensemble est assignée indépendamment des autres. Notons qu'en pratique, nous ne pouvons obtenir des éléments focaux que dans l'ensemble

1. ici, un exemple partiellement étiqueté est un exemple étiqueté par un sous-ensemble de Θ

des nouvelles étiquettes \mathbb{A} . Ainsi limiter la taille de \mathbb{A} permet de manipuler moins d'éléments focaux.

Enfin, nous proposons le critère de décision utilisé pour le choix d'une sortie imprécise, i.e., sous forme d'ensemble de classes, basée sur une minimisation de risque tenant compte de l'erreur de spécialisation (lorsqu'une sortie est plus précise que ce que l'information disponible n'autorise à conclure).

La section est organisée comme suit. D'abord, dans la section 4.1.1, nous verrons comment l'approche proposée manipule l'apprentissage classique d'un classifieur probabiliste pour construire une fonction de masse via un ré-étiquetage imprécis. Ensuite, nous verrons dans la sous-section 4.1.2, notre proposition de choix de la sortie imprécise.

4.1.1 Construction générique d'une fonction de masse via un ré-étiquetage imprécis

Dans cette sous-section, nous détaillons la construction d'une fonction de masse par l'approche proposée. Comme nous l'avons introduit précédemment, cette construction repose sur une manipulation de l'espace de discernement d'un classifieur probabiliste.

Afin de réaliser la manipulation, les exemples d'apprentissage dont le vecteur de caractéristiques et/ou la classe est incertaine et/ou imprécise doivent être ré-étiquetés en fonction de leur niveau d'imprécision. Cette phase qui correspond au ré-étiquetage est, aussi considérée comme une phase d'explicitation de l'imprécision des données dans la suite du chapitre. Avant de détailler les méthodes, pour expliquer cette idée de l'explicitation de l'imprécision nous proposons une illustration dans l'exemple 1.

Exemple 1. *La Fig. 4.1 illustre, sur un exemple fictif de classification avec l'ensemble des classes $\Theta = \{\theta_1, \theta_2, \theta_3\}$, la manipulation des exemples d'apprentissage. Lorsqu'un exemple d'apprentissage \mathbf{x}^l , $l \in \{1, \dots, L\}$ étiqueté θ_1 est dans une zone de chevauchement des classes θ_1 et θ_2 , alors l'exemple \mathbf{x}^l doit être ré-étiqueté comme $\{\theta_1, \theta_2\}$. En effet, comme ce chevauchement peut induire une mauvaise discrimination entre $\{\theta_1\}$, $\{\theta_2\}$ dans la phase d'apprentissage, nous avons préféré que le classifieur apprenne que \mathbf{x}^l n'est ni un bon exemple de θ_1 ni de θ_2 . Dans cet exemple fictif, trois nouvelles classes se forment ($\{\theta_1, \theta_2\}$, $\{\theta_1, \theta_3\}$ et Θ) après la phase d'explicitation de l'imprécision. En haut de la figure 4.1, le niveau de discernement est indiqué pour les deux états de la base de données; en dessous la répartition des exemples d'apprentissage est montrée pour les deux états de la base de données (avant et après le ré-étiquetage imprécis).*

L'explicitation de l'imprécision par un expert n'est pas toujours possible (tâche laborieuse en cas de très nombreux exemples). C'est pour cette raison que nous avons envisagé plusieurs alternatives de ré-étiquetage basées sur des outils d'apprentissage automatique. Notons que nous considérons que l'imprécision dans le problème de classification peut provenir de zone de chevauchement et/ou de zone avec peu d'exemples.

Comme cela a été illustré dans l'exemple 1, la phase d'explicitation de l'imprécision consiste à ré-étiqueter les exemples d'apprentissage sujets à de l'imprécision. Plus précisément, un exemple d'apprentissage représenté par ses caractéristiques \mathbf{x}^l initialement étiqueté $\theta^l \in \Theta$ devrait être ré-étiqueté par le sous-ensemble $A^l \subseteq \Theta$ de plus petit cardinal possible selon le principe de "l'engagement minimal", *least commitment* en anglais. Puisque que nous faisons l'hypothèse que les étiquettes initiales sont fiables, il semble raisonnable que les nouvelles classes imprécises contiennent les classes initiales : $A^l \ni \theta^l$. Autrement dit, la nouvelle classe est cohérente avec la classe d'origine. La Fig. 4.2 montre la détection des exemples à ré-étiqueter dans des plus grands sous-ensembles dans l'exemple 1. Sur la figure, ces exemples (entourés en rouge) se trouvent sur des zones de chevauchement des classes, et donc ne sont pas de "bons" représentants de

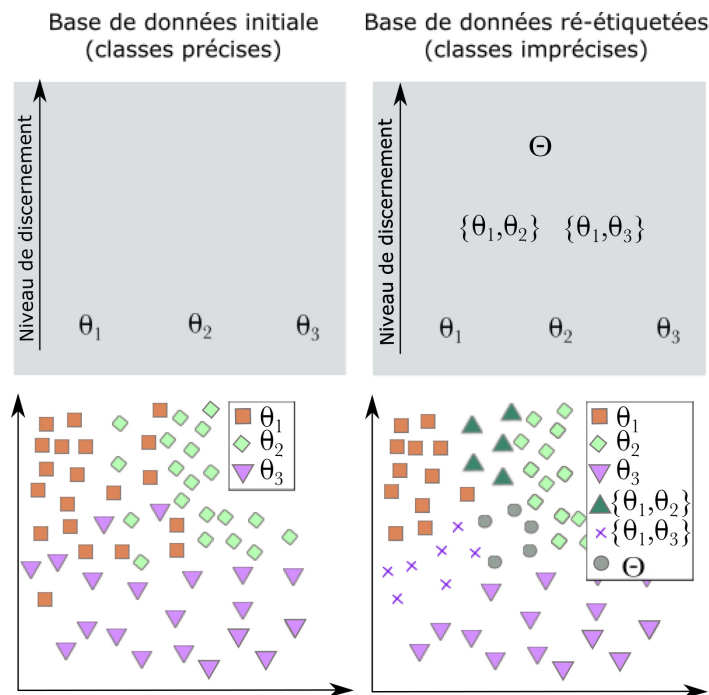


FIG. 4.1 Manipulation de la base d'apprentissage pour un exemple fictif

leur classe. Une fois que les exemples difficiles sont identifiés, il faut faire le choix des nouvelles étiquettes à leur associer.

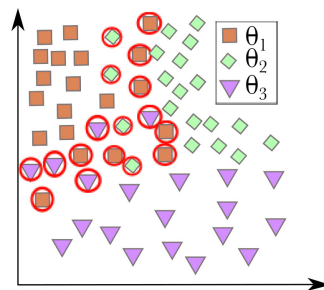


FIG. 4.2 Détection des exemples à ré-étiqueter

Nous avons considéré quatre méthodes pour effectuer l'explicitation de l'imprécision via un ré-étiquetage imprécis. La première méthode est d'adapter une procédure de ré-étiquetage classique, i.e., changement d'étiquette ou suppression des exemples, en procédure de ré-étiquetage imprécis. La seconde méthode se situe dans les méthodes d'"ensemble learning" et elle consiste à séparer chaque exemple de la base d'apprentissage et de le considérer comme un exemple à classifier, puis faire voter un ensemble de classifieurs standards pour lui attribuer un ensemble de classes candidates. La troisième méthode considérée est d'utiliser un classifieur imprécis pour effectuer cette tâche. La quatrième, est similaire à la deuxième méthode et à la troisième dans le principe mais au lieu de faire voter des classifieurs, ou de prendre la sortie d'un classifieur imprécis, elle s'appuie sur un calcul de l'entropie associée à une probabilité a posteriori sur l'ensemble des classes.

Dans la suite nous donnons pour ces quatre méthodes, plus de détails sur la façon dont les exemples à ré-étiqueter sont détectés puis sur le choix de la nouvelle étiquette :

- **Adaptation d'un ré-étiquetage classique** : la première méthode envisagée pour effectuer le ré-étiquetage imprécis est d'adapter une méthode existante de ré-étiquetage. Dans la plupart des études, le ré-étiquetage est effectué dans des situations de classes bruitées (sujette à de l'incertitude) alors que nous considérons le cas inverse (les classes sont fiables et l'imprécision et l'incertitude sont dans les caractéristiques). Cependant certaines stratégies proposées peuvent convenir aux deux situations. Par exemple, comme nous l'avons mentionné dans le chapitre 2, sous-sous-section 2.2.4.2, (Muhlenbach et al., 2004) propose une stratégie de filtrage qui identifie les exemples "suspects" pour améliorer la performance de généralisation de l'algorithme d'apprentissage.

Une façon simple d'adapter ce ré-étiquetage en ré-étiquetage imprécis serait de conserver la classe initiale de \mathbf{x}^l, θ^l , et d'y adjoindre une nouvelle étiquette, notée θ^{maj} , comme proposé dans l'approche de (Muhlenbach et al., 2004). Ainsi, nous obtenons le ré-étiquetage imprécis suivant :

$$A^l = \begin{cases} \{\theta^l, \theta^{maj}\} & \text{si } \mathbf{x}^l \text{ est un exemple suspect,} \\ \{\theta^l\} & \text{sinon.} \end{cases}$$

- **Techniques d'apprentissage ensembliste** : la deuxième méthode envisagée est de considérer que les exemples dont les caractéristiques sont sujettes à de l'imprécision, sont les plus difficiles à classer par des classifieurs standards. Sous cette hypothèse, le ré-étiquetage d'un exemple d'apprentissage est effectué quand une majorité de classifieurs s'accordent à lui donner une autre étiquette. Considérons un ensemble de \mathcal{C} classifieurs standards $\delta_{\Theta}^1, \dots, \delta_{\Theta}^{\mathcal{C}}$ où pour chaque classifieur $\delta_{\Theta}^c : \mathcal{X} \rightarrow \Theta, c \in \{1, \dots, \mathcal{C}\}$ nous calculons des prédictions de validation croisée (CV) pour les données d'apprentissage $(\mathbf{x}^l, \theta^l)_{l=1, \dots, L}$. Le ré-étiquetage de l'exemple \mathbf{x}^l initialement étiqueté par la classe θ^l est basé sur la procédure suivante de vote des prédictions CV des \mathcal{C} classifieurs :

- Si plus de 50% des classifieurs prédisent une même classe θ^{maj} , l'exemple est ré-étiqueté comme l'union $A^l = \{\theta^l, \theta^{maj}\}$. Notons que si $\theta^{maj} = \theta^l$ la classe d'origine est conservée, c'est à dire, $A^l = \{\theta^l\}$.
- Si aucune des classes prédites par les \mathcal{C} classifieurs n'obtient la majorité, alors l'ignorance est privilégiée et l'exemple est ré-étiqueté comme suit : $A^l = \Theta$.

Nous pouvons résumer cette procédure de ré-étiquetage par :

$$A^l = \begin{cases} \{\theta^l, \theta^{maj}\} & \text{si plus de 50\% des classifieurs s'accordent sur une classe } \theta^{maj} , \\ \Theta & \text{sinon.} \end{cases}$$

Le fait que plusieurs classifieurs (\mathcal{C}) soient utilisés pour caractériser l'imprécision permet une meilleure objectivité quant à l'évaluation de l'imprécision réelle des caractéristiques. En effet, le ré-étiquetage est effectué quand l'exemple n'est pas seulement difficile pour un seul classifieur. Avec cette proposition nous limitons le cardinal maximal des nouvelles classes à 2 éléments, sauf en cas d'ignorance. Ainsi nous évitons de créer trop de nouvelles classes ce qui pourrait déséquilibrer trop fortement la base d'apprentissage, et induire de trop nombreux éléments focaux. Cependant, un ré-étiquetage plus général pourrait être facilement envisagé.

- **Utilisation d'un classifieur imprécis** : la troisième méthode envisagée consiste à faire appel à un classifieur imprécis existant $\delta_{2\Theta}$ en validation croisée. Par exemple avec la technique CV, nous pourrions ré-étiqueter l'exemple d'apprentissage difficile (\mathbf{x}^l, θ^l)

par la procédure suivante. D'abord le classifieur imprécis $\delta_{2\Theta}$ est entraîné sur la base d'apprentissage privée de l'exemple (\mathbf{x}^l, θ^l) , puis ce classifieur est appliqué à \mathbf{x}^l et nous obtenons une prédiction sous la forme d'un ensemble $A^{cv} \subseteq \Theta$. Enfin, l'exemple est ré-étiqueté comme

$$A^l = A^{cv} \cup \{\theta^l\},$$

afin d'assurer la cohérence avec l'étiquette initiale.

- **Utilisation de l'entropie :** la dernière méthode envisagée pour le ré-étiquetage imprécis des exemples est basée sur la notion d'*entropie de Shannon* ici utilisée comme indicateur du doute du classifieur probabiliste dont on cherche à analyser l'espace de discernement. Nous pouvons considérer l'entropie de Shanon :

$$H(p(\cdot|\mathbf{x}^l)) = - \sum_{i=1}^n p(\theta_i|\mathbf{x}^l) \log(p(\theta_i|\mathbf{x}^l)),$$

où $p(\theta_i|\mathbf{x}^l)$ est la probabilité calculée en validation croisée par un classifieur probabiliste δ pour la classe $\theta_i \in \Theta$ pour l'exemple l'apprentissage \mathbf{x}^l . Lorsque l'entropie $H(p(\cdot|\mathbf{x}^l))$ est grande, cela signifie que la distribution de probabilité $p(\cdot|\mathbf{x}^l)$ représente de nombreuses informations différentes. Plus $H(p(\cdot|\mathbf{x}^l))$ est grande, plus le classifieur δ a des raisons de douter. Nous proposons une procédure de ré-étiquetage basée sur $H(p(\cdot|\mathbf{x}^l))$. Dans cette procédure, nous appliquons le classifieur probabiliste δ à la base d'apprentissage en validation croisée. Pour chaque exemple d'apprentissage \mathbf{x}^l , nous obtenons une distribution de probabilité $p(\cdot|\mathbf{x}^l)$ sur les classes de Θ . Ensuite, nous basons notre condition de ré-étiquetage sur la comparaison de la valeur de $H(p(\cdot|\mathbf{x}^l))$ avec un seuil fixé, noté ρ . Si $H(p(\cdot|\mathbf{x}^l))$ est supérieure au seuil ρ alors nous considérons que \mathbf{x}^l est un exemple difficile et lui assignons une nouvelle étiquette en adjoignant à sa classe θ^l , la classe ayant la plus grande probabilité d'être la vraie classe de \mathbf{x}^l . Si l'entropie est inférieure au seuil ρ alors nous conservons la classe initiale. Finalement, nous donnons à \mathbf{x}^l la nouvelle étiquette A^l comme suit :

$$A^l = \begin{cases} \{\theta^l, \arg \max_{\theta \in \Theta \setminus \{\theta^l\}} p(\theta|\mathbf{x}^l)\} & \text{si } H(p(\cdot|\mathbf{x}^l)) > \rho, \\ \{\theta^l\} & \text{sinon.} \end{cases}$$

Notons que si le ré-étiquetage permet de mieux gérer l'incertitude des données, il peut aussi affecter les pré-traitements classiques de l'apprentissage automatique, augmenter la complexité algorithmique et créer un déséquilibre dans la représentation des classes dans la base d'apprentissage par l'ajout de nouvelles classes. Dans les sous-sections suivantes nous tentons d'apporter des réponses à ces préoccupations.

4.1.1.1 Ré-étiquetage imprécis et extraction de caractéristiques

Lors du pré-traitement des exemples d'apprentissage, nous pouvons également extraire des caractéristiques afin de faciliter l'apprentissage et diminuer la complexité (voir chapitre 2, sous-sous-section 2.2.4.3). L'extraction de caractéristiques est compatible avec notre ré-étiquetage imprécis. Dans le cas où l'extraction est non supervisée, il suffit d'extraire de nouvelles caractéristiques \mathbf{x}'' pour $l = 1, \dots, L$ avant de procéder au ré-étiquetage. Dans le cas où nous préférons une extraction de caractéristiques supervisée, comme LDA, les processus d'extraction et de ré-étiquetage sont dépendants et nous avons plusieurs possibilités :

- appliquer une première extraction (Extraction 1 "pour le ré-étiquetage" à valeurs dans \mathcal{X}') sur les exemples d'apprentissage initiaux afin de faciliter la tâche de ré-étiquetage (nous notons x'' les caractéristiques d'un exemple d'apprentissage \mathbf{x} après l'Extraction 1),

- et/ou appliquer une seconde extraction (Extraction 2 "pour l'apprentissage avec les nouvelles classes" à valeurs dans \mathcal{X}'') en considérant les caractéristiques initiales et les nouvelles étiquettes afin de faciliter la classification (nous notons x''' les caractéristiques d'un exemple d'apprentissage x après l'Extraction 2).

En effet si l'extraction supervisée vise à faciliter la classification, l'extraction 2 "pour l'apprentissage avec les nouvelles classes" semble appropriée car elle renforce l'indépendance de traitement des nouvelles étiquettes. La Fig. 4.3 résume les pré-traitements de la base d'apprentissage avec ou sans extraction de caractéristiques.

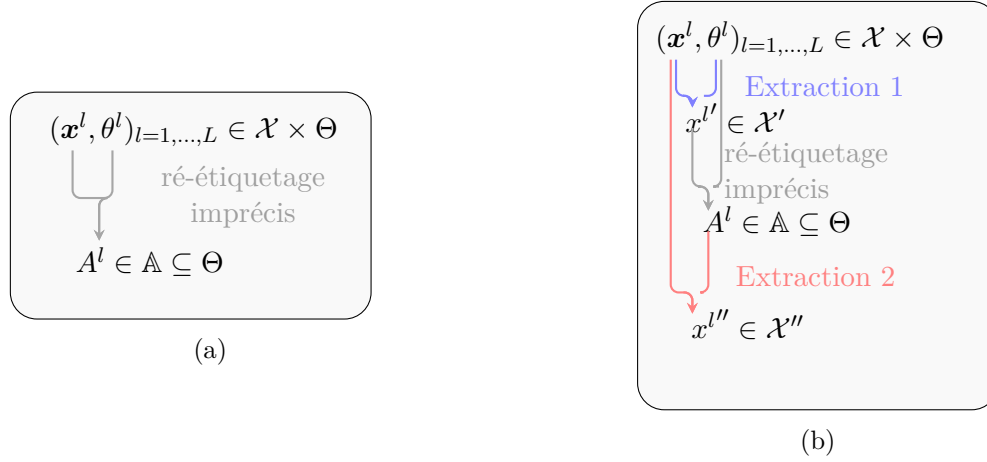


FIG. 4.3 Pré-traitements de la base d'apprentissage dans le cas sans extraction (a) / avec extraction (b)

Une fois que tous les exemples d'apprentissage ont été ré-étiquetés $(x^l, A^l)_{l=1,...,L}$, le classifieur probabiliste δ peut être entraîné. L'entraînement sur cette base d'apprentissage permet d'isoler les informations attribuables à chaque sous-ensemble indépendamment des autres. De plus, bien que le nombre de classes considérées augmente, la séparation entre les nouvelles classes $A \in \mathbb{A}$ est plus claire (il y a moins de zones de chevauchement entre les classes) ce qui facilite la tâche de la classification.

4.1.1.2 Ré-étiquetage imprécis et déséquilibre des classes

Puisque les nouvelles classes suite à la phase de ré-étiquetage sont plus nombreuses, elles sont moins bien fournies en exemples et peuvent également être davantage déséquilibrées.

Exemple 2. Reprenons le problème fictif de classification utilisé dans l'exemple 1. La TABLE 4.1 donne les répartitions des exemples dans l'ensemble des classes initiales Θ et dans l'ensemble des nouvelles classes \mathbb{A} ici égal à l'ensemble $\{\{\theta_1\}, \{\theta_2\}, \{\theta_3\}, \{\theta_1, \theta_2\}, \{\theta_1, \theta_3\}, \Theta\}$.

Base de données initiale		
θ_1	θ_2	θ_3
20	20	20

Base de données ré-étiquetées					
$A_1 = \{\theta_1\}$	$A_2 = \{\theta_2\}$	$A_3 = \{\theta_3\}$	$A_4 = \{\theta_1, \theta_2\}$	$A_5 = \{\theta_1, \theta_3\}$	$A_6 = \Theta$
12	16	15	5	7	5

TABLE 4.1 – Répartitions des exemples d'apprentissage

Lorsque les classes sont trop déséquilibrées, un classifieur a tendance à moins bien classifier les exemples des classes minoritaires. En effet, les classes minoritaires sont moins déterminantes dans l'évaluation interne de l'apprentissage car elles ne représentent qu'une petite partie des exemples. De plus ces classes peuvent également être mal connues justement en raison du manque d'exemples. Pour gérer ce déséquilibre des classes que le ré-étiquetage peut avoir engendré, nous proposons d'utiliser la méthode SMOTE (Synthetic Minority Over-sampling Technique en anglais) proposée dans (Chawla et al., 2002) qui consiste à créer des exemples synthétiques pour augmenter les classes minoritaires. Dans la première version SMOTE proposée, les exemples synthétiques sont introduits sur les segments reliant les exemples d'une même classe minoritaire. En pratique pour chaque exemple de la classe minoritaire, on considère ses k plus proches voisins. Pour chacun des k voisins, l'exemple synthétique est obtenu par la différence entre les vecteurs de caractéristiques de l'exemple et de son voisin multipliée par un nombre aléatoire. En fonction du nombre d'exemples synthétiques souhaité, des exemples sont aléatoirement choisis dans les k plus proches voisins.

Exemple 3. Reprenons l'exemple 1 de classification présentant une base de données ré-étiquetée. Nous avons donné dans la suite de cet exemple (exemple 2) la répartition des exemples dans les nouvelles classes. Montrons dans cet exemple comment la méthode de sur-échantillonnage SMOTE peut être utilisée pour ré-équilibrer la base d'apprentissage. La procédure est présentée dans la Fig. 4.4. La première sous-figure 4.4a représentant la base d'apprentissage ré-étiquetée déséquilibrée est rappelée. La classe la mieux fournie en exemples est $A_2 = \{\theta_2\}$ avec ses 16 exemples. Dans la seconde sous-figure 4.4b, la création des exemples synthétiques (petits points rouges) nécessaires pour atteindre 16 exemples dans chaque classe est illustrée. La sous-figure 4.4c montre la base d'apprentissage ré-étiquetée ré-équilibrée.

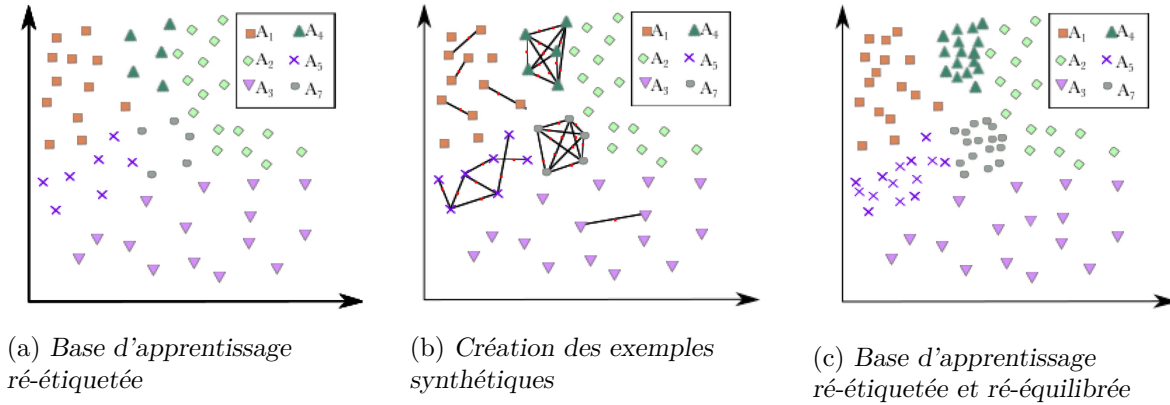


FIG. 4.4 Création d'exemples synthétiques avec la méthode SMOTE

Lorsque la base d'apprentissage est ré-étiquetée et ré-équilibrée, nous entraînons un classifieur probabiliste. Nous notons ce classifieur $\delta_{\mathbb{A}}$ pour insister sur le fait que les classes considérées par le classifieur sont des éléments de \mathbb{A} . Pour un nouvel exemple à traiter \mathbf{x} , $\delta_{\mathbb{A}}$ fournit une distribution de probabilité sur les éléments de \mathbb{A} . Puisque les éléments de \mathbb{A} sont en pratique aussi des éléments de 2^{Θ} (*i.e.* des ensembles de classes de Θ), la probabilité a posteriori fournie par $\delta_{\mathbb{A}}$ s'interprète comme une fonction de masse $m(\cdot|\mathbf{x})$. Notons que les éléments focaux sont forcément des éléments de \mathbb{A} .

Dans les deux précédentes sections nous avons donné un moyen générique d'assigner une fonction de masse à partir d'un classifieur probabiliste. Dans la section suivante, nous supposons

que nous disposons de la fonction de masse a posteriori $m(\cdot|\mathbf{x})$ pour prendre la décision du choix de l'ensemble à fournir pour répondre à un problème de classification imprécise.

4.1.2 Choix d'une sortie imprécise à partir d'une fonction de masse et d'un critère de décision

Dans cette section nous supposons que nous disposons d'une information, fournie par un classifieur noté δ_{2^Θ} , concernant un nouvel exemple \mathbf{x} dont la classe $\theta_{\mathbf{x}}^*$ n'est pas connue, sous la forme d'une fonction de masse a posteriori $m(\cdot|\mathbf{x})$ sur 2^Θ . L'objectif est maintenant de fournir une sortie imprécise $\delta_{2^\Theta}(\mathbf{x}) \subseteq \Theta$, la plus adéquate pour l'exemple \mathbf{x} sur la base de $m(\cdot|\mathbf{x})$ et d'un critère de décision que nous allons définir. Le choix de $\delta_{2^\Theta}(\mathbf{x})$ signifie que tous ses éléments sont candidats à la véritable classe $\theta_{\mathbf{x}}^*$ et le manque d'information empêche de renvoyer une sortie plus pertinente. Pour ce faire, nous allons nous baser sur la théorie de la décision en construisant une fonction *perte*, puis une fonction de *risque* qui constituera notre critère de décision pour le choix de la sortie pour \mathbf{x} . Comme nous faisons le choix d'une sortie imprécise, deux adaptations sont nécessaires dans la définition de la fonction *perte*.

D'une part, nous considérons l'ensemble des actions possibles $\mathcal{A} = 2^\Theta \setminus \{\emptyset\}$, où l'action $A \in \mathcal{A}$ signifie "attribuer le sous-ensemble A à l'exemple \mathbf{x} ". D'autre part, l'erreur classique qui consiste à vérifier si la classe prédite est la même que la vraie classe de l'exemple, est remplacée par les erreurs de spécialisation et de généralisation (chapitre 3, sous-section 3.3.4). Nous avons déjà montré comment tenir compte de l'erreur de généralisation dans l'état de l'art. Notre approche se distingue par la prise en compte de l'erreur de spécialisation, par l'association d'une perte à la sortie possible A lorsque celle-ci est plus précise que $B_{\mathbf{x}}^*$, où $B_{\mathbf{x}}^*$ est le plus petit ensemble en accord avec l'information disponible sur $\theta_{\mathbf{x}}^*$. $B_{\mathbf{x}}^*$ représente l'information dont nous disposons sur la vraie classe $\theta_{\mathbf{x}}^*$ de \mathbf{x} . Par exemple si nous savons seulement que $\theta_{\mathbf{x}}^*$ est dans l'ensemble $\{\theta_1, \theta_3\}$ alors nous avons $B_{\mathbf{x}}^* = \{\theta_1, \theta_3\}$.

Nous proposons de construire une telle fonction de perte en se basant sur la mesure F_β initialement définie en recherche d'information (IR). Dans le chapitre précédent, nous avons vu que F_β pouvait être utilisée comme mesure d'évaluation de sortie imprécise (Coz et al., 2009). Dans (Coz et al., 2009), F_β est définie sur $\Theta \times 2^\Theta$ et la valeur $F_\beta(\theta_{\mathbf{x}}^*, \delta_{2^\Theta}(\mathbf{x}))$ est la récompense donnée à la sortie $\delta_{2^\Theta}(\mathbf{x})$ lorsque la vraie classe est $\theta_{\mathbf{x}}^*$. Dans notre approche, nous adoptons la formulation de F_β initialement définie sur $2^\Theta \times 2^\Theta$ en IR. Dans notre contexte, nous définissons la fonction F_β , $\beta \geq 0$, pour l'information la plus précise disponible concernant $\theta_{\mathbf{x}}^*$ représentée par $B_{\mathbf{x}}^*$ et la sortie possible A comme suit :

$$F_\beta(B_{\mathbf{x}}^*, A) = \frac{(1 + \beta^2)|A \cap B_{\mathbf{x}}^*|}{\beta^2|B_{\mathbf{x}}^*| + |A|}.$$

$F_\beta(B_{\mathbf{x}}^*, A)$ représente la récompense donnée au classifieur quand il prédit un ensemble de classe A alors que l'information la plus précise disponible concernant $\theta_{\mathbf{x}}^*$ est $B_{\mathbf{x}}^*$. Nous retrouvons la définition de la mesure d'évaluation F_β donnée dans le chapitre précédent, sous-section 3.3.2, lorsque la vraie classe $\theta_{\mathbf{x}}^*$ est précisément connue. La TABLE 4.2 montre les récompenses obtenues lorsque $\Theta = \{\theta_1, \theta_2, \theta_3\}$. Notons, que nous obtenons la TABLE 3.3, sous-section 3.3.2 du chapitre précédent pour $\Theta = \{Covid, Grippe, Angine\}$, avec les quatre dernières lignes en plus.

Dans ce qui suit, nous proposons de commenter certaines situations en fonction du sous-ensemble représentant l'information disponible $B_{\mathbf{x}}^*$ et de l'ensemble prédit A . Le pire scénario de prédiction se produit lorsqu'il n'y a pas d'intersection entre A et $B_{\mathbf{x}}^*$. Cela est toujours sanctionné par $F_\beta(B_{\mathbf{x}}^*, A) = 0$. En revanche, quand $A = B_{\mathbf{x}}^*$, nous avons $F_\beta(B_{\mathbf{x}}^*, A) = 1$ pour toutes les valeurs de β . Remarquons que nous avons l'occasion de donner la récompense maximale à

		Sortie envisagée A						
		$\{\theta_1\}$	$\{\theta_2\}$	$\{\theta_3\}$	$\{\theta_1, \theta_2\}$	$\{\theta_1, \theta_3\}$	$\{\theta_2, \theta_3\}$	Θ
B_x^*	$\{\theta_1\}$	1	0	0	$\frac{1+\beta^2}{2+\beta^2}$	$\frac{1+\beta^2}{2+\beta^2}$	0	$\frac{1+\beta^2}{3+\beta^2}$
	$\{\theta_2\}$	0	1	0	$\frac{1+\beta^2}{2+\beta^2}$	0	$\frac{1+\beta^2}{2+\beta^2}$	$\frac{1+\beta^2}{3+\beta^2}$
	$\{\theta_3\}$	0	0	1	0	$\frac{1+\beta^2}{2+\beta^2}$	$\frac{1+\beta^2}{2+\beta^2}$	$\frac{1+\beta^2}{3+\beta^2}$
	$\{\theta_1, \theta_2\}$	$\frac{1+\beta^2}{1+2\beta^2}$	$\frac{1+\beta^2}{1+2\beta^2}$	0	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1+\beta^2}{3/2+\beta^2}$
	$\{\theta_1, \theta_3\}$	$\frac{1+\beta^2}{1+2\beta^2}$	0	$\frac{1+\beta^2}{1+2\beta^2}$	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1+\beta^2}{3/2+\beta^2}$
	$\{\theta_2, \theta_3\}$	0	$\frac{1+\beta^2}{1+2\beta^2}$	$\frac{1+\beta^2}{1+2\beta^2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	$\frac{1+\beta^2}{3/2+\beta^2}$
	Θ	$\frac{1+\beta^2}{1+3\beta^2}$	$\frac{1+\beta^2}{1+3\beta^2}$	$\frac{1+\beta^2}{1+3\beta^2}$	$\frac{1+\beta^2}{1+3/2\beta^2}$	$\frac{1+\beta^2}{1+3/2\beta^2}$	$\frac{1+\beta^2}{1+3/2\beta^2}$	1

TABLE 4.2 – Utilité F_β généralisée.

des sous-ensembles de cardinal > 1 , ce qui n'était pas envisageable avec aucune des mesures d'évaluation de classifieurs imprécis évoquées dans le chapitre précédent (à part acc qui ne prend en compte que la prudence). Entre ces deux cas extrêmes, les erreurs de généralisation (manque de pertinence) i.e., $B_x^* \subset A$ et les erreurs de spécialisation i.e., $A \subset B_x^*$ sont contrôlées différemment en fonction de β :

- Lorsque β est petit, le manque de pertinence est davantage pénalisé. En particulier pour $\beta = 0$, nous avons

$$F_0(B_x^*, A) = \frac{|A \cap B_x^*|}{|A|}.$$

Dans ce cas, une erreur de spécialisation n'est pas pénalisée (récompense maximale 1), alors qu'une erreur de généralisation reçoit la récompense $|B_x^*|/|A|$.

- Lorsque β est grand, c'est l'erreur de spécialisation qui est davantage pénalisée. Pour $\beta \rightarrow \infty$, nous avons

$$\lim_{\beta \rightarrow \infty} F_\beta(B_x^*, A) = \frac{|A \cap B_x^*|}{|B_x^*|}.$$

Cette fois, c'est uniquement l'erreur de spécialisation qui est sanctionnée avec la récompense $|A|/|B_x^*|$.

Pour un exemple \mathbf{x} à classifier de fonction de masse $m(\cdot|\mathbf{x})$, nous prédisons l'ensemble $\delta_{2\Theta}(\mathbf{x})$ minimisant le risque suivant :

$$r_\beta(A|\mathbf{x}) = \sum_{B \subseteq \Theta} m(B|\mathbf{x})L_\beta(B, A), \quad (27)$$

où la fonction de perte est donnée par $L_\beta = 1 - F_\beta$. Finalement la sortie imprécise fournie par notre approche est :

$$\delta_{2\Theta}(\mathbf{x}) = \arg \min_{A \subseteq \Theta} r_\beta(A|\mathbf{x}),$$

Nous aurions pu construire un critère basé sur l'espérance de gain à la place du risque mais nous avons préféré passer par la fonction perte pour faciliter la comparaison avec l'état de l'art.

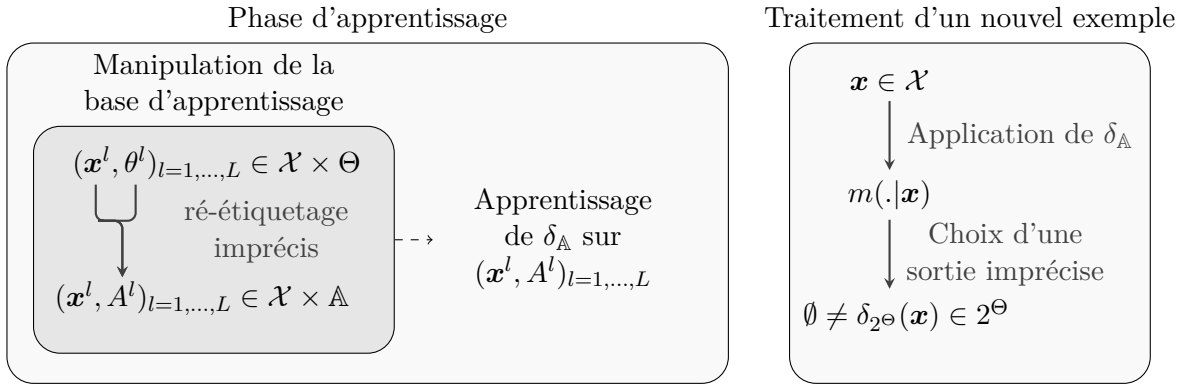


FIG. 4.5 Étapes de la classification imprécise proposée par ECLAIR

Dans cette section, nous avons présenté notre approche de classification évidentielle imprécise. Le résumé de la procédure que nous appelons ECLAIR est proposé dans la Fig. 4.5.

4.2 Discussion

Dans cette section, nous lions l'approche proposée aux travaux de l'état de l'art. En ce qui concerne les procédures de ré-étiquetage, comme cela a été évoqué dans le chapitre 2 de nombreuses recherches ont été menées pour identifier les exemples suspects dans l'intention de les supprimer ou de les ré-étiqueter par une classe concurrente plus appropriée (Lallich et al., 2002). Ceci est généralement fait pour améliorer la performance. D'autres travaux consistent à ré-étiqueter dans des classes imprécises. Cela a été fait pour tester une approche de classification évidentielle sur des données ré-étiquetées imprécises dans (Zhang et al., 2004). Mais, comme nous l'avons déjà indiqué, notre ré-étiquetage imprécis a un objectif différent, à savoir expliciter l'imprécision des données (zones de chevauchement ou avec peu d'exemples) dans l'espace des caractéristiques.

Certains travaux comme (Hüllermeier et al., 2019) s'intéressent à l'apprentissage à partir d'exemples ayant des étiquettes imprécises. À ce sujet une adaptation du k -nn évidentiel ((Denoeux, 1995), présenté dans le chapitre 2, section 2.3.2) au cas d'exemples ayant des étiquettes imprécises était évoquée dans (Zouhal et Denoeux, 1998). Dans ce cas, l'apport d'évidence par un voisin dont le vecteur de caractéristiques serait \mathbf{x}^v de classe imprécise A^v est représenté sous la forme d'une fonction de masse m_v :

$$\forall A \subseteq \Theta, m_v(A|\mathbf{x}) = \begin{cases} \alpha \phi_i(d(\mathbf{x}, \mathbf{x}^v)) & \text{si } A = A^v, \\ 1 - \alpha \phi_i(d(\mathbf{x}, \mathbf{x}^v)) & \text{si } A = \Theta, \\ 0 & \text{si } A \in 2^\Theta \setminus \{\Theta, \{A^v\}\}. \end{cases}$$

où $d(\mathbf{x}, \mathbf{x}^v)$ est la distance entre \mathbf{x} et son voisin \mathbf{x}^v , α un paramètre tel que $0 < \alpha < 1$ et ϕ_i une fonction décroissante vérifiant $\phi_i(0) = 1$ et $\lim_{z \rightarrow \infty} \phi_i(z) = 0$. Remarquons que les éléments focaux ne sont plus réduits aux singletons et à l'ignorance totale, comme ils l'étaient pour le cas de données précisément étiquetées. Ce classifieur évidentiel pourrait s'intégrer dans notre approche de classification (après le ré-étiquetage imprécis, au lieu d'utiliser un classifieur probabiliste pour assigner les masses). Notons que ce classifieur est "conscient" des relations entre les ensembles alors que l'apprentissage que nous proposons ignore ces relations.

D'autre part, la procédure de décision par minimisation de risque a été proposée pour répondre aux attentes de la classification imprécise. Dans la littérature de nombreuses procédures existent pour choisir un ensemble à partir d'une fonction de masse a posteriori. Dans le chapitre 2, nous avons vu que la plupart des décisions basées sur les fonctions de croyance consistent à choisir l'action $a \in \mathcal{A}$ qui minimise le risque suivant (Denoeux, 2019) :

$$\sum_{B \subseteq \Theta} m(B|\mathbf{x})\mathcal{L}(B, a),$$

où la quantité $\mathcal{L}(B, a)$ est exprimée en fonction d'une fonction de perte $L : \Theta \times \mathcal{A} \rightarrow \mathbb{R}$.

Pour ces critères de décision, la quantité $\mathcal{L}(B, a)$ concerne la perte encourue en choisissant l'action $a \in \mathcal{A}$, quand la vraie classe est comprise dans $B \subseteq \Theta$. Pour quantifier cette perte, des pertes définies pour des réalités précises sont agrégées sur B . Au lieu d'agréger des pertes définies pour des réalités précises, notre procédure définit directement une perte en fonction de l'information disponible B . Nous avons $\mathcal{L}(B, a) = L(B, a)$ où L est une perte définie sur $2^\Theta \times \mathcal{A}$. Cette perte généralisée est plus difficilement interprétable mais permet de fixer les erreurs de spécialisation et également de mieux récompenser des ensembles de cardinal > 1 lorsque l'information n'est pas suffisante pour décider précisément. L'exemple suivant pointe l'avantage de notre approche dans cette situation.

Exemple 4. *Considérons un exemple dont les caractéristiques observées \mathbf{x} sont affectées par de l'imprécision de telle sorte que le vecteur de caractéristiques \mathbf{x} est tel qu'un expert est incapable de donner une information plus précise que "la vraie classe est soit θ_1 soit θ_2 ". Cela se traduit par $m(\{\theta_1, \theta_2\}|\mathbf{x}) = 1$. Supposons que nous envisagions l'action $a = \{\theta_1, \theta_2\}$ (i.e., fournir la sortie imprécise $\{\theta_1, \theta_2\}$ pour cet exemple). Dans ce cas, le calcul du risque suivant (Denoeux, 2019) repose sur $\mathcal{L}(\{\theta_1, \theta_2\}, \{\theta_1, \theta_2\})$ qui est calculé en fonction des pertes $L(\{\theta_1\}, \{\theta_1, \theta_2\})$ et $L(\{\theta_2\}, \{\theta_1, \theta_2\})$. Pourtant, nous savons que l'ensemble $\{\theta_1, \theta_2\}$ répond aux attentes de la classification imprécise. Nous devrions n'avoir aucune perte associée à cette action étant données les informations disponibles.*

4.3 Illustration de l'approche proposée ECLAIR

Dans cette section nous donnons une illustration détaillée, du déroulement de l'approche proposée ECLAIR. Pour simplifier la présentation, nous considérons le problème de classification multi-classes utilisé dans le chapitre 3, sous-section 3.1.2, configuration 2. Pour rappel, il s'agit d'un problème simulé de classification pour lequel l'ensemble des classes précises possible est $\Theta = \{\theta_1, \theta_2, \theta_3\}$ et les trois classes sont simulées par des distributions gaussiennes de paramètres respectifs $\mu_1 = (0, 0)$, $\mu_2 = (0, 1)$, $\mu_3 = (1/2, \sqrt{3}/2)$ et $\sigma_1 = \sigma_2 = \sigma_3 = 0.01$. De plus un bruit gaussien ϵ de variance 0.1 (écart type $\sqrt{0.1} \simeq 0.316$) est ajouté (configuration 2 dans l'exemple d'origine). Nous considérons une base d'apprentissage constituée de 600 exemples (200 exemples pour chaque classes) ainsi générés. La Fig 4.6 représente cette base d'apprentissage.

Dans cette illustration, nous montrons comment à partir de cette base d'apprentissage, le classifieur évidentiel et imprécis ECLAIR, effectue ses prédictions pour de nouveaux exemples.

Nous commençons par la phase de ré-étiquetage de la base d'apprentissage. Nous avons proposé plusieurs manières d'identifier puis de ré-étiqueter les exemples d'apprentissage dans des classes sous forme d'ensembles correspondant à leur niveau d'imprécision. Dans cette illustration nous considérons l'approche basée sur l'entropie obtenue en validation croisée (10 blocs) par le classifieur Naive Bayes. Nous fixons le seuil d'entropie à $\rho = 0.8$. La Fig 4.7a montre les valeurs d'entropie obtenues pour chaque exemple d'apprentissage. Les exemples entourés sont

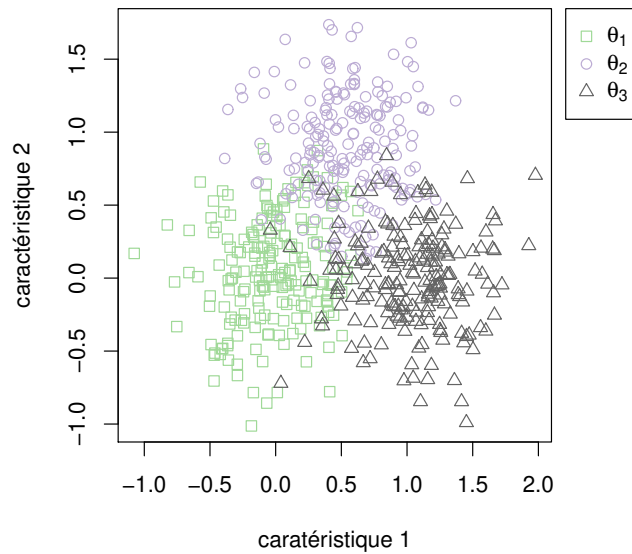
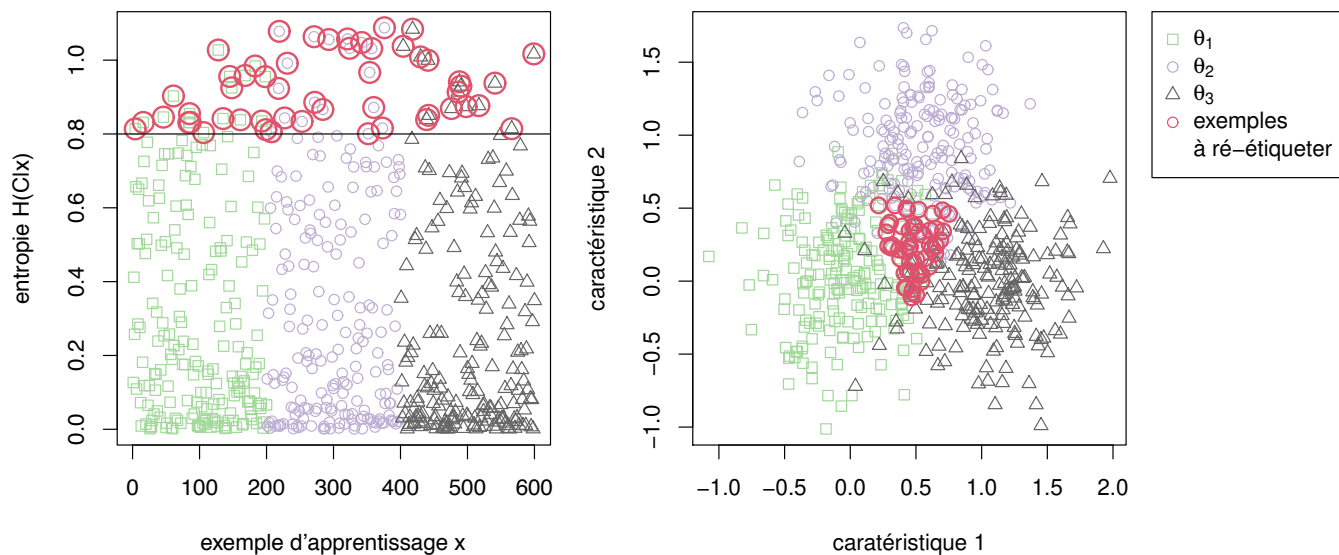


FIG. 4.6 Base d'apprentissage initiale

ceux dont la distribution de probabilité obtenue a une entropie supérieure à 0.8, et qui doivent par conséquent être ré-étiquetés. Ces exemples à ré-étiqueter par des sous-ensembles de classes sont reportés sur la Fig. 4.7b. Nous constatons que ces exemples correspondent à des zones de chevauchement de classes ce qui est en accord avec l'objectif du ré-étiquetage pour expliciter l'imprécision des exemples.



(a) Valeurs d'entropie obtenues pour chaque exemple d'apprentissage.

(b) Représentation des exemples d'apprentissage dans l'espace des caractéristiques

FIG. 4.7 Sélection des exemples d'apprentissage à ré-étiqueter dans des ensembles de classes

Les exemples dont la valeur d'entropie associée est inférieure à 0.8, sont ré-étiquetés par

le singleton de la classe d'origine. La TABLE 4.3, montre le ré-étiquetage de deux exemples d'apprentissage dont la classe d'origine est θ_1 . Le premier exemple obtient une valeur d'entropie inférieure à 0.8 et n'est pas considéré comme un exemple *difficile*. Le second exemple au contraire est considéré comme un exemple *difficile*, il est ré-étiqueté dans l'ensemble $\{\theta_1, \theta_3\}$.

Exemple d'apprentissage	Vraie classe	Probabilité a posteriori			Entropie	→	Nouvelle étiquette
		$p(\theta_1 \mathbf{x})$	$p(\theta_2 \mathbf{x})$	$p(\theta_3 \mathbf{x})$			
\mathbf{x}	θ						\mathbb{A}
(-0.159, 0.171)	θ_1	0.975	0.020	0.004	0.127		$\{\theta_1\}$
(0.502, -0.095)	θ_1	0.406	0.037	0.557	0.814		$\{\theta_1, \theta_3\}$

TABLE 4.3 – Ré-étiquetage de deux exemples d'apprentissage

Lorsque tous les exemples d'apprentissage sont ré-étiquetés avec cette procédure, nous obtenons une nouvelle base d'apprentissage représentée dans la Fig. 4.8.

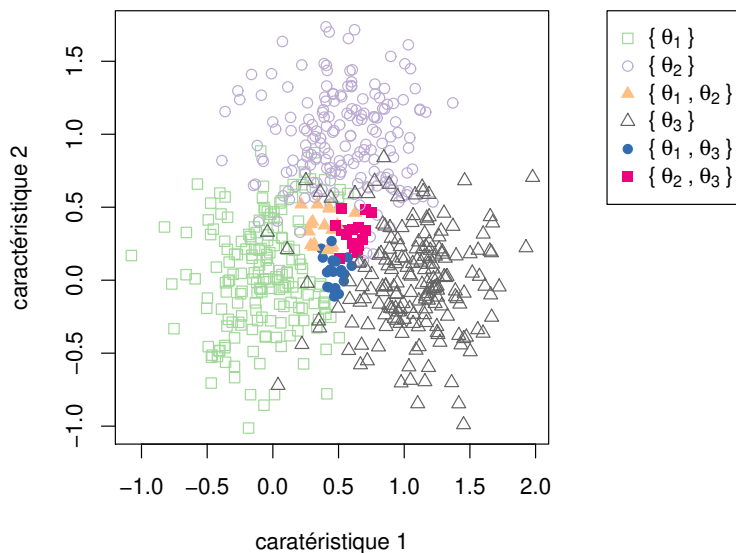


FIG. 4.8 Base d'apprentissage ré-étiquetée

Cette nouvelle base d'apprentissage est déséquilibrée comme le montre la répartition des exemples donnée dans la TABLE 4.4.

$\{\theta_1\}$	$\{\theta_2\}$	$\{\theta_2\}$	$\{\theta_1, \theta_2\}$	$\{\theta_2, \theta_3\}$	$\{\theta_1, \theta_3\}$
182	182	185	17	19	15

TABLE 4.4 – Effectifs des exemples dans les nouvelles classes

En utilisant la méthode SMOTE (sous-sous-section 4.1.1.2), en considérant tous les exemples comme potentiels voisins ($k = \infty$), nous créons des exemples synthétiques afin de ré-équilibrer les classes. Nous obtenons la base de données définitive représentée dans la Fig. 4.9 avec 185 exemples dans chaque classe.

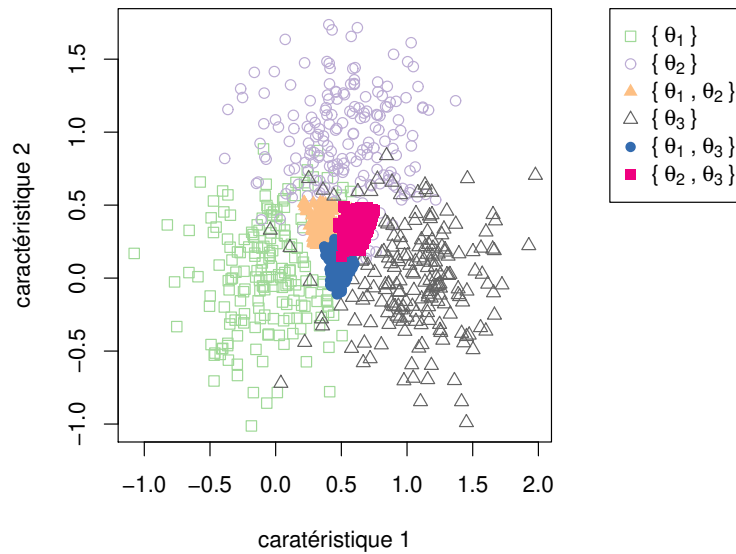


FIG. 4.9 Base d'apprentissage définitive (après ré-étiquetage et ré-équilibrage des classes)

Nous entraînons ensuite le classifieur probabiliste Naive Bayes sur les six classes.

Détaillons le traitement d'un nouvel exemple x' représenté dans la base de test Fig. 4.10. Lorsque le classifieur probabiliste Naive Bayes est appliqué à x' , nous obtenons la fonction de masse a posteriori $m(.|x')$ donnée dans Fig. 4.10.

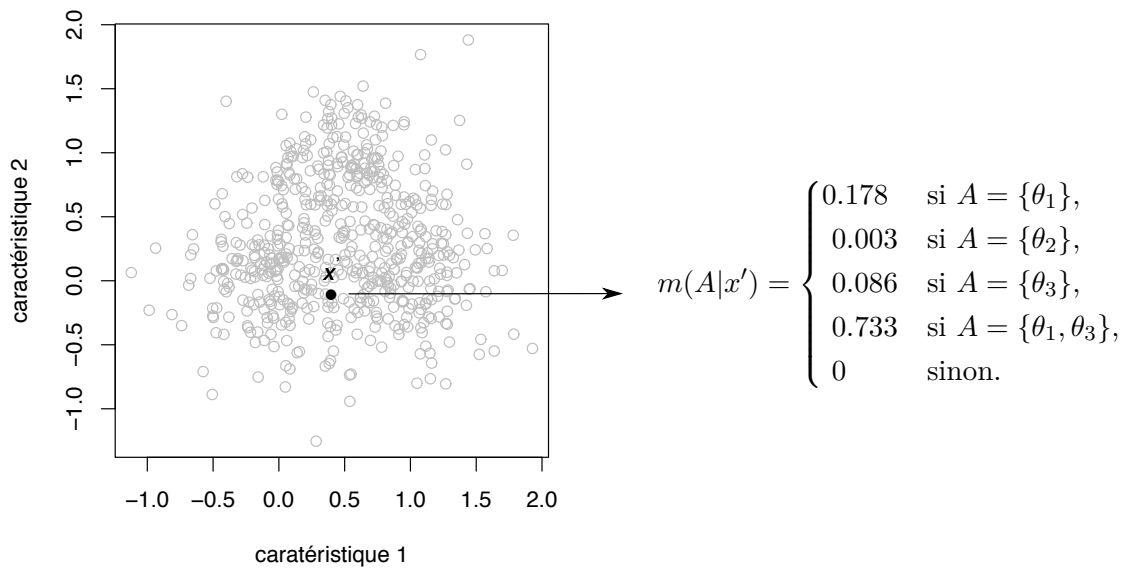


FIG. 4.10 Nouvel exemple x' à classifier dans la base de test et sa fonction de masse associée $m(.|x')$

Nous effectuons ensuite la décision à partir de la fonction de masse $m(.|x')$ par la minimisation du risque (défini équation (27)) en prenant ici $\beta = 1$. La fonction de perte pour $\beta = 1$ est donnée dans la TABLE 4.5. La dernière ligne de la table donne le risque associé à chaque action (sortie) possible dans 2^Θ .

		Sortie possible A						
		$\{\theta_1\}$	$\{\theta_2\}$	$\{\theta_3\}$	$\{\theta_1, \theta_2\}$	$\{\theta_1, \theta_3\}$	$\{\theta_2, \theta_3\}$	Θ
Ensemble observé B	$\{\theta_1\}$	0	1	1	0.33	0.33	1	0.5
	$\{\theta_2\}$	1	0	1	0.333	1	0.333	0.5
	$\{\theta_3\}$	1	1	0	1	0.333	0.333	0.5
	$\{\theta_1, \theta_2\}$	0.333	0.333	1	0	0.5	0.5	0.2
	$\{\theta_1, \theta_3\}$	0.333	1	0.333	0.5	0	0.5	0.2
	$\{\theta_2, \theta_3\}$	1	0.333	0.333	0.5	0.5	0	0.2
	Θ	0.5	0.5	0.5	0.2	0.2	0.2	0
R(A)		0.333	0.997	0.425	0.513	0.091	0.574	0.280

TABLE 4.5 – Perte L_β

D'après ces calculs, nous concluons que la sortie du classifieur imprécis ECLAIR doit être $\{\theta_1, \theta_3\}$ (car cet ensemble a obtenu la plus petite valeur de risque). Détaillons le calcul du risque pour une autre sortie possible, disons $\{\theta_1\}$, afin de commenter la prise en compte de l'erreur de spécialisation :

$$\begin{aligned}
R(\{\theta_1\}) &= \sum_{B \subseteq \Theta} m(B) L_1(\{\theta_1\}, B), \\
&= 0.178 L_1(\{\theta_1\}, \{\theta_1\}) + 0.003 L_1(\{\theta_1\}, \{\theta_2\}) + 0.086 L_1(\{\theta_1\}, \{\theta_3\}) + 0.733 \underbrace{L_1(\{\theta_1\}, \{\theta_1, \theta_3\})}_{\text{Perte associée à une erreur de spécialisation}}, \\
&= 0.178 \times 0 + 0.003 \times 1 + 0.086 \times 1 + 0.733 \times 0.333, \\
&= 0.333.
\end{aligned}$$

Nous voyons que la valeur de risque associé à la sortie possible $\{\theta_1\}$ est fortement pénalisée par $L_1(\{\theta_1\}, \{\theta_1, \theta_3\})$ qui quantifie la perte associée à une erreur de spécialisation consistant à prédire $\{\theta_1\}$ alors que l'information disponible permettrait seulement de conclure $\{\theta_1, \theta_3\}$. Étant donné que la masse de croyance allouée exactement à l'ensemble $\{\theta_1, \theta_3\}$ (et non à des sous-ensembles plus spécifiques de $\{\theta_1, \theta_3\}$) est élevée (0.733), l'influence de cette pénalisation est importante.

Nous obtenons les sorties représentées dans la Fig. 4.11, en appliquant la même procédure à chaque exemple de la base de test constitué de 600 exemples générés de la même façon que les exemples d'apprentissage.

Nous pouvons comparer les sorties fournies par le classifieur imprécis ECLAIR avec celles obtenues avec le classifieur imprécis NONDET (chapitre précédent, sous-section 3.1.2, configuration 2). Nous redonnons les sorties obtenues avec le classifieur NONDET (instancié avec le classifieur Naive Bayes et $\beta = 1$) dans la Fig. 4.12. Les résultats des deux procédures sont donnés dans la TABLE 4.6.

	$acc_{discounted}$ ($= u_{50}$)	u_{65}	u_{80}	acc
ECLAIR	0.864	0.886	0.908	0.939
NONDET	0.870	0.886	0.902	0.925

TABLE 4.6 – Résultats de classification imprécise pour ECLAIR et NONDET

Les résultats de la TABLE 4.6 ne montrent pas de dominance complète entre les performances d'ECLAIR et celles de NONDET pour ce problème de classification simulé. Nous voyons que pour des évaluations où la pertinence est importante, NONDET est plus performant.

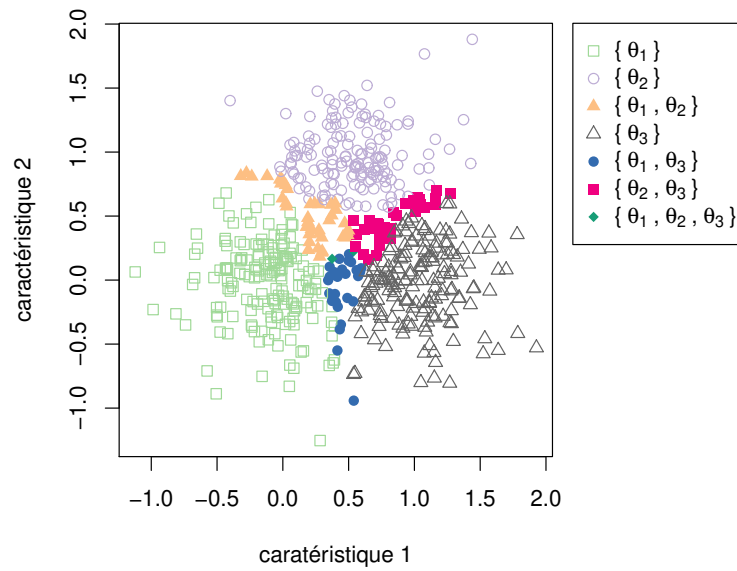


FIG. 4.11 Sorties ECLAIR pour les exemples de test

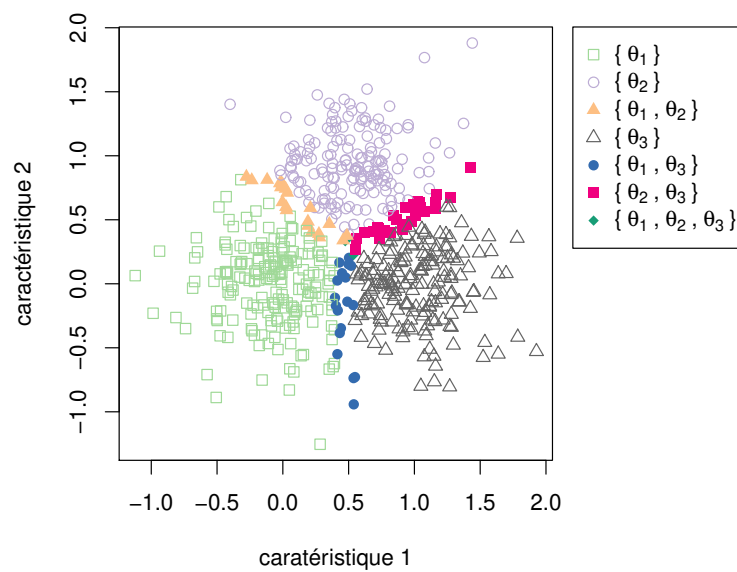


FIG. 4.12 Sorties NONDET pour les exemples de test

Lorsque c'est la prudence qui est valorisée, ECLAIR est plus performant. Plus généralement, nous remarquons que, dans les deux situations, les performances restent très proches.

4.4 Performances d'ECLAIR par rapport à l'état de l'art

Dans cette section, nous comparons les performances de l'approche classification imprécise proposée, ECLAIR, avec deux classifieurs imprécis de l'état de l'art, sur une sélection de jeux de données provenant du site *UCI Machine Learning Repository* et un jeu de données provenant de l'entreprise Pellenc ST que nous avons appelé *plastics* (voir Annexe A). Ces classifieurs sont le classifieur NONDET (Coz et al., 2009) et le classifieur NCC (Zaffalon, 2001), tous deux introduits dans le chapitre précédent. Nous considérons ces deux classifieurs comme des représentants des deux courants majoritaires de classifieurs imprécis. Le premier courant est celui des classifieurs imprécis basés sur la théorie des probabilités qui utilisent une règle de décision par minimisation de risque. Le second courant est celui des classifieurs imprécis basés sur les probabilités imprécises et qui effectuent leurs décisions par une règle de dominance. Nous donnons ci-dessous quelques précisions concernant les classifieurs NONDET et NCC :

- **NONDET** : Ce classifieur imprécis repose sur la minimisation de la perte $1 - F_\beta$, telle que définie dans le chapitre précédent (non généralisée au cas d'erreur de spécialisation). Comme l'approche proposée ECLAIR, ce classifieur générique, permet de donner des versions imprécises de classifieurs probabilistes. Le classifieur NONDET est également paramétré par β qui joue le rôle de degré de prudence tout comme pour ECLAIR.
- **NCC** : Ce classifieur est une extension du classifieur probabiliste Naive Bayes aux probabilités imprécises. Une introduction de ce classifieur est disponible dans le chapitre précédent et plus de détails sont donnés en Annexe D. Nous avons implémenté la version donnée dans l'article (Zaffalon, 2001) qui fait l'hypothèse que les caractéristiques sont à valeurs dans des espaces discrets et finis. Afin de rendre la procédure compatible avec les jeux de données dont les caractéristiques sont à valeurs dans des espaces continus, nous avons effectué une étape de discrétisations (comme cela a été fait dans l'article (Zaffalon, 2001)). A cette fin, nous avons utilisé une méthode de discrétisation basée sur le principe de la longueur minimale de description (Fayyad et Irani, 1993), avec la routine *mdlp* du package *discretization* de R. Le classifieur NCC tel que présenté dans (Zaffalon, 2001), est paramétré par $s > 0$ qui s'interprète comme un degré de prudence et que (Walley et Peter, 1991) suggère de choisir dans l'intervalle $[1, 2]$.

Le classifieur NCC a été comparé au classifieur NONDET (comme version imprécise du classifieur Naive Bayes) dans des travaux comme (Coz et al., 2009; Zaffalon et al., 2012). Ces travaux s'accordent sur les bonnes performances de NONDET par rapport à NCC pour des problèmes de classification de bio-informatique pour lesquels le nombre de caractéristiques est important et le nombre d'exemples disponibles est faible (Zaffalon et al., 2012).

Précisions que nous avons choisi d'effectuer ECLAIR avec le ré-étiquetage basé sur l'utilisation de l'entropie² et que nous avons effectué un ré-équilibre des classes avec la méthode SMOTE³ à chaque utilisation d'ECLAIR tout au long de la section. Puisque ces trois approches sont paramétrées et que les approches NONDET et ECLAIR offrent des versions pour n'importe quel classifieur probabiliste, nous commençons par commenter les résultats obtenus sur un unique jeu de données, par exemple, les données *wine*. Dans un second temps, nous présentons

2. en prenant $\rho = 0.8$ dans tous les cas excepté pour la version imprécise du classifieur SVM où nous avons considéré $\rho = 0.9$.

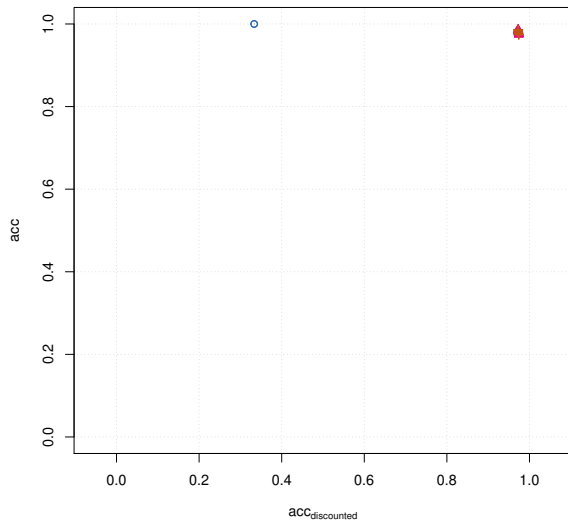
3. en considérant tous les exemples comme potentiels voisins ($k = \infty$).

des résultats plus généraux, sur la sélection des jeux de données pour des valeurs de paramètres fixées et pour les versions NONDET et ECLAIR du classifieur probabiliste Naive Bayes.

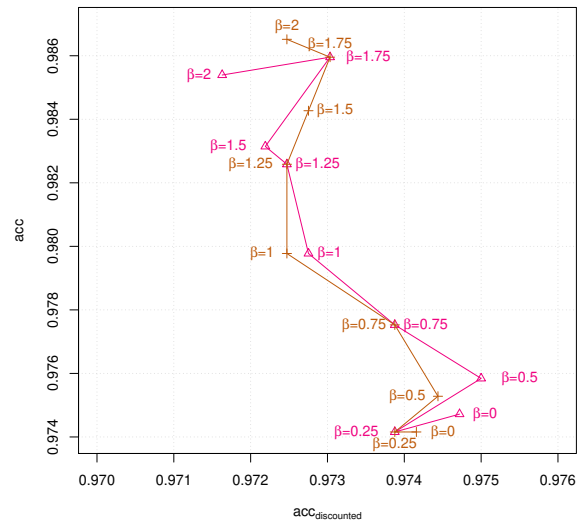
Nous commençons donc par donner des résultats des approches de classification imprécise NCC, NONDET et ECLAIR sur le jeu de données *wine* qui a peu de classes (3), peu de caractéristiques (13) et peu d'exemples (178), ce qui permet de réduire les temps de calculs. Nous avons appliqué NCC en choisissant $s = 1, 1.5, 2$ au jeu de données *wine*, ce qui a donné dans les trois cas les mêmes résultats. Pour les classifieurs NONDET et ECLAIR, nous avons choisi trois versions probabilistes des classifieurs standards, Naive Bayes, k -nn et SVM. Nous avons fourni des versions imprécises de ces classifieurs pour des valeurs de β dans $[0, 2]$. Les résultats obtenus sont présentés dans la Fig. 4.13. Les trois sous-figures de gauche, Fig. 4.13a, 4.13c et 4.13e représentent les récompenses acc en fonctions des récompenses $acc_{discounted}$ obtenues par les trois classifieurs imprécis, instanciés pour les classifieurs standards probabilistes Naive Bayes, k -nn et SVM. Nous avons vu dans le chapitre précédent que pour la mesure d'évaluation acc , seule la prudence de la sortie compte, une sortie reçoit la récompense maximale 1 à partir du moment où elle contient la vraie classe. Nous avons vu qu'au contraire, la mesure d'évaluation $acc_{discounted}$ valorise la pertinence et juge la prudence au même titre que la décision par l'aléatoire. Dans une application donnée, un classifieur imprécis doit trouver le compromis souhaité entre la prudence et la pertinence. En l'absence d'expression de compromis, nous pouvons considérer qu'un bon classifieur imprécis devrait être à la fois suffisamment prudent et pertinent. Un bon classifieur imprécis doit donc obtenir des récompenses élevées avec acc et $acc_{discounted}$, et donc se trouver dans la partie haute à droite dans les sous-figures qui suivent. Les trois sous-figures de droite Fig. 4.13b, 4.13d et 4.13f sont des zooms sur les points correspondants aux approches NONDET et ECLAIR pour les différentes valeurs de β .

Nous constatons que les approches NONDET et ECLAIR rivalisent sur des valeurs de récompenses proches alors que NCC est nettement moins pertinent mais obtient le score maximal de justesse ($acc = 1$), ce qui veut dire que toutes les sorties contiennent systématiquement la vraie classe. Dans les cas de NONDET et ECLAIR, nous voyons que le choix du classifieur probabiliste a un fort impact sur les performances de ces classifieurs imprécis. Pour ce jeu de données, parmi les trois classifieurs probabilistes, SVM est celui qui permet de maximiser à la fois la prudence et la pertinence. Enfin, les Fig. 4.13b, 4.13d et 4.13f nous permettent de constater l'influence du paramètre β . Cette influence s'explique par le fait que les classifieurs NONDET et ECLAIR sont optimisés pour F_β (ou F_β généralisé dans le cas d'ECLAIR), qui valorise la pertinence pour les petites valeurs de β et valorise la prudence pour les grandes valeurs de β . Nous retrouvons cette dynamique dans les résultats : pour les petites valeurs de β , $acc_{discounted}$ tend à être meilleur et acc moins bon, au contraire, pour les grandes valeurs de β , acc baisse et $acc_{discounted}$ augmente.

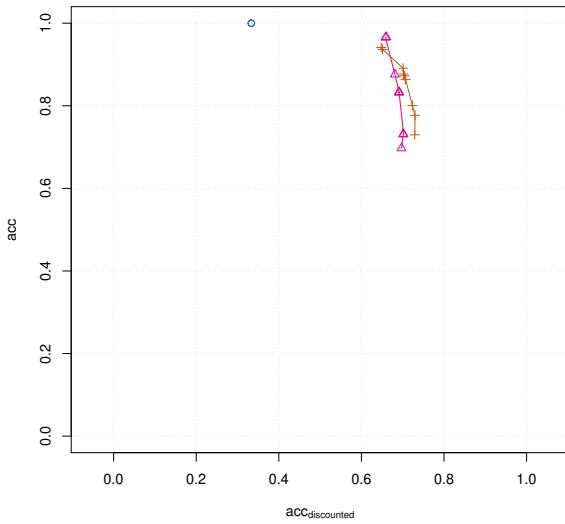
Afin de mieux évaluer les performances de l'approche proposée, nous proposons de la comparer à NONDET et NCC sur une sélection de problèmes de classification définis par 8 jeux de données. Chaque jeu de données ayant des particularités propres à son domaine d'application, nous pouvons ainsi vérifier la robustesse de l'approche proposée. Pour simplifier les comparaisons, nous prenons les choix de paramétrages suivants. Les versions de NONDET et NCC utilisent le classifieur Naive Bayes, et nous fixons $\beta = 1$, ce qui revient à optimiser la décision pour la mesure d'évaluation F_1 . D'après (Zaffalon et al., 2012), la mesure F_1 donne des récompenses comprises entre celles données par u_{65} et u_{80} , ce qui nous paraît être un bon compromis. Pour NCC, nous choisissons $s = 1$. La Fig. 4.14 donne les comparaisons deux à deux des trois classifieurs imprécis, en fonction des mesures d'évaluation $acc_{discounted}$ (qui évalue seulement la pertinence), u_{80} qui évalue un compromis entre la pertinence et la prudence, acc (qui ne tient compte que de la prudence). La légende des jeux de données est donnée dans la sous-figure



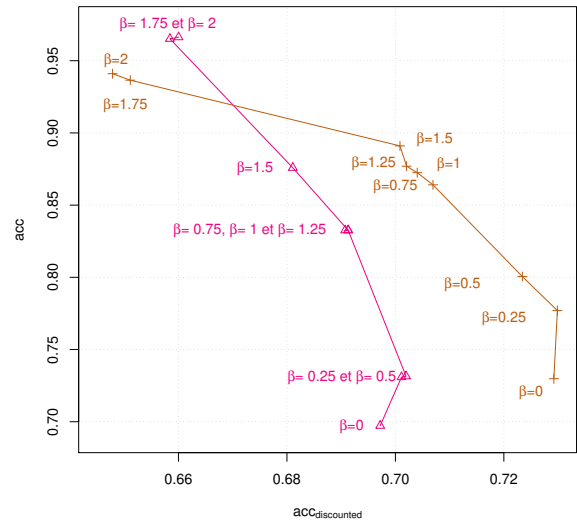
(a) Classifieur Naive Bayes



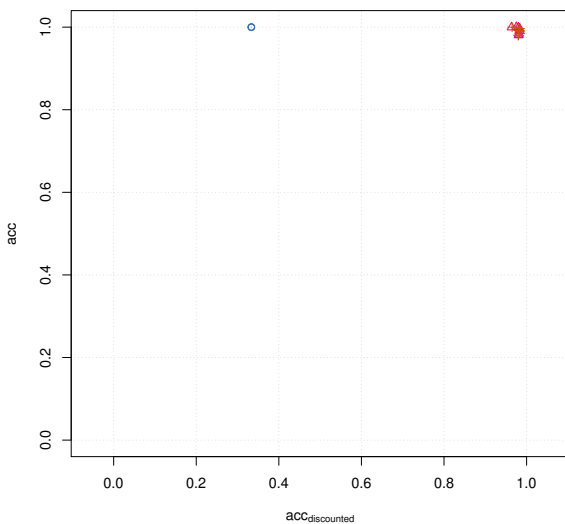
(b) Classifieur Naive Bayes (zoom)



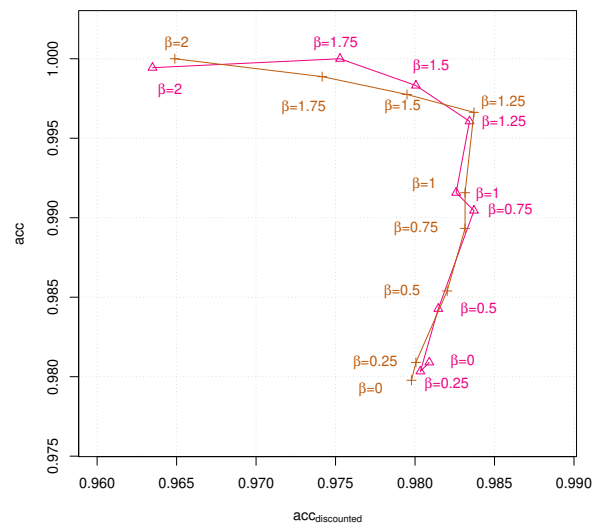
(c) Classifieur k-nn



(d) Classifieur k-nn (zoom)



(e) Classifieur SVM



(f) Classifieur SVM (zoom)

FIG. 4.13 Récompenses acc contre $acc_{discounted}$ obtenues sur le jeu de données *wine* pour les approches NONDET (Δ), ECLAIR ($+$) avec les classifieurs Naives Bayes, k -nn et SVM, et l'approche NCC (\circ) - résultats moyens obtenus sur 10 expériences de validations croisées à 10 blocs

4.14j. Dans toutes autres sous-figures, la droite $y = x$ est tracée pour départager les performances entre les deux classifieurs imprécis comparés. Par exemple, la sous-figure 4.14a montre que NONDET est plus pertinent que NCC pour tous les jeux de données. Commentons ces comparaisons deux à deux :

- **NONDET contre NCC** (sous-figures 4.14a, 4.14b, 4.14c) : c'est NONDET qui a l'avantage sur NCC en terme de pertinence et également pour le compromis fixé par u_{80} . En revanche NONDET ne domine pas complètement NCC puisque, en terme de prudence, NCC obtient de meilleures récompenses que NONDET pour tous les classifieurs.
- **ECLAIR contre NCC** (sous-figures 4.14d, 4.14e, 4.14f) : comme NONDET, ECLAIR est plus pertinent que NCC, et se trouve mieux récompensée avec u_{80} . NCC est toujours meilleure en terme de prudence, à l'exception du jeu de données *glass* pour lequel ECLAIR et NCC ont une récompense *acc* proche.
- **ECLAIR contre NONDET** (sous-figures 4.14g, 4.14h, 4.14i) : nous voyons que les performances de NONDET et ECLAIR sont plus proches. En effet, les points représentant les différentes récompenses obtenues pour des jeux de données, sont pour la plupart autour de la droite identité. Cela s'explique en partie par le fait que ECLAIR peut être considérée comme une extension de NONDET. Lorsqu'aucun exemple d'apprentissage n'est considéré comme suffisamment difficile, le ré-étiquetage n'a pas lieu, et dans ce cas nous retrouvons la procédure de NONDET. Lorsque peu d'exemples sont ré-étiquetés, ECLAIR se différencie moins de NONDET. Nous notons tout de même certaines différences de performances entre les deux classifieurs. En terme de pertinence, NONDET est équivalent à ECLAIR pour les jeux de données utilisés. En terme de prudence, c'est clairement ECLAIR qui a l'avantage pour 3 jeux de données. Du point de vue du compromis fixé par u_{80} , ECLAIR montre une légère avance sur NONDET pour les trois mêmes jeux de données.

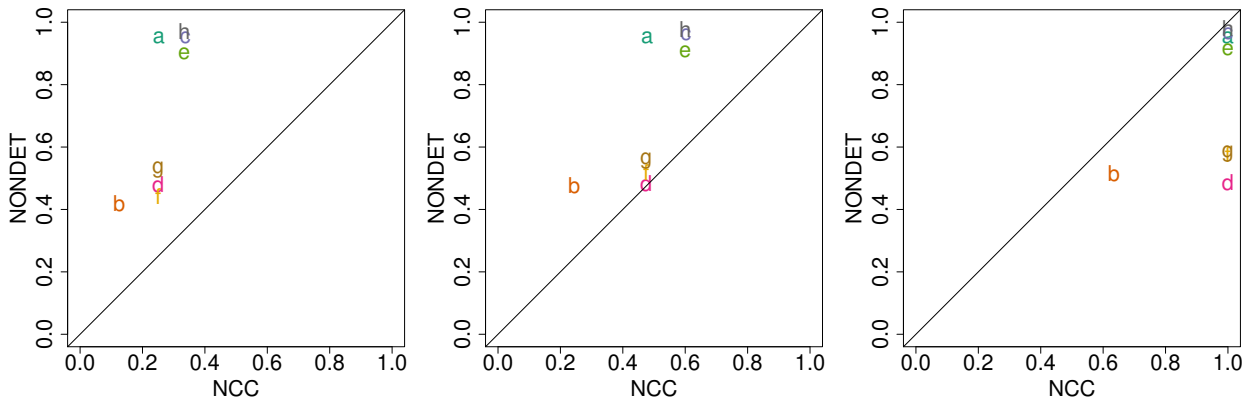
Finalement, l'approche proposée ECLAIR est compétitive avec les deux approches de l'état de l'art NONDET et NCC sur l'ensemble de jeux de données considéré.

4.5 Conclusion du chapitre

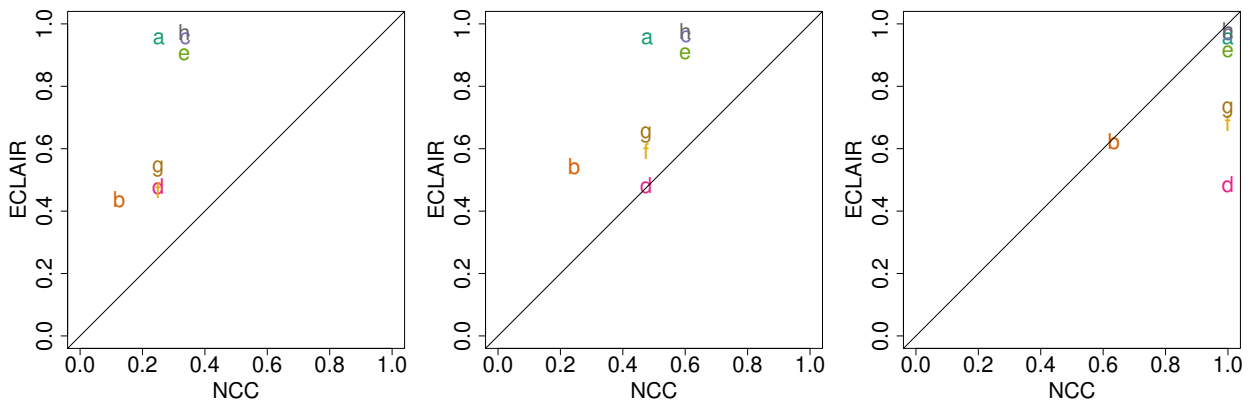
Dans ce chapitre, nous avons exploité les avantages de la classification probabiliste et du cadre des fonctions de croyance pour représenter l'imprécision et l'incertitude de façon générique sous la forme d'une fonction de masse a posteriori. L'approche proposée permet de transformer un classifieur probabiliste existant en un classifieur évidentiel en modifiant l'apprentissage du classifieur (explicitation de l'imprécision présente dans les données). Dans un second temps, nous avons proposé une règle de décision pour fournir une sortie imprécise à partir d'une fonction de masse. Cette règle de décision repose sur une formulation d'un risque en accord avec l'attente de la classification imprécise qui consiste à ne pas prédire un ensemble plus précis que l'observation/l'information disponible ne le permet.

L'approche proposée répond à certains critères énoncés dans le chapitre précédent. D'abord la prise en compte de l'information est adaptée à la présence d'incertitude et/ou d'imprécision. Ensuite l'approche est flexible. En effet, le paramètre β permet de fixer le compromis entre pertinence et prudence requis. L'approche proposée est également générique puisque n'importe quel classifieur peut être employé pour effectuer la classification imprécise. Enfin les expérimentations de ce chapitre montrent que les performances de l'approche proposée sont compétitives avec l'état de l'art pour une sélection de jeux de données.

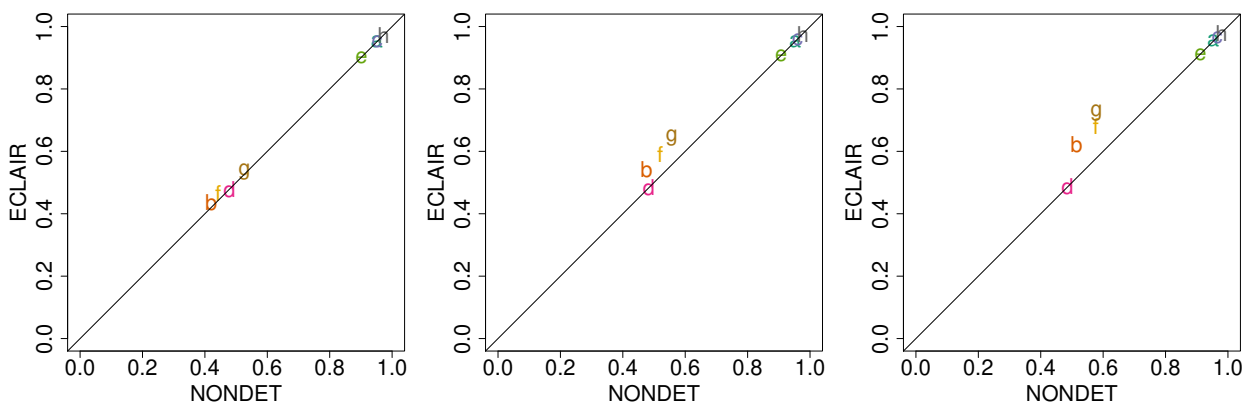
Nous envisageons les perspectives suivantes pour l'approche proposée dans ce chapitre. Premièrement, le critère énoncé dans le chapitre précédent concernant la complexité, n'a pas été



(a) u_{50} NONDET vs NCC (b) u_{80} NONDET vs NCC (c) acc NONDET vs NCC



(d) u_{50} ECLAIR vs NCC (e) u_{80} ECLAIR vs NCC (f) acc ECLAIR vs NCC



(g) u_{50} ECLAIR vs NONDET (h) u_{80} ECLAIR vs NONDET (i) acc ECLAIR vs NONDET



(j) légende des jeux de données

FIG. 4.14 Récompenses u_{50} , u_{80} , et acc obtenues par les approches NONDET, ECLAIR (avec le classifieur Naive Bayes, $\beta = 1$) et NCC sur 8 jeux de données - résultats moyens obtenus sur 10 expériences de validations croisées à 10 blocs

abordé. En l'absence d'algorithmes efficaces pour la minimisation du risque proposé, nous devons parcourir les $2^{|\Theta|} - 1$ sorties imprécises envisageables. C'est à cause de cette complexité exponentielle que nous n'avons pas pu tester l'approche proposée sur des jeux de données impliquant un grand nombre de classes. A ce sujet, une perspective serait d'étudier des algorithmes efficaces pour la minimisation du risque proposée à l'instar de (Mortier et al., 2019) dans le cadre probabiliste. Par ailleurs, nous pourrions envisager d'étendre d'autres fonctions de pertes sur l'espace de définition $2^\Theta \times 2^\Theta$ à partir des mesures d'évaluation de classification imprécise présentées dans le chapitre précédent comme u_{65} et u_{80} . Concernant les performances de l'approche proposée, la qualité des probabilités a posteriori a une grande influence sur les performances des prédictions, nous pourrions envisager une étape de calibration des probabilités pour limiter cette influence. Enfin, notre approche repose essentiellement sur la procédure de ré-étiquetage pour l'explication de l'imprécision des données. Nous envisageons d'étudier de façon encore plus approfondie les avantages et inconvénients des méthodes de ré-étiquetage imprécis proposés.

Notons que l'approche proposée basée sur les fonctions de croyances exploite la présence d'incertitude et d'imprécision pour fournir une sortie prudente. Dans un problème de classification seulement affecté par de l'incertitude, il est préférable d'utiliser un classifieur plus efficace en terme de complexité comme le classifieur NONDET basé sur la théorie des probabilités.

Rappelons que le cadre évidentiel est avantageux dans le problème du tri de plastiques pour les raisons suivantes. D'abord, il permet de manipuler les informations imprécises et incertaines. De plus, il rend possible un large choix de combinaison pour fusionner les informations provenant des pixels à l'échelle de l'objet. Dans le chapitre suivant nous verrons comment intégrer les sorties prudentes (sous forme de fonction de masse ou de sortie imprécises), fournies par le classifieur évidentiel et imprécis ECLAIR dans un processus de tri de plastiques. Dans ce processus, les décisions prudentes correspondent à des bacs dédiés non pas à un plastique spécifique, mais à des ensembles de plastiques (familles de plastiques). Les contenus des bacs les plus précis devraient être maximisés tout en respectant des contraintes connues sur les proportions d'impuretés maximales. Pour le présenter comme une approche plus générale, le processus de tri sera formalisé comme un mélange séquentiel de composés avec des contraintes sur les proportions de certains composés.

Manipulation de vecteurs de proportions imprécises et incertaines :
Application au tri des plastiques sombres.

Sommaire

5.1	Éléments théoriques pour la gestion d’information évidentielle sur des vecteurs de proportions	115
5.1.1	Vecteurs de proportions imprécises et/ou incertaines	116
5.1.2	Vecteur de proportions d’un mélange	119
5.1.3	Inférence sur le respect des contraintes	124
5.2	Stratégies de tri	126
5.2.1	Mise à jour de la connaissance concernant la composition du bac	128
5.2.2	Inférence sur le respect des contraintes dans le bac positif	129
5.2.3	Complexité	131
5.3	Illustration	132
5.3.1	Présentation des données disponibles pour le problème de tri	132
5.3.2	Simulations	133
5.4	Discussion	139
5.4.1	Généralisation au cas de fragments qui ne sont pas monomatières	139
5.4.2	Généralisation au cas de plusieurs bacs	140
5.5	Conclusion du chapitre	142

Dans le chapitre précédent, nous avons proposé une approche permettant de construire un classifieur évidentiel à partir d’un classifieur probabiliste. Nous avons ensuite proposé une procédure de décision permettant de transformer la fonction de masse fournie par le classifieur évidentiel en sortie imprécise. La procédure complète (classification évidentielle puis procédure de décision) constitue donc un classifieur évidentiel imprécis. La procédure de décision nous a permis d’évaluer les capacités prédictives des fonctions de croyance fournies. Nous pourrions employer cette même procédure telle quelle pour résoudre le problème du tri de fragments de plastiques. Dans ce cas, nous soufflerions chaque fragment arrivant en continu, dans le bac dédié à l’ensemble de plastiques prédit par le classifieur évidentiel imprécis. Cependant cette procédure de décision n’est pas optimisée pour répondre aux attentes du tri de plastiques. En effet, dans le problème du tri, les fragments doivent être traités séquentiellement dans l’ordre où ils sont amenés par le tapis roulant, et nous devons respecter des contraintes dans les bacs

sur des proportions d'impuretés fixées par des experts. Notons que nous faisons l'hypothèse d'un flux de fragments arrivant en continu indéfiniment. Nous verrons dans ce chapitre que la stratégie, i.e., le comportement du décideur au cours du remplissage, a une réelle influence sur la qualité, i.e., le respect des contraintes, et le rendement du tri, i.e., la vitesse de remplissage des bacs. L'objectif du tri est double, d'une part, nous devons maximiser le remplissage des bacs dédiés aux plastiques précis (*ABS*, *HIPS*, etc.), i.e., minimiser la quantité de plastiques soufflés dans le bac de rejet réservé aux fragments dont le classifieur ignore la nature précise. D'autre part, le contenu des bacs doit respecter des seuils d'impuretés imposés sur les proportions de certains types de plastiques.

Pour plus de généralité lors de la présentation des éléments théoriques, nous parlerons de "pièce", au lieu de "fragment", de "mélange de pièces" au lieu de "ensemble de fragments contenus dans un bac" et de "composant" au lieu de "type de plastiques". Nous parlerons également de "pièce" pour désigner une pièce composée de plusieurs composants et de "pièce monomatière" lorsqu'une pièce est composée d'un unique composant. Dans cette thèse les fragments sont considérés comme des pièces monomatières, cependant la partie théorique de ce chapitre est exposée pour le cas plus général de pièces qui ne sont pas forcément monomatières.

Dans ce chapitre nous nous intéressons à la phase de décision du problème de tri de fragments de plastiques à partir d'informations sous la forme de fonctions de masse telles que fournies par le classifieur évidentiel dans le cas monomatière (chapitre précédent). A cette fin, nous avons besoin d'estimer les proportions de chacun des plastiques présents dans les bacs de tri. Sous une forme plus générale, ce problème est celui d'un mélange séquentiel (ajouts successifs de pièces dans un mélange) sous contraintes sur les proportions du mélange à partir d'informations évidentielles. Chaque bac de tri est donc vu comme un mélange de fragments dont la nature est décrite par des informations évidentielles. Puisque nous nous intéressons maintenant aux proportions des plastiques contenus dans les bacs, nous sommes amenés à manipuler des informations évidentielles décrivant ces proportions dans un espace continu. Habituellement, les fonctions de croyance sont définies sur des espaces discrets finis, mais de nombreux problèmes nécessitent de maintenir la croyance sur un espace continu, même si elle est basée sur un nombre fini et discret d'informations, voir par exemple ([Pellicanò et al., 2018](#)) qui traite de l'espace réel bivarié. Nos fonctions de masse utilisées pour décrire des proportions dans un mélange de pièces pourraient être utiles dans d'autres domaines que celui du tri de plastiques, comme la surveillance de polluants, ou le contrôle de la qualité dans l'agro-alimentaire. Nous présentons donc les éléments théoriques permettant la manipulation de fonctions de croyance sur des vecteurs de proportions dans le cadre général d'un mélange séquentiel de pièces, avant de nous intéresser au problème spécifique du tri des plastiques avec l'hypothèse de pièces monomatières.

L'organisation du chapitre est la suivante. Dans une première section [5.1](#), nous donnons des éléments théoriques pour représenter les proportions d'un mélange de pièces avec les fonctions de croyance. Nous verrons notamment comment modéliser le respect de contraintes sur des proportions décrites par des fonctions de masse. Dans la section suivante [5.2](#), nous nous baserons sur les éléments théoriques concernant la manipulation de vecteurs de proportions pour proposer des stratégies de tri. Nous illustrerons ces stratégies de tri dans la section [5.3](#) avec des simulations de tri réalisées à partir de données réelles sur des fragments de plastiques. Enfin, dans la section [5.4](#), nous donnerons une suggestion pour obtenir des informations permettant de traiter le cas de fragments qui ne sont pas monomatières, puis nous proposerons des éléments de résolution du problème de tri de plastiques monomatières généralisé au cas de plusieurs bacs.

5.1 Éléments théoriques pour la gestion d'information évidentielle sur des vecteurs de proportions

Dans cette section, nous présentons des éléments théoriques nécessaires basés sur les fonctions de croyance, pour manipuler des informations imprécises et/ou incertaines à propos de vecteurs de proportions. Comme mentionné dans l'introduction, nous restons dans le cadre général en considérant des vecteurs de proportions pour des pièces, des éclats de roches par exemple, auxquelles peuvent être associées des proportions de composants minéraux dans un sol donné (ou encore la quantité de polluants à surveiller dans un sol ou une nappe phréatique). Ces vecteurs de proportions peuvent encore correspondre aux concentrations de différents composants dans un mélange (dont la composition doit être maîtrisée), ou encore à des histogrammes mal connus (par exemple, des pixels d'une couleur donnée dans une image). Nous considérons que les pièces étudiées sont composées de composants appartenant à un espace fini Θ . Par abus de langage nous désignons par "poids" la masse physique des pièces afin qu'il n'y ait pas de confusion possible avec les masses de fonctions de croyances.

Commençons par préciser ce que nous désignons par *vecteur de proportions*. Un vecteur de proportions associé à une pièce est un vecteur $\tau = (\tau_1, \dots, \tau_n) \in [0, 1]^n$ tel que chaque τ_i , $0 \leq i \leq n$, est la proportion de la composante $\theta_i \in \Theta$ dans la pièce.

Exemple 1. Prenons un exemple de proportions dans le domaine de la cuisine. Dans la recette de la crème aux œufs, les ingrédients sont $1/2$ L de lait, 100g de sucre et 3 œufs. Nous avons $\Theta = \{\text{lait, sucre, œuf}\}$. En supposant qu'un œuf représente 55g, nous pouvons déduire que le vecteur de proportions associé à la pièce "crème aux œufs" est $\tau \simeq (0.65, 0.13, 0.22)$.

Par définition, un vecteur de proportions τ est inclus dans le simplexe n -unitaire, c'est-à-dire l'ensemble de vecteurs positifs de n éléments sommant à un :

$$\mathbb{U} = \{(\tau_1, \dots, \tau_n) \in [0, 1]^n : \sum_{i=1}^n \tau_i = 1\}.$$

Dans la suite de la section, nous utilisons la représentation simplifiée du simplexe présentée dans la Fig. 5.1 pour illustrer nos exemples avec $n = 3$.

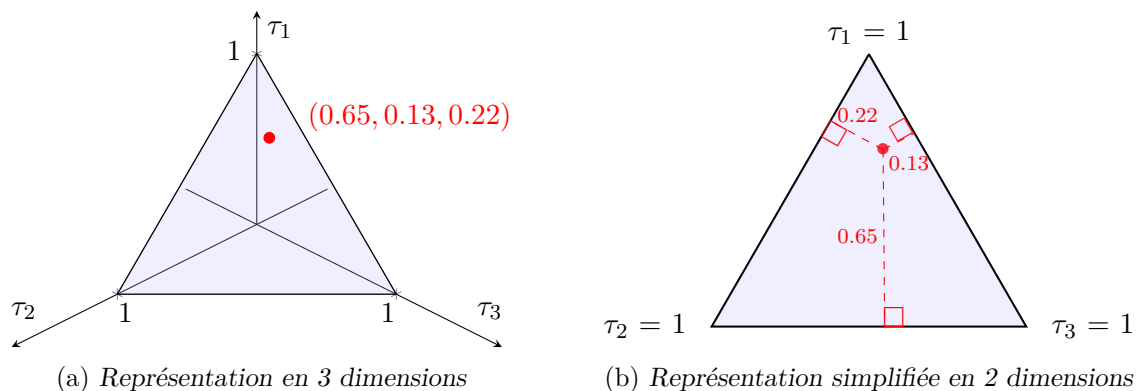


FIG. 5.1 Représentation du vecteur de proportions de la crème aux œufs $\tau \simeq (0.65, 0.13, 0.22)$ dans le simplexe unitaire

Dans cette section, nous nous intéressons au traitement des informations relatives aux vecteurs de proportions. Il s'agit de caractériser l'information concernant le mélange d'un ensemble de pièces à partir d'informations concernant les vecteurs de proportions de chaque pièce. Par

exemple, quel est le vecteur de proportions associé au mélange des trois pièces suivantes $1/2 L$ de lait, $100g$ de sucre et 3 œufs ? Dans ce cas, la réponse est simple (exemple 1) car les informations données sont parfaitement connues (précises et certaines) : avec $\Theta = \{\text{lait}, \text{sucre}, \text{œuf}\}$, le mélange des trois pièces respectivement caractérisées par des poids de $500g$, $100g$, et $165g$ et des vecteurs de proportions, $(1, 0, 0)$, $(0, 1, 0)$ et $(0, 0, 1)$, est une nouvelle pièce, la recette de la crème aux œufs, caractérisée par un poids de $765g$ et le vecteur de proportions $(0.65, 0.13, 0.22)$. L'opération est moins évidente lorsque l'information concernant les vecteurs de proportions n'est plus parfaitement connue. Par exemple si l'on décide de mélanger la crème aux œufs maison caractérisée par le vecteur de proportions ci-avant parfaitement connu avec de la crème aux œufs industrielle (dont le dosage des ingrédients précis est gardé secret), il n'est alors plus possible de caractériser précisément les proportions de la nouvelle recette. Supposons par exemple que nous savons avec certitude que la proportion de sucre dans la recette industrielle est dans l'intervalle $[0.2, 0.4]$. Nous pouvons en déduire que la proportion de sucre dans le mélange de $100g$ de crème aux œufs maison et de $100g$ de crème aux œufs industrielle est dans l'intervalle $[0.165, 0.265]$. Si nous ne sommes pas certain que la proportion de sucre dans la recette industrielle est dans l'intervalle $[0.2, 0.4]$, alors comment cette incertitude se traduit-elle sur la proportion en sucre du mélange ? Dans la suite de la section, nous formalisons, l'opération correspondant au mélange de plusieurs pièces lorsque les informations disponibles ne sont plus parfaitement connues (présence d'imprécision et d'incertitude). Nous nous intéressons d'autre part à faire de l'inférence concernant certains événements qui peuvent être utilisés pour surveiller les proportions de composants dans un mélange. Dans notre exemple de la crème aux œufs, nous pourrions être intéressés, pour des raisons de santé, par l'évènement "la proportion de sucre est inférieure à 0.3".

Dans les prochaines sous-sections, nous clarifions les formes d'imprécision et d'incertitude que nous considérons et présentons dans le cadre des fonctions de croyance : le traitement de l'imprécision (intervalles), indépendante des fonctions de croyance, et le traitement de l'incertitude. Ensuite, nous présentons nos propositions d'opérations caractérisant les informations résultant de mélanges de différentes pièces. Enfin, nous présentons nos outils qui permettent d'inférer de l'information sur des événements d'intérêt, i.e., le respect de contraintes sur les proportions.

5.1.1 Vecteurs de proportions imprécises et/ou incertaines

Dans le cas général, le traitement de proportions inconnues peut être difficile car les éléments focaux peuvent être infinis et prendre diverses formes. Par exemple, nous aurions pu dans l'absolu, considérer des proportions imprécises exprimées comme des unions d'intervalles telles que $[0, 0.2] \cup [0.8, 1]$, ou des proportions imprécises exprimées par une infinité de valeurs disjointes, comme tous les nombres rationnels dans $[0, 1]$, etc.. En pratique, il est donc nécessaire de simplifier le problème et de traiter des sous-ensembles spécifiques de l'espace initial continu. Ainsi nous considérerons que les informations imprécises sur les proportions s'expriment sous forme d'intervalles spécifiques de proportions. Nous simplifions la représentation de l'imprécision sur les proportions en ne considérant pas des vecteurs imprécis mais des vecteurs dont les composantes (proportions) sont imprécises. Dans la suite, nous montrons comment nous avons choisi d'exprimer dans notre étude, les informations incertaines et imprécises concernant des vecteurs de proportions. Notons que dans tous les cas, nous connaissons parfaitement le poids physique, noté $w \in \mathbb{R}_+^*$ de la pièce en question. Dans le cas de notre application au tri de plastiques, ce poids correspond à la quantité de matière d'un fragment : il est estimé par le nombre de pixels du fragment dans son image hyperspectrale. Dans d'autres domaines, il peut s'agir d'un extrait rocheux, d'un volume d'air ou d'eau, etc ..

Pour représenter les informations concernant les vecteurs de proportions, nous proposons une représentation par intervalles et une représentation dans le cadre des fonctions de croyance.

Définition 1. Vecteur de proportions imprécises Un vecteur de proportions imprécises s'exprime comme :

$$\mathcal{I} := (\mathcal{I}_1, \dots, \mathcal{I}_n)$$

où $\mathcal{I}_i, i \in \{1, \dots, n\}$ est une information imprécise sur la proportion de $\theta_i \in \Theta$ dans une pièce. Notons que nous introduisons la notation $(\mathcal{I}_1, \dots, \mathcal{I}_n)$ pour simplifier les notations mais qu'il s'agit en réalité du produit cartésien $\mathcal{I}_1 \times \dots \times \mathcal{I}_n$. Ces informations imprécises $\mathcal{I}_i, i \in \{1, \dots, n\}$ considérées ici sont exprimées sous la forme d'intervalles $[\ell_i, u_i] \subseteq [0, 1]$ signifiant :

"La proportion τ_i du matériau de type θ_i dans la pièce est comprise entre ℓ_i et u_i ".

Étant donné que les intervalles \mathcal{I}_i doivent décrire de vraies proportions, nous exigeons les deux conditions suivantes proposées originellement par (De Campos et al., 1994) au sujet des probabilités imprécises, pour vérifier qu'un ensemble crédal ne soit pas vide et que ses bornes soient bien atteintes par des probabilités de l'ensemble crédal. Dans notre cas, ces conditions s'appliquent au vecteur \mathcal{I} de proportions imprécises :

- Le vecteur \mathcal{I} de proportions imprécises doit recouvrir au moins un vecteur de proportions précises. En d'autres termes, il doit exister au moins un vecteur de proportions précises (τ_1, \dots, τ_n) tel que $\tau_i \in [\ell_i, u_i]$ pour tous les $i \in \{1, \dots, n\}$. Cette condition est satisfaite si et seulement si :

$$\sum_{i=1}^n \ell_i \leq 1 \leq \sum_{i=1}^n u_i. \quad (28)$$

Cette condition permet d'éviter que \mathcal{I} ne contienne aucun vecteur de proportions réaliste, i.e., les vecteurs de \mathbb{U} .

- Le vecteur \mathcal{I} doit être tel que les bornes des intervalles \mathcal{I}_i soient atteignables. Cela signifie qu'il doit exister au moins un vecteur de proportions (τ_1, \dots, τ_n) recouvert par \mathcal{I} avec $\tau_i = \ell_i$ ou $\tau_i = u_i$, et ce pour chaque $i \in \{1, \dots, n\}$. Cette condition est satisfaite si et seulement si, $\forall i \in \{1, \dots, n\}$:

$$u_i + \sum_{j \neq i} \ell_j \leq 1 \text{ et } \ell_i + \sum_{j \neq i} u_j \geq 1. \quad (29)$$

Notons que lorsque les limites proviennent de quantités physiques, ces exigences sont automatiquement satisfaites.

Exemple 2. Soit $\Theta = \{\theta_1, \theta_2, \theta_3\}$. Prenons $\mathcal{I} = ([0, 1/2], [0, 1/2], [0, 1/2])$ un exemple de vecteur de proportions imprécises. Notons que ce vecteur correspond à l'évènement "aucune proportion n'est majoritaire". Nous pouvons facilement vérifier que \mathcal{I} respecte bien les deux conditions. La première condition est respectée car il existe un vecteur de proportions précises qui vérifie la contrainte (28), par exemple $(0, 1/2, 1/2)$. La condition (29) est également respectée car toutes les bornes sont accessibles : $(0, 1/2, 1/2)$ pour la borne inférieure de la première proportion, $(1/2, 1/2, 0)$ pour la borne supérieure de la première proportion, etc.

En revanche, le vecteur $J = ([1/2, 1], [1/2, 1], [1/2, 1])$ n'est pas considéré comme un vecteur de proportions imprécises car, par exemple, la première borne $1/2$ n'est pas accessible.

Remarque 1. Dans certaines applications, et en particulier pour le tri de plastiques, la pièce à ajouter au mélange peut-être supposée monomatière. Dans un tel cas, une information imprécise

sur la composition d'une pièce ne s'exprime qu'avec des *proportions binaires*, i.e., des proportions dans l'ensemble $\{0, 1\}$, 0 si le composant n'est pas présent dans la pièce et 1 s'il constitue la pièce. En effet, considérons l'information imprécise suivante concernant une pièce monomatière :

"Le type de matériau de la pièce est l'un des types de l'ensemble E inclus dans Θ ."

Traduisant cette dernière information en terme de proportions imprécises, nous obtenons, pour chaque $i \in \{1, \dots, n\}$:

$$\mathcal{I}_i = \begin{cases} \{0\} & \text{si } \theta_i \notin E, \\ \{1\} & \text{si } E = \{\theta_i\}, \\ \{0, 1\} & \text{si } \theta_i \in E \text{ et } |E| > 1. \end{cases}$$

Exemple 3. Par exemple dans le cas d'un fragment monomatière de plastiques avec $\Theta = \{ABS, HIPS, PE, PP\}$, l'information "le fragment est un monomatière ABS ou un monomatière HIPS" s'exprime comme $\mathcal{I} = (\{0, 1\}, \{0, 1\}, \{0\}, \{0\})$. En effet, nous devons représenter les deux possibilités, dans le cas où le fragment est un monomatière ABS, nous aurions le vecteur de proportions précises $(1, 0, 0, 0)$, s'il s'agissait un monomatière HIPS, alors nous aurions $(0, 1, 0, 0)$. Les proportions de PE et PP sont nulles dans les deux cas, donc représentées par $\{0\}$ dans le vecteur. Les proportions de ABS et HIPS, elles peuvent être soit nulles, soit maximales, elles sont donc représentées par $\{0, 1\}$ dans le vecteur.

Remarque 1. (suite) Remarquons que cette forme de vecteur n'est pas conforme à notre définition 1 des vecteurs imprécis (nous n'avons pas d'intervalles mais des singletons). Cependant, puisque seuls les points extrêmes interviennent dans les manipulations et inférences considérées dans ce chapitre, nous considérerons l'enveloppe convexe¹ :

$$\mathcal{I}_i = \begin{cases} [0, 0] & \text{si } \theta_i \notin E, \\ [1, 1] & \text{si } E = \{\theta_i\}, \\ [0, 1] & \text{si } \theta_i \in E \text{ et } |E| > 1. \end{cases}$$

Cette considération ne changera pas les calculs ni les résultats et simplifiera considérablement notre exposé.

Poursuivons l'exemple 3. D'après la remarque 1. (suite), nous représenterons donc l'information "le fragment est un monomatière ABS ou un monomatière HIPS" par le vecteur imprécis $\mathcal{I} = ([0, 1], [0, 1], [0, 0], [0, 0])$.

Définition 2. Vecteur de proportions entachées d'imprécision et d'incertitude

La notion de *vecteur de proportions entachées d'imprécision et d'incertitude* est une généralisation de celle d'un *vecteur de proportions imprécises*. Elle concerne un vecteur de proportions inconnues pour lequel les informations disponibles sont un ou plusieurs vecteur de proportions imprécises associés à des degrés de certitude.

Nous caractérisons un tel vecteur par une fonction de masse m définie sur $[0, 1]^n$ pour laquelle les éléments focaux autorisés sont des vecteurs de proportions imprécises. La fonction de masse $m : 2^{[0,1]^n} \rightarrow [0, 1]$ est ainsi définie telle que pour tout $A \subseteq [0, 1]^n$:

$$\begin{aligned} m(A) &\geq 0 \text{ si } A \text{ est un vecteur de proportions imprécises,} \\ m(A) &= 0 \text{ sinon.} \end{aligned}$$

1. Un ensemble convexe est tel que pour n'importe quels de ses points a, b , tous les points du segment $[a, b]$ appartiennent aussi à l'ensemble. L'enveloppe convexe d'un objet est le plus petit ensemble convexe qui contient cet objet.

Nous notons (m, \mathbb{F}) le corps d'évidence où $\mathbb{F} \subseteq 2^{[0,1]^n}$ est un sous-ensemble fini de vecteurs de proportions imprécises. Dans le cas de l'ignorance totale, nous avons la masse vide $m?$, qui alloue une masse de croyance maximale au vecteur de proportions imprécises $[0, 1]^n$: $m?([0, 1]^n) = 1$.

Exemple 4. *Considérons $\Theta = \{\theta_1, \theta_2, \theta_3\}$. Prenons l'exemple d'une pièce dont l'information sur la proportion de chacun des θ_i est représentée par la fonction de masse suivante : $m(\left([0, 1/2], [0, 1/2], [0, 1/2]\right)) = 0.4$ et $m(\left([0, 1], [0, 1], [0, 1]\right)) = 0.6$. Les vecteurs de proportions imprécises sont donnés dans la Fig. 5.2. Cette fonction de masse signifie que la masse allouée exactement à l'évènement "aucune proportion n'est majoritaire" est 0.4 et la masse allouée exactement à l'évènement "ignorance totale (toutes les proportions sont possibles)" est 0.6.*

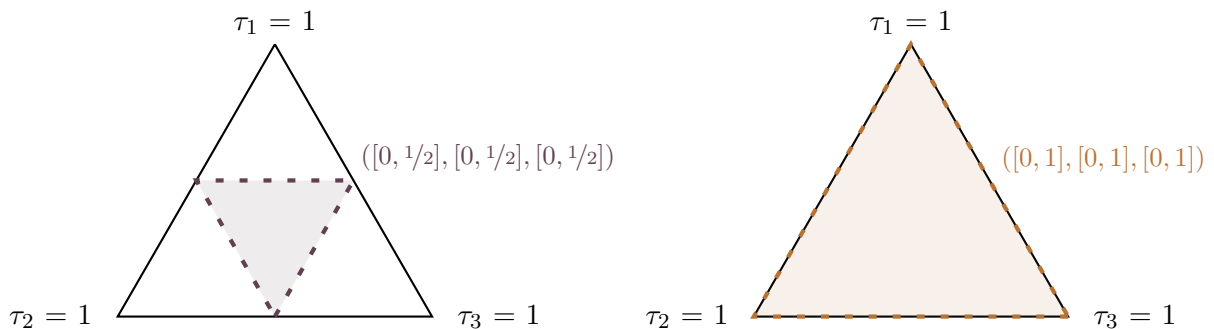


FIG. 5.2 Représentation des vecteurs de proportions imprécises dans le simplexe unitaire.

Dans la suite du chapitre, nous distinguerons trois cas différents en fonction du type d'information pris en compte pour caractériser un vecteur de proportions :

- Cas déterministe : lorsque nous disposons d'informations certaines et précises sous la forme d'un vecteur précis $\tau \in \mathbb{U}$.
- Cas probabiliste : lorsque l'information disponible est une distribution de probabilités discrète définie sur $\mathbb{U} : p : \mathbb{U} \rightarrow [0, 1]$ où seulement des sous-ensembles discrets et finis d'éléments de \mathbb{U} ont une probabilité non nulle. Par exemple, $p(\left(1/3, 1/3, 1/3\right)) = 0.4$ et $p(\left(0, 0, 1\right)) = 0.6$.
- Cas d'imprécision sous forme d'intervalles : lorsque nous disposons d'informations imprécise sous la forme d'un intervalle pour chaque composante. Il s'agit du cas de vecteur de proportions imprécises de la définition 1.
- Cas évidentiel : lorsque nous disposons d'un vecteur de proportions entachées d'imprécision et d'incertitude caractérisé par une fonction de masse telle que définie dans la définition 2.

Cette distinction nous permettra de faire des comparaisons entre des stratégies de tri dans la section 5.2.

5.1.2 Vecteur de proportions d'un mélange

Nous nous intéressons à présent à la représentation de l'information sur un mélange de pièces en fonction des types d'informations disponibles sur chacune des pièces.

5.1.2.1 Cas déterministe

Dans le cas déterministe, l'information concernant le vecteur de proportions d'un mélange est trivial. En effet, en supposant les informations suivantes concernant deux pièces :

- les poids w^1 et w^2 ,
- les vecteurs de proportions précisément connus $\tau^1 = (\tau_1^1, \dots, \tau_n^1)$ et $\tau^2 = (\tau_1^2, \dots, \tau_n^2)$,

nous déduisons que le poids du mélange des deux pièces est $w^1 + w^2$ et que son vecteur de proportions est composé des moyennes des proportions de chaque pièce pondérées par les poids des pièces. Ainsi, le mélange est caractérisé par la notation et le vecteur suivants :

$$\tau^{1 \boxplus} \tau^2 = \left(\frac{\tau_1^1 w^1 + \tau_1^2 w^2}{w^1 + w^2}, \dots, \frac{\tau_n^1 w^1 + \tau_n^2 w^2}{w^1 + w^2} \right).$$

5.1.2.2 Cas probabiliste

Dans le cas probabiliste, il suffit de calculer la probabilité de chaque vecteur de proportions résultant d'une combinaison des vecteurs de proportions ayant des probabilités non nulles des deux pièces mélangées.

Exemple 5. *Considérons $\Theta = \{\theta_1, \theta_2\}$ et deux pièces de même poids 1kg. Supposons que le vecteur de proportions de la première pièce soit caractérisé par la distribution de probabilité p^1 telle que :*

$$p^1((1/2, 1/2)) = 0.5 \text{ et } p^1((1, 0)) = 0.5,$$

et que le vecteur de proportions de la seconde pièce soit caractérisé par la distribution de probabilité p^2 identique à p^1 :

$$p^2((1/2, 1/2)) = 0.5 \text{ et } p^2((1, 0)) = 0.5.$$

La distribution de probabilité caractérisant le vecteur de proportions du mélange de ces deux pièces, noté $p^{1 \boxplus 2}$ est :

$$p^{1 \boxplus 2}((1/2, 1/2)) = 0.5 \times 0.5 = 0.25,$$

*(probabilité que la première **et** la seconde pièce aient le vecteur de proportions $(1/2, 1/2)$)*

$$p^{1 \boxplus 2}((3/4, 1/4)) = 0.5 \times 0.5 + 0.5 \times 0.5 = 0.5,$$

*(probabilité que la première ait le vecteur de proportions $(1/2, 1/2)$ **et** que la seconde ait le vecteur de proportions $(1, 0)$ **ou** inversement)*

$$p^{1 \boxplus 2}((1, 0)) = 0.5 \times 0.5 = 0.25$$

*(probabilité que la première **et** la seconde pièce aient le vecteur de proportions $(1, 0)$)*

Nous remarquons dans cet exemple que le mélange de deux pièces dont les proportions sont caractérisées par la même distribution de probabilité, est caractérisé par une distribution différente. Cette remarque est importante pour la suite, lorsque nous nous intéresserons à la confiance en des événements spécifiques. Supposons que nous nous intéressions à l'évènement "la pièce est monomatière de type θ_1 ". La probabilité de cet évènement est 0.5 pour les deux pièces. Lorsque les deux pièces sont mélangées, la probabilité de cet évènement n'est plus que 0.25.

De façon générale, considérons deux pièces avec les poids w^1 et w^2 et les distributions de probabilités p^1 et p^2 . Le mélange de ces deux pièces est caractérisé par le poids $w^1 + w^2$ et par la distribution de probabilité $p^{1\boxplus 2}$ suivante. Pour tout vecteur de proportion τ ,

$$p^{1\boxplus 2}(\tau) = \sum_{\tau^1 \boxplus \tau^2 = \tau} p^1(\tau^1) \times p^2(\tau^2)$$

où les vecteurs τ^1 (resp. les vecteurs τ^2) concernent seulement les vecteurs ayant la probabilité non-nulle avec p^1 (resp. p^2).

5.1.2.3 Cas imprécis sous forme d'intervalles

Nous donnons dans la suite la définition de la somme de vecteurs de proportions imprécises correspondant au mélange de deux pièces pour lesquelles nous disposons d'informations imprécises sur leurs proportions. Nous verrons ensuite comment cette somme se généralise au cas de vecteur de proportions entachées d'imprécision et d'incertitude.

Définition 3. Somme de vecteurs de proportions imprécises Considérons que nous avons une information initiale w^1 , $\mathcal{I}^1 = ([\ell_1^1, u_1^1], \dots, [\ell_n^1, u_n^1])$ représentant le poids total et le vecteur de proportions imprécises connu d'une première pièce, et que nous voulons obtenir la nouvelle information suite à l'ajout d'une autre pièce de poids w^2 , et de vecteur de proportions imprécises $\mathcal{I}^2 = ([\ell_1^2, u_1^2], \dots, [\ell_n^2, u_n^2])$. Le poids de leur mélange est $w^1 + w^2$, et le nouveau vecteur de proportions imprécises du mélange est défini comme suit :

$$\mathcal{I}^1 \boxplus \mathcal{I}^2 := (\mathcal{K}_1, \dots, \mathcal{K}_n) \tag{30}$$

où pour chaque $i = 1, \dots, n$:

$$\mathcal{K}_i = \left[\frac{w^1}{w^1 + w^2} \ell_i^1 + \frac{w^2}{w^1 + w^2} \ell_i^2, \frac{w^1}{w^1 + w^2} u_i^1 + \frac{w^2}{w^1 + w^2} u_i^2 \right].$$

Exemple 6. *Considérons que nous avons trois types de matériaux $\theta_1, \theta_2, \theta_3$ et que nous souhaitons suivre l'état de notre connaissance sur la composition d'un mélange dans un bac lorsque nous y ajoutons une pièce. Supposons que le bac contienne initialement $w^1 = 1\text{kg}$ de matière et que l'information sur sa composition soit imprécise : une proportion $\tau_1 \in [1/4, 1/2]$ de θ_1 , une proportion $\tau_2 \in [0, 1/3]$ de θ_2 et une proportion $\tau_3 \in [1/6, 3/4]$ de θ_3 . Concernant la nouvelle pièce à ajouter au bac, supposons qu'elle ait un poids $w^2 = 1\text{kg}$ et que les proportions des trois matériaux soient connues de façon imprécise : $\tau_1 \in [0, 1/2]$, $\tau_2 \in [0, 1/2]$, $\tau_3 \in [0, 1/2]$ (aucun matériau n'est majoritaire dans la composition de la pièce). Les vecteurs de proportions imprécises, $\mathcal{I}^1 = ([1/4, 1/2], [0, 1/2], [1/6, 3/4])$ décrivant le bac et $\mathcal{I}^2 = ([0, 1/2], [0, 1/2], [0, 1/2])$ décrivant la nouvelle pièce sont représentés dans la Fig. 5.3.*

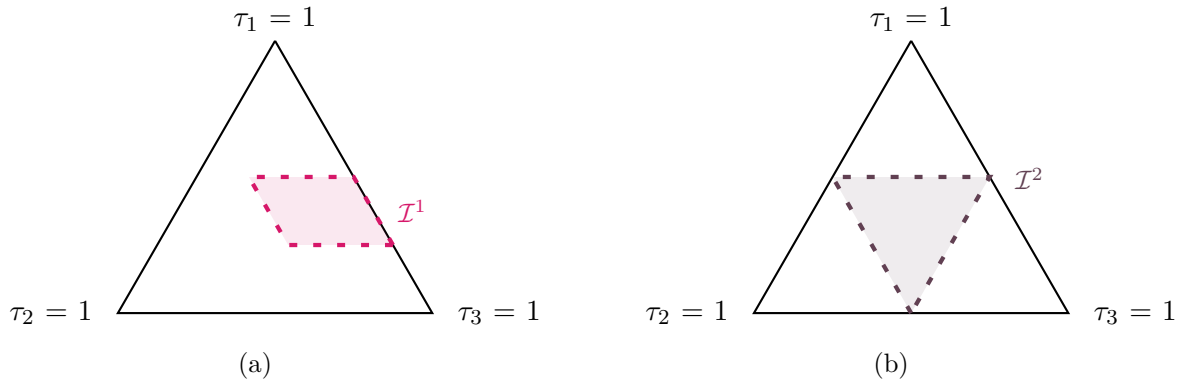


FIG. 5.3 Représentation du vecteur de proportions imprécises décrivant la composition du bac 5.3a, et de la nouvelle pièce 5.3b dans le simplexe.

Lorsque la pièce est ajoutée au bac, nous disposerons des informations imprécises suivantes sur la nouvelle composition du bac :

$$\mathcal{I}^1 \boxplus \mathcal{I}^2 = \left(\left[\frac{1}{8}, \frac{1}{2} \right], \left[0, \frac{5}{12} \right], \left[\frac{1}{12}, \frac{5}{8} \right] \right).$$

Ce dernier vecteur de proportions imprécises est représenté dans la Fig. 5.4.

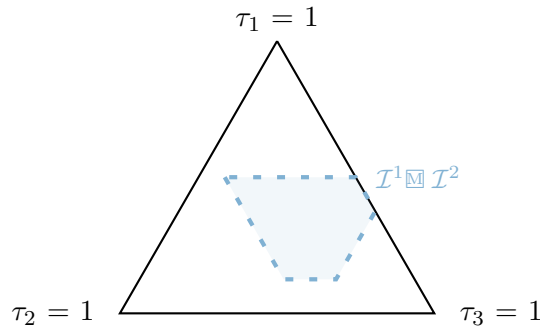


FIG. 5.4 Représentation du vecteur de proportions imprécises décrivant la composition du mélange (bac + nouvelle pièce) dans le simplexe.

5.1.2.4 Cas évidentiel

Définition 4. Somme de fonctions de masse représentant l'information sur des vecteurs de proportions Étant donné deux corps d'évidence (m^1, \mathbb{F}^1) et (m^2, \mathbb{F}^2) associés aux poids w^1 et w^2 de deux pièces, nous définissons la masse $m^{1 \boxplus 2} : 2^{[0,1]^n} \rightarrow [0, 1]$ résultant du mélange des deux pièces pour $\mathcal{I} \subseteq [0, 1]^n$ comme suit :

$$m^{1 \boxplus 2}(\mathcal{I}) = \sum_{\substack{\mathcal{I}' \in \mathbb{F}^1, \mathcal{I}'' \in \mathbb{F}^2 \\ \mathcal{I} = \mathcal{I}' \boxplus \mathcal{I}''}} m^1(\mathcal{I}') \cdot m^2(\mathcal{I}''). \quad (31)$$

où \boxplus est défini dans la formule (30) et le poids total associé à la fonction de masse $m^{1 \boxplus 2}$ est $w^1 + w^2$. Cette opération peut bien sûr être répétée de manière itérative si nous avons plusieurs pièces à mélanger. Il convient de noter que si toutes les fonctions de masse ont un nombre fini

d'éléments focaux, il en est de même pour $m^{1\textcircled{M}2}$, ce nombre étant au maximum $|\mathbb{F}^1| \cdot |\mathbb{F}^2|$. En pratique, il est impératif d'approximer les résultats pour contrôler le nombre d'éléments focaux si ce nombre devient trop grand.

Exemple 7. Reprenons l'exemple 6, en considérant des informations évidentielles au lieu des simples vecteurs de proportions imprécises. La TABLE 5.1 donne les information concernant le bac et la nouvelle pièce.

TABLE 5.1 – Les deux informations évidentielles

bac ($w^1 = 1kg$)		Nouvelle pièce ($w^2 = 1kg$)	
\mathbb{F}^1	m^1	\mathbb{F}^2	m^2
\mathcal{I}'_1	0.5	\mathcal{I}''_1	0.6
$([1/4, 1/2], [0, 1/3], [1/6, 3/4])$		$([0, 1/2], [0, 1/2], [0, 1/2])$	
\mathcal{I}'_2	0.5	\mathcal{I}''_2	0.4
$([0, 1/6], [0, 1/6], [5/6, 1])$		$([0, 1/6], [0, 1/6], [5/6, 1])$	

Lorsque la nouvelle pièce est ajoutée au bac, nous obtenons les éléments focaux suivants pour $m^{1\textcircled{M}2}$:

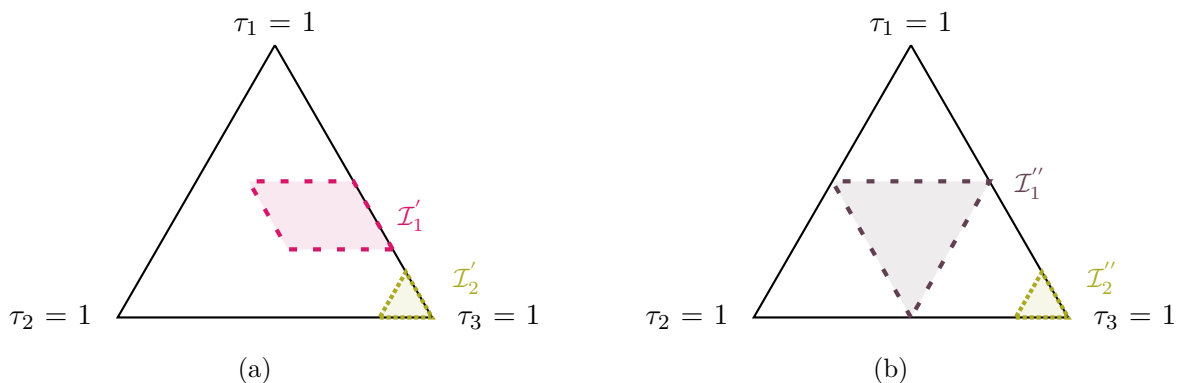
$$\mathcal{I}'_1 \boxtimes \mathcal{I}''_1 = ([1/8, 1/2], [0, 5/12], [1/12, 5/8]), \quad m^{1\textcircled{M}2}(\mathcal{I}'_1 \boxtimes \mathcal{I}''_1) = 0.3,$$

$$\mathcal{I}'_1 \boxtimes \mathcal{I}''_2 = ([1/8, 1/3], [0, 1/4], [1/2, 7/8]), \quad m^{1\textcircled{M}2}(\mathcal{I}'_1 \boxtimes \mathcal{I}''_2) = 0.2,$$

$$\mathcal{I}'_2 \boxtimes \mathcal{I}''_1 = ([0, 1/3], [0, 1/3], [5/12, 3/4]), \quad m^{1\textcircled{M}2}(\mathcal{I}'_2 \boxtimes \mathcal{I}''_1) = 0.3,$$

$$\mathcal{I}'_2 \boxtimes \mathcal{I}''_2 = ([0, 1/6], [0, 1/6], [5/6, 1]), \quad m^{1\textcircled{M}2}(\mathcal{I}'_2 \boxtimes \mathcal{I}''_2) = 0.2.$$

La représentation sur le simplexe des éléments focaux pour le bac, la nouvelle pièce et leur mélange est donnée dans la Fig. 5.5.



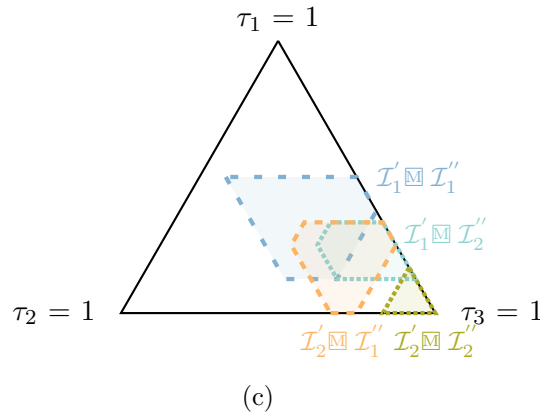


FIG. 5.5 Représentation des éléments focaux décrivant, le bac 5.5a, la nouvelle pièce 5.5b et leur mélange 5.5c dans le simplexe unitaire.

Avec la somme que nous venons de définir, nous sommes capables de suivre l'évolution de notre connaissance imparfaite sur les proportions d'un mélange itérativement. En effet, à chaque fois que l'on rajoute une pièce, nous mettons à jour notre information sur le mélange constituant le bac en sommant le vecteur associé à la nouvelle pièce avec le vecteur associé au mélange antérieur.

5.1.3 Inférence sur le respect des contraintes

Intéressons-nous à présent à des événements spécifiques exprimant le respect de contraintes sur les proportions. Un type particulier d'événements (et son complément) qui nous intéresse est :

"La proportion des composants d'un sous-ensemble A est en dessous d'un seuil α ",

où $A \subseteq \Theta$ est un sous-ensemble de types de matériaux spécifiques. On peut penser par exemple à des contraintes normatives sur la concentration de polluants dans le sol ou à une limitation d'impuretés dans un mélange pétrolier. Dans le cas du tri de plastiques, les compositions des bacs de tri sont soumises à des contraintes visant à limiter les impuretés. Ce type d'événements, que nous désignerons par A_α est un sous-ensemble de \mathbb{U} tel que :

$$A_\alpha = \{(\tau_1, \dots, \tau_n) \in \mathbb{U} : \text{et } \sum_{\theta_i \in A} \tau_i \leq \alpha\}. \quad (32)$$

Dans le cas déterministe : il suffit de vérifier l'appartenance du vecteur de proportions à A_α pour respecter la contrainte.

Dans le cas probabiliste : il est question d'inférer la probabilité de l'évènement A_α à partir de la distribution de probabilité p représentant le mélange. Dans ce cas :

$$\mathbb{P}(A_\alpha) = \sum_{\tau \in A_\alpha} p(\tau)$$

il est possible de se prononcer sur le respect de la contrainte par exemple en calculant l'espérance $\mathbb{E}[\mathbb{1}_{A_\alpha}] = \mathbb{P}(A_\alpha)$.

Dans le cas évidentiel : considérons un corps d'évidence (m, \mathbb{F}) défini sur $[0, 1]^n$. L'inférence dans ce cas consiste à calculer les plausibilités/croyances des événements du type A_α . Pour calculer $Bel(A_\alpha)$ et $Pl(A_\alpha)$, nous devons vérifier si les éléments focaux sont inclus, se chevauchent

ou sont disjoints de A_α . Pour un élément focal donné $\mathcal{I} \in \mathbb{F}$ tel que $\mathcal{I} = ([\ell_1, u_1], \dots, [\ell_n, u_n])$, cela peut être vérifié comme suit (De Campos et al., 1994) :

- Nous aurons $\mathcal{I} \subseteq A_\alpha$ si les sommes de toutes les proportions possibles dans \mathcal{I} des composants dans A sont inférieures au seuil α . Cela peut se vérifier par :

$$\mathcal{I} \subseteq A_\alpha \Leftrightarrow \min\left(\sum_{\theta_i \in A} u_i, 1 - \sum_{\theta_i \notin A} \ell_i\right) \leq \alpha. \quad (33)$$

Le premier terme de l'inégalité à droite de cette équivalence étant la proportion maximale de composants de A dans le mélange étant données nos connaissances imprécises sur les proportions.

- Nous aurons $\mathcal{I} \cap A_\alpha \neq \emptyset$ s'il existe au moins un vecteur de proportions pour lequel la proportion de matériaux affectée à A est inférieure à α . Cela peut être vérifié par :

$$\mathcal{I} \cap A_\alpha \neq \emptyset \Leftrightarrow \max\left(\sum_{\theta_i \in A} \ell_i, 1 - \sum_{\theta_i \notin A} u_i\right) \leq \alpha. \quad (34)$$

Le premier terme de l'inégalité à droite de cette équivalence étant la proportion minimale de composants de A dans le mélange compte tenu de nos connaissances imprécises sur les proportions. Si cette proportion est inférieure à α , cela prouve qu'au moins un vecteur se trouve dans $\mathcal{I} \cap A_\alpha$, et si elle est supérieure, aucun vecteur ne se trouve dans A_α , et donc $\mathcal{I} \cap A_\alpha = \emptyset$.

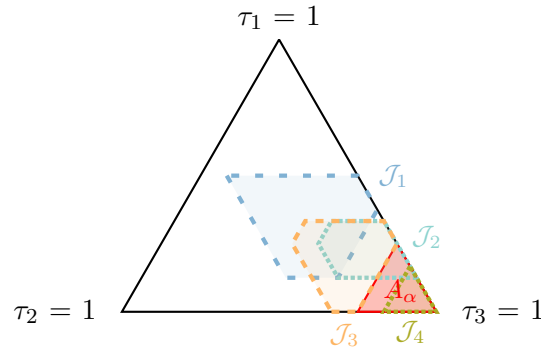
Notons que dans la pratique, de tels événements peuvent facilement être combinés. Soit A et B deux sous-ensembles de Θ . Supposons par exemple que les éléments de A ne devraient pas dépasser une proportion α et que les éléments de B , ne devraient pas dépasser une proportion β , alors vérifier si $\mathcal{I} \subseteq A_\alpha \cap B_\beta$ et $\mathcal{I} \cap (A_\alpha \cap B_\beta) \neq \emptyset$ revient à effectuer les vérifications (33) et (34) pour A_α et B_β séparément. Nous sommes intéressés par leur intersection (ce ne serait pas vrai pour leur union). Notons également que vérifier si les éléments d'un sous-ensemble E dépassent bien un certain seuil $\lambda \in [0, 1]$ revient simplement à vérifier que le vecteur de proportions résultant appartient à $(E^c)_{1-\lambda}$ (où E^c désigne le complémentaire de E dans Θ). Cette remarque s'étend bien sûr à un nombre illimité d'événements de ce type. En particulier, on pourrait vérifier si la proportion d'un matériau particulier nuisible ou polluant n'est pas trop élevée (ou si la quantité d'un élément intéressant ou nécessaire est suffisamment élevée).

Pour calculer $[Bel(A_\alpha), Pl(A_\alpha)]$, il nous suffit d'effectuer les $|\mathbb{F}|$ vérifications décrites ci-dessus, ou même moins si \mathbb{F} a une structure spécifique (par exemple, l'imbrication des éléments focaux). Nous pourrions étendre l'étude pour calculer le nombre de calculs en cas, par exemple, d'éléments focaux emboîtés, etc..

Exemple 8. Revenons à l'exemple 7 et considérons le résultat du mélange de la nouvelle pièce dans le bac. Considérons $\Theta = \{\theta_1, \theta_2, \theta_3\}$, $A = \{\theta_1, \theta_2\}$ et $\alpha = 0.25$ (la proportion de θ_1 et θ_2 ne devrait pas dépasser 0.25, ou de manière équivalente la proportion de θ_3 devrait être supérieure à 0.75), la TABLE 5.2 donne les résultats des différentes vérifications décrites ci-dessus concernant A_α et la Fig. 5.6 montre la représentation des vecteurs de proportions imprécises dans le simplexe unitaire.

TABLE 5.2 – Calcul de $Bel(A_\alpha)$ et $Pl(A_\alpha)$

		Croyance Bel	Plausibilité Pl
\mathcal{J}	$m^1 \textcircled{\mathbb{M}}^2(\mathcal{J})$	$\mathcal{J} \subseteq A_\alpha$	$\mathcal{J} \cap A_\alpha \neq \emptyset$
$\mathcal{J}_1 = \mathcal{I}'_1 \boxtimes \mathcal{I}''_1$	0.3	non	non
$\mathcal{J}_2 = \mathcal{I}'_1 \boxtimes \mathcal{I}''_2$	0.2	non	oui
$\mathcal{J}_3 = \mathcal{I}'_2 \boxtimes \mathcal{I}''_1$	0.3	non	oui
$\mathcal{J}_4 = \mathcal{I}'_2 \boxtimes \mathcal{I}''_2$	0.2	oui	oui
		$Bel(A_\alpha) = 0.2$	$Pl(A_\alpha) = 0.7$

FIG. 5.6 Représentation des éléments focaux du mélange et de l'ensemble A_α dans le simplexe.

Remarque 2. Remarquons qu'avoir un niveau minimal de croyance avec une contrainte sur les proportions vérifiées sur plusieurs pièces, ne garantit pas que nous aurions le même niveau minimal de croyance à propos du mélange de ces pièces. Reprenons l'exemple 7, où nous mélangeons le contenu d'un bac à une nouvelle pièce. La croyance que la proportion des composants θ_1 et θ_2 est inférieure à 0.25, $Bel(A_\alpha)$, avec $A = \{\theta_1, \theta_2\}$, $\alpha = 0.25$, pour ce bac est 0.5 et cette croyance est 0.4 pour la nouvelle pièce. En revanche la croyance $Bel(A_\alpha)$ pour le mélange du bac et de la nouvelle pièce est bien moins élevée (0.2).

Dans la section suivante, nous appliquons les outils proposés pour manipuler des vecteurs de proportions dans le problème du tri de plastiques.

5.2 Stratégies de tri

Dans cette sous-section, nous proposons des stratégies de tri de fragments de plastiques basées sur les outils d'inférences introduits précédemment. Les *pièces* sont désormais des fragments de plastiques à trier. Comme exposé dans le premier chapitre de cette thèse, les conditions industrielles du tri des plastiques impliquent la présence d'imprécision et d'incertitude. De plus pour des raisons physico-chimiques liées à leur non-miscibilité, les plastiques doivent être triés par types de matériaux avant d'être recyclés. Étant donné qu'une petite quantité de plastiques mal triés peut entraîner une forte diminution de la résistance aux chocs (Signoret et al., 2019c) et de la valeur monétaire, les impuretés doivent être limitées. Ainsi, les experts définissent des seuils de tolérance sur les proportions d'impuretés.

En pratique, des fragments de plastiques, arrivent en continu sur un convoyeur avant d'être enregistrés par une caméra infrarouge. Les informations acquises pour chaque fragment sont ensuite utilisées pour décider du tri des fragments, c'est-à-dire dans quel bac les fragments doivent être envoyés. Deux étapes sont considérées dans le tri d'un fragment. La première étape consiste à appliquer un classifieur intégré au dispositif de tri donnant une information sur le type de plastique du fragment. A partir de cette information, la deuxième étape consiste à effectuer la décision du tri en ligne du fragment en respectant des contraintes sur les compositions des bacs. Dans cette section, nous considérons que les résultats de la première étape sont disponibles et utilisés comme entrées pour la deuxième étape.

Notons que nous faisons l'hypothèse d'une alimentation continue de fragments, autrement dit le tri s'arrête quand un bac est plein (et non pas quand il n'y a plus rien à trier). Afin d'optimiser le rendement du tri il est donc nécessaire de maximiser la vitesse de remplissage des bacs avec la plus grande valeur marchande. Dans cet objectif, le décideur n'aura pas la même attitude face au risque de se tromper en fonction de l'avancement du remplissage du bac et de la confiance en la qualité du mélange réalisé dans le bacs.

Exemple 9. *Pour illustrer la différence entre un problème de classification et notre problème de tri, considérons que nous sommes sur la fin du tri, et que nous sommes en train de déterminer l'intérêt d'ajouter un fragment dans un bac plutôt que dans un autre. Nous aurions des décisions différentes pour les deux situations suivantes :*

- *Dans la première situation, nous sommes très confiants sur le remplissage courant et nous estimons, avec les informations disponibles, qu'ajouter le fragment ne perturbera pas la qualité du bac, même s'il s'avérait que le fragment était une impureté. Donc nous ajouterions le fragment pour remplir le bac au plus vite.*
- *Dans la seconde situation, supposons que nous sommes bien moins confiants quant à la qualité du tri. Nous ne prendrions pas le risque de polluer davantage le bac, selon les informations disponibles, en y ajoutant le fragment.*

Dans un problème de classification, l'état de remplissage ne serait pas pris en compte et nous aurions une seule décision pour les deux situations.

A des fins d'évaluation nous simplifions le problème de tri en considérant seulement deux bacs et des éléments de résolution du problème de tri généralisé au cas de plusieurs bacs sont donnés dans la section 5.4. Nous considérons des conditions de tri impliquant un bac (que nous appellerons *bac positif*) dédié aux plastiques respectant certaines contraintes et un bac pour le *rejet* qui n'est soumis à aucune contrainte sur les proportions de sa composition. Le dispositif de tri considéré offre donc deux options de tri pour un fragment fr :

1. le souffler dans le bac positif, i.e., *tri positif*,
2. ou le rejeter, i.e., *tri négatif*.

Pour faciliter notre exposé, nous parlons de 'fragment positif' pour désigner un fragment envoyé dans le bac positif. Nous considérons que les proportions de matériaux de types $A \subseteq \Theta$ ne doivent pas dépasser le seuil α dans le bac positif. En d'autres termes, le bac est dédié à A^c et les éléments de A sont des impuretés. Nous utilisons l'ensemble A_α , défini dans l'équation (32), pour exprimer cette contrainte.

Précisons que nous faisons l'hypothèse que les fragments sont monomatières, i.e., avec des *proportions binaires*. L'information dont nous disposons pour un fragment fr est fournie par un classifieur évidentiel sous la forme d'une fonction de masse sur Θ . Une telle fonction de masse peut se traduire en une fonction de masse sur $[0, 1]^n$ comme illustré dans l'exemple 10.

Exemple 10. Par exemple si un classifieur évidentiel retourne la fonction de masse suivante concernant fr :

$$m^\Theta(\{\theta_1\}|fr) = 0.4, \quad m^\Theta(\{\theta_1, \theta_3\}|fr) = 0.6.$$

Alors, prenant en compte la Remarque 1 (suite) pour convertir une information imprécise, i.e., un sous-ensemble de Θ , en un vecteur de proportions imprécises, nous obtenons :

$$m^{\{0,1\}^n}(\{\{1\}, \{0\}, \{0\}\}|fr) = 0.4, \quad m^{\{0,1\}^n}(\{\{0, 1\}, \{0\}, \{0, 1\}\}|fr) = 0.6.$$

Ensuite, toujours d'après la Remarque 1 (suite), nous pouvons considérer la fermeture convexe de manière équivalente dans la suite des inférences :

$$m^{[0,1]^n}(\{[1, 1], [0, 0], [0, 0]\}|fr) = 0.4, \quad m^{[0,1]^n}(\{[0, 1], [0, 0], [0, 1]\}|fr) = 0.6.$$

Dans la suite de la section, nous montrons comment employer les outils proposés dans ce chapitre pour suivre au cours du tri, la croyance que les contraintes sur les impuretés dans le bac positif soient respectées. Ensuite, nous présentons trois stratégies de tri basées sur ces outils. Enfin nous donnons des précisions sur la complexité impliquée par ces stratégies.

5.2.1 Mise à jour de la connaissance concernant la composition du bac

Afin de trier successivement les fragments transportés par le convoyeur, nous considérerons les informations suivantes. D'abord concernant le fragment fr en lui même, nous notons :

- son poids W^{fr} ,
- sa fonction de masse $m(\cdot|fr)$.

Concernant le bac positif, nous avons les informations suivantes avant de trier le $k + 1$ -ième fragment passant sur le convoyeur :

- la capacité totale du bac W^{bac} , c'est-à-dire son poids lorsqu'il est rempli de fragments,
- le poids actuel W_k^{bac} des fragments qu'il contient,
- la fonction de masse m_k décrivant notre connaissance sur les proportions actuelles de matériaux qu'il contient.

Si fr est ajouté au bac, le poids actuel W_k^{bac} et la fonction de masse m_k sont mis à jour à l'aide de l'opération de somme (31) :

$$(m_{k+1}, W_{k+1}^{bac}) = (m_k \circledast m(\cdot|fr), W_k^{bac} + W^{fr})$$

A partir de m_{k+1} , nous pouvons calculer $Bel_{k+1}(A_\alpha)$, i.e., le degré de croyance que le nouveau mélange, le bac positif et fr , ne contient pas une proportion des composants éléments de A supérieure à α . Dans la suite nous notons $final$, l'indice k tel que $W_k^{bac} = W^{bac}$. Nous notons donc m_{final} la fonction de masse décrivant les proportions dans le bac positif à l'étape finale du tri.

5.2.2 Inférence sur le respect des contraintes dans le bac positif

La stratégie de tri est cruciale car elle cherche à optimiser le *taux de tris positifs final*, c'est-à-dire le poids total des matériaux triés positivement, i.e., soufflés dans le bac dédié aux plastiques respectant les contraintes, par rapport au poids total de tous les fragments traités. Plus ce taux est élevé, plus le bac est rempli rapidement et la matière inutilement rejetée est minimisée. Cependant, afin de respecter les contraintes sur les impuretés, le décideur a également besoin de limiter l'ajout de fragments susceptibles d'être des impuretés dans le bac positif.

Jusqu'à présent, nous sommes en mesure de suivre l'évolution de nos connaissances sur la composition du bac à chaque étape k du tri, et de faire des inférences sur la satisfaction des contraintes sur les impuretés. Nous pourrions maintenant baser la décision du tri sur l'inférence de la croyance de l'événement A_α . Par exemple, une procédure de décision pour le tri de fr pourrait s'appuyer sur la potentielle composition du bac si nous y ajoutons fr . Ainsi, le dispositif de tri ne soufflerait le fragment dans le bac positif que si :

$$Bel_k(\otimes)Bel(A_\alpha|fr) \geq \eta, \quad (35)$$

η étant un seuil de confiance appartenant à $[0, 1]$ qui devrait être suffisamment grand pour limiter le risque d'enfreindre les contraintes associées au bac positif. Considérer cette décision pour chaque fragment successif garantira qu'à tout moment du remplissage du bac positif, le degré de croyance que les contraintes d'impureté soient respectées sera supérieure à η . Il s'agit d'une contrainte très forte lorsque $Bel_k(A_\alpha)$ est proche de η , seulement les fragments avec une très forte croyance de ne pas être une impureté pourraient maintenir $Bel_{k+1}(A_\alpha)$ au dessus de η .

Une procédure plus flexible pourrait être envisagée dans le cas où les contraintes seraient exigées seulement à la fin du remplissage du bac positif, c'est-à-dire lorsque la quantité de matière dans le bac atteint le poids total W^{bac} . L'objectif serait alors de vérifier :

$$Bel_{final}(A_\alpha) \geq \eta. \quad (36)$$

Dans ce cas, nous pourrions autoriser $Bel_k(A_\alpha) < \eta$ au cours du tri. L'idée derrière cette procédure flexible est de faire des hypothèses sur les futurs fragments à venir afin de mieux évaluer les avantages de l'ajout du fragment courant dans le bac positif. Aussi $Bel_{final}(A_\alpha)$ devra alors être estimée à tout moment.

Nous pensons qu'une tolérance raisonnable sur la qualité des fragments est préférable au début du tri, alors que la fin du tri doit être plus exigeante car seule la tolérance sur le contenu final compte. Cela nous amène à être plutôt optimistes quant aux futurs fragments positifs, en d'autres termes, nous pensons que les futurs fragments positifs seront des fragments pour lesquels notre confiance qu'ils ne soient pas des impuretés est élevée, et donc qu'ils pourraient compenser les possibles erreurs du début du remplissage. Une hypothèse simple est de considérer qu'à court terme, les futurs fragments positifs seraient de la même nature que le fragment actuel, i.e, nous piochons dans le même gisement, et qu'à long terme les derniers fragments seront de «bons» fragments, i.e., des fragments pour lesquels nous disposons de preuves qu'ils sont composés de matériaux dans A^c .

Notons $W_{reste} = W^{bac} - W_k^{bac}$ la capacité résiduelle du bac. Nous considérerons les hypothèses suivantes pour inférer la croyance de l'événement A_α à toute étape k :

- D'abord, la première moitié des futurs fragments positifs correspond à un premier fragment fictif $fr^{partie1}$ de poids νW_{reste} , $\nu > 0$ ayant une fonction de masse $m^{partie1}$ identique à la fonction de masse du fragment à venir, c'est-à-dire $m^{partie1} = m(\cdot|fr)$.

- Ensuite pour la seconde partie du bac à remplir, nous considérons un deuxième fragment fictif $f^{partie2}$ de poids $(1 - \nu) W_{reste}$. Traduisant notre optimisme sur cette deuxième partie, nous avons choisi de représenter la croyance à propos de $f^{partie2}$ par une fonction de masse qui attribue une grande masse pour l'événement " $f^{partie2}$ est de type A^c " (le bac étant dédié aux plastiques de A^c). Pour obtenir une telle fonction de masse, nous pouvons considérer l'élément focal $\mathcal{I}^{partie2} = (\mathcal{I}_1^{partie2}, \dots, \mathcal{I}_n^{partie2})$, où chaque $\mathcal{I}_i^{partie2}$, $i \in \{1, \dots, n\}$ est tel que :

$$\mathcal{I}_i^{partie2} = \begin{cases} 0 & \text{si } \theta_i \notin A^c, \\ 1 & \text{si } A^c = \{\theta_i\}, \\ [0, 1] & \text{si } \theta_i \in A^c \text{ et } |A^c| > 1. \end{cases}$$

Nous définissons ensuite la fonction de masse sur la proportion de $f^{partie2}$ par :

$$\begin{aligned} m^{partie2}(\mathcal{I}^{partie2}) &= \gamma, \\ m^{partie2}(\mathbb{U}) &= 1 - \gamma, \end{aligned}$$

$\gamma \in [0, 1]$, choisi en fonction de notre degré d'optimisme.

Fig. 5.7 représente un récapitulatif de toutes les informations et hypothèses utilisées pour estimer les croyances concernant la composition finale du bac.

$m^{partie2} (1 - \nu) W_{reste}$	}	Futur à long terme
$m^{partie1} \quad \nu W_{reste}$		
$m(\cdot fr)$	}	Présent, fragment actuel
m_k		
W^{fr}	}	Passé
W_k^{bac}		

FIG. 5.7 Estimation de la composition finale du bac d'après notre hypothèse simple

En considérant ces fragments fictifs, nous permettons une certaine prise de risque au début du tri. Notons que plus γ sera élevé, c'est à dire, plus nous espérons que les futurs fragments pourront compenser les décisions précédentes, plus la prise de risques sera accentuée. Cette propension au risque s'atténuera au cours du remplissage. La Fig. 5.8, montre comment est estimée la composition finale au début du tri (k petit) et sur la fin (k grand). En effet, au début W_{reste} est élevé donc notre hypothèse optimiste a un rôle important dans l'évaluation de (36). Cependant, au fur et à mesure de l'avancement du tri, le volume disponible pour les futurs fragments diminuera, de même que leur influence dans l'estimation de la croyance concernant le respect des contraintes de la composition finale. En contrepartie, l'état actuel de la composition du bac aura une influence croissante dans l'évaluation de (36). Par conséquent, le tri sera plus exigeant sur la fin.



FIG. 5.8 Estimation de la composition finale du bac d'après notre hypothèse simple, au début du remplissage (k petit), 5.8a et sur la fin du remplissage (k grand), 5.8b avec $\nu = 0.05$.

Dans la prochaine section, nous comparerons les stratégies suivantes pour trier un fragment fr :

- Procédure naïve : la décision n'est basée que sur les informations concernant le fragment actuel. Si $Bel(A_\alpha|fr) \geq \eta$ est vérifié alors le fragment fr est trié positivement.
- Procédure prudente : la décision est basée sur les informations concernant le fragment actuel $m(\cdot|fr)$ et la composition actuelle du bac, c'est-à-dire m_{k-1} . Si (35) est vérifiée, alors le fragment fr est trié positivement.
- Procédure flexible : comme décrit précédemment, cette procédure est basée sur une estimation des croyances concernant la composition finale du bac. Finalement, la procédure correspond au tri positif de fr uniquement si notre estimation de m_{final} par $m_k \otimes m(\cdot|fr) \otimes m^{partie1} \otimes m^{part2}$ implique que (36) est vérifiée.

5.2.3 Complexité

Dans cette sous-section, la complexité des procédures consistant à trier l fragments est analysée. Nous notons \mathbb{F}^k l'ensemble des éléments focaux de la masse associée au k -ième fragment.

5.2.3.1 Procédure naïve

Dans le cas de la procédure naïve, l'ensemble A_α est comparé à $|F^k|$ vecteurs de proportions imprécises pour chaque étape k afin de déterminer $Bel(A_\alpha|fr)$. Ainsi, la complexité est linéaire dans ce cas, c'est-à-dire que la complexité pour cette procédure est $\mathcal{O}(l)$.

5.2.3.2 Procédure prudente ou flexible

Concernant les approches qui prennent en compte des mises à jour des croyances sur la composition du bac, comme les procédures prudente et flexible, le nombre d'éléments focaux associés à m_l , i.e. la fonction de masse décrivant les proportions du bac à la fin du tri, croît de façon exponentielle au nombre de fragments dans le dépôt. En effet, nous avons au plus

$$|\mathbb{F}^1| \times |\mathbb{F}^2| \times \dots \times |\mathbb{F}^l|$$

éléments focaux pour m_l . Considérons le cas où $|\mathbb{F}^k| = 2$ pour tous les fragments f_k . Ainsi, m_l aurait au plus 2^l éléments focaux. Avec la procédure prudente, il est nécessaire de déterminer $Bel_k(A_\alpha)$ pour tous les l fragments, y compris ceux rejetés. Le coût de ces opérations est inférieur à $\sum_{j=1}^l 2^j = 2^{l+1} - 2$. Ainsi, la procédure prudente aurait une complexité exponentielle $\mathcal{O}(2^l)$. La

procédure flexible est plus coûteuse que la procédure prudente et sa complexité est également exponentielle. En effet, deux éléments focaux pour chaque masse $m^{partie1}$ et $m^{partie2}$ doivent être pris en compte à chaque étape. Ainsi, 2^{k+2} éléments focaux totaux sont considérés et la complexité est $\mathcal{O}(2^l)$.

Afin de réduire le nombre d'éléments focaux, nous pouvons effectuer des approximations, en considérant par exemple que les éléments focaux "très proches" sont identiques. Par exemple, les intervalles $\mathcal{I}^1 = [l^1, u^1]$ et $\mathcal{I}^2 = [l^2, u^2]$ pourraient être considérés comme "très proches" si la distance euclidienne entre toutes les bornes des intervalles de \mathcal{I}^1 et \mathcal{I}^2 est inférieure à une petite valeur réelle positive ϵ :

$$\sqrt{\sum_{i=1}^n [(l_i^1 - l_i^2)^2 + (u_i^1 - u_i^2)^2]} < \epsilon.$$

D'autre part, nous pourrions envisager que les éléments focaux auxquels une très petite masse est affectée seraient supprimés et leurs masses transférées à l'ensemble total (ignorance).

5.3 Illustration

Nous illustrerons dans cette section, les procédures de tri naïve, prudente et flexible, via des simulations de tri. Dans un premier temps, nous présenterons les données utilisées pour nos expériences, puis présenterons nos résultats de simulations de tri avec les trois procédures.

5.3.1 Présentation des données disponibles pour le problème de tri

Les données utilisées pour les expériences de ce chapitre sont les images hyperp spectrales de centaines de fragments de plastiques fournies par l'entreprise Pellenc st. La répartition des fragments par type de plastique dans $\Theta = \{ABS, HIPS, PE, PP\}$ et par appartenance aux bases d'apprentissage et de test, est donnée dans la TABLE 5.3. En moyenne, le nombre de pixels par fragment est environ 83.5.

	Apprentissage				Test			
	<i>Styréniques</i>		<i>Polyoléfines</i>		<i>Styréniques</i>		<i>Polyoléfines</i>	
	<i>ABS</i>	<i>HIPS</i>	<i>PE</i>	<i>PP</i>	<i>ABS</i>	<i>HIPS</i>	<i>PE</i>	<i>PP</i>
Nombre de pixels	2958	2946	2984	2910	2877	2913	2878	2923
Nombre de fragments	35	35	35	35	35	35	35	35

TABLE 5.3 – Base de données

Afin d'obtenir des fonctions de masse sur les 140 fragments de test, un classifieur évidentiel a été entraîné sur les 11798 pixels de la base d'apprentissage. Le classifieur utilisé est le classifieur évidentiel ECLAIR proposé dans cette thèse (voir chapitre 4), effectué à partir du classifieur probabiliste par régression logistique (chapitre 2, sous-section 2.2.3)². Une fois entraîné, le classifieur évidentiel a été appliqué aux 11591 pixels de la base de test. Finalement, pour chaque fragment de test, nous avons combiné les masses provenant de chaque pixel par l'opérateur moyenne arithmétique \oplus (chapitre 2, sous-section 2.1.3) car cette combinaison a une faible complexité, est adaptée au conflit et n'implique pas de nouveaux éléments focaux. La Fig. 5.9

2. Dans la procédure ECLAIR utilisée ici, le ré-étiquetage imprécis des exemples d'apprentissage a été effectué sur des validations croisées 10-blocs par la méthode utilisant l'entropie avec $\rho = 0.5$.

résume le procédé qui nous a permis d’obtenir une fonction de masse pour un exemple fictif de fragment de test fr .

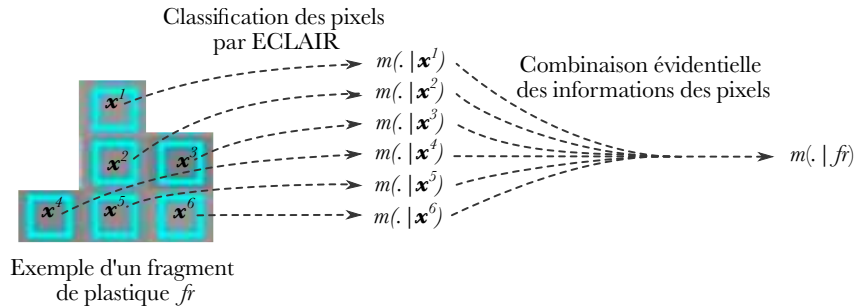


FIG. 5.9 Procédure utilisée pour obtenir une fonction de masse concernant un fragment fr à partir de ses 6 pixels x^1, \dots, x^6 .

Par ailleurs, le poids W^{fr} de fr est approché par le nombre de pixels dans son image hyperspectrale. Dans l’exemple de la Fig. 5.9, nous avons $W^{fr} = 6$.

Finalement, nous disposons pour nos expériences de 140 couples $(m(.|fr), W^{fr})$ concernant 140 fragments de test.

5.3.2 Simulations

Afin d’évaluer les procédures de tri, nous avons effectué 30 simulations de flux de fragments, i.e., suite aléatoire de fragments à trier comme s’ils arrivaient sur le convoyeur. La simulation d’un flux a été effectuée en sélectionnant des ordres d’indexation aléatoires dans $\{1, \dots, 140\}$ correspondant aux 140 fragments de test. Pour des raisons de temps de calcul, nous avons fixé la capacité du bac positif à $W^{bac} = 800$. Pour chaque simulation, nous avons arrêté la procédure de tri lorsque la capacité W^{bac} était atteinte. Nous avons considéré que le bac positif était dédié aux plastiques styréniques (ensemble $\{ABS, HIPS\}$) et que la proportion de polyoléfines (ensemble $\{PE, PP\}$) ne devait pas dépasser 0.04 dans le bac. Nous avons donc considéré un bac positif exigeant que la proportion des composantes de A doit être inférieure à α , i.e, la proportion appartient à A_α , avec $A = \{PE, PP\}$ et $\alpha = 0.04$. Le bac positif est ainsi dédié aux plastiques de $A^c = \{ABS, HIPS\}$. De plus, nous avons fixé le seuil de confiance $\eta = 0.6$. Afin d’introduire les expériences menées, nous détaillons dans la Fig. 5.10, le suivi de $Bel_k(A_\alpha)$ et de W_k^{bac} pour une simulation de remplissage du bac avec la procédure naïve. Dans 5.10a et 5.10c, les fragments ayant été ajoutés au bac sont colorés en rose. Dans toutes les sous-figures, les pointillés représentent les étapes de la procédure où un fragment a été ajouté au bac. Dans 5.10b, et 5.10d la puce à l’extrémité de la courbe représente l’arrêt de la procédure (la capacité du bac 800 est atteinte). Avec cette procédure naïve, la décision est uniquement basée sur $Bel(A_\alpha|fr)$. Cela s’observe dans la sous-figure 5.10a, qui montre que les fragments envoyés sont systématiquement ceux tels que $Bel(A_\alpha|fr) > \eta$. Dans la sous-figure 5.10b, nous voyons que $Bel_k(A_\alpha)$ est mise à jour après chaque ajout d’un nouveau fragment.

Les simulations de remplissage des bacs avec les procédures prudente et flexible utilisent le même flux de fragments simulé (la même sélection aléatoire des ordres de passage des fragments). Rappelons que dans le cas de l’approche prudente, la décision est basée sur $Bel_k(\mathbb{M})Bel(A_\alpha|fr)$ et dans le cas de l’approche flexible, sur $Bel_{final}(A_\alpha)$. Pour l’approche flexible, l’estimation de l’état final du bac est calculée à chaque itération avec les paramètres $\gamma = 1$ et $\nu = 0.05$ (voir l’estimation de la composition finale résumée dans la Fig. 5.7). La valeur du paramètre γ traduit l’hypothèse la plus optimiste possible sur le fragment fictif $f^{partie2}$, i.e., la croyance que $f^{partie2}$

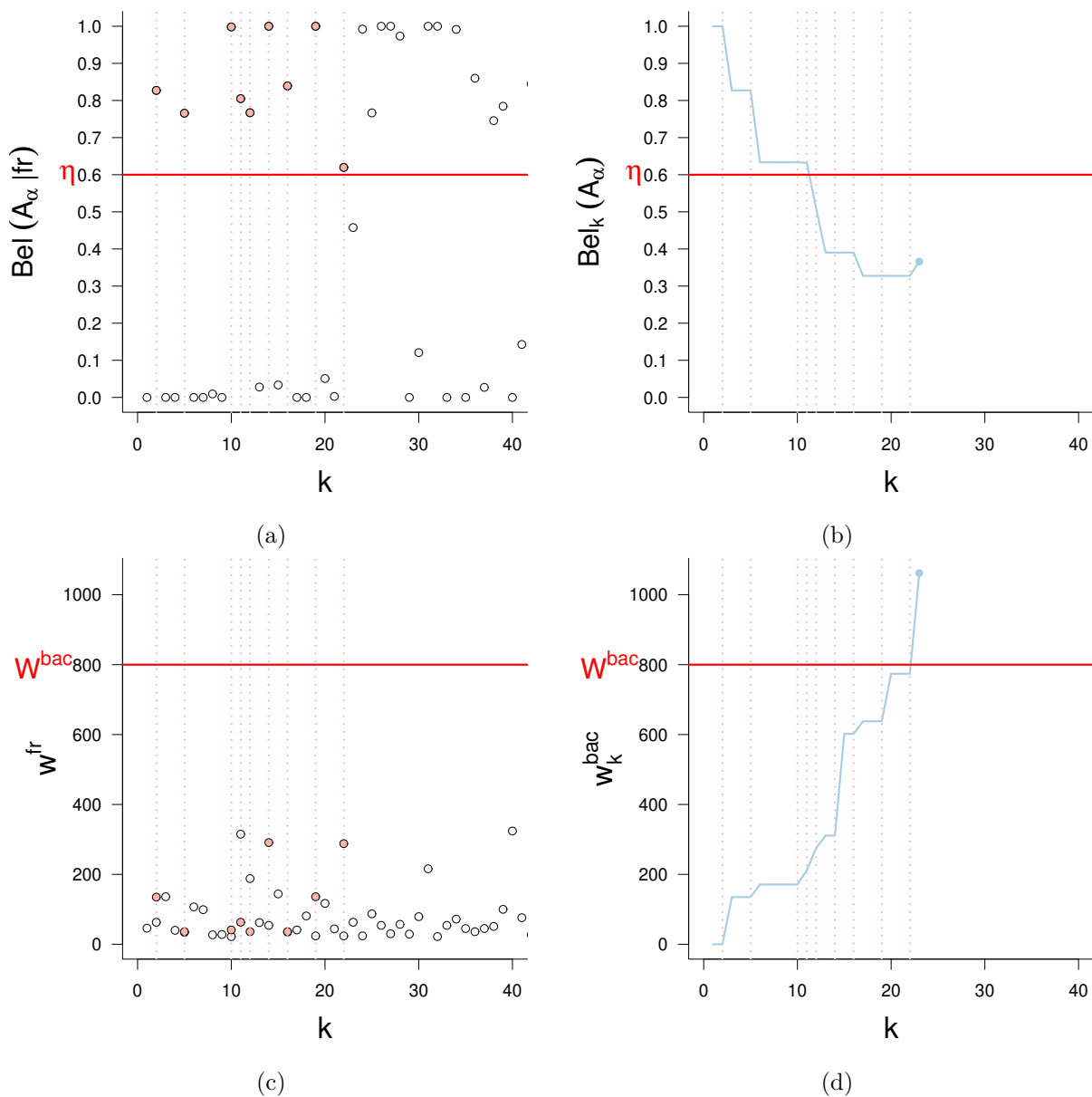


FIG. 5.10 *Détails des informations sur les fonctions de masse et les poids du flux de fragments simulé (5.10a, 5.10c) et du bac positif (5.10b, 5.10d) pour une simulation de remplissage avec la procédure naïve.*

	Naïf	Prudent	Flexible
$Bel_{final}(A_\alpha)$	0.3141	0.6285	0.6458
<i>taux de tris positifs final</i>	0.5263	0.2901	0.3259

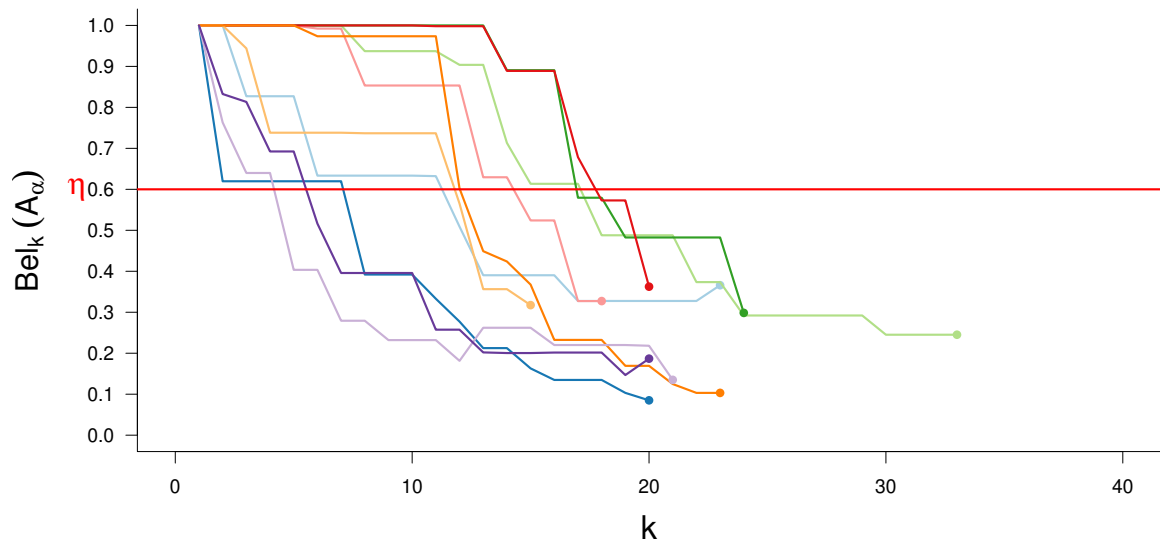
TABLE 5.4 – Valeurs moyennes de $Bel_{final}(A_\alpha)$ et *taux de tris positifs final* sur 30 simulations

est un styrénique vaut 1. D'autre part la faible valeur du paramètre ν renforce l'influence de $f^{partie2}$ dans l'estimation de l'état final de la composition du bac.

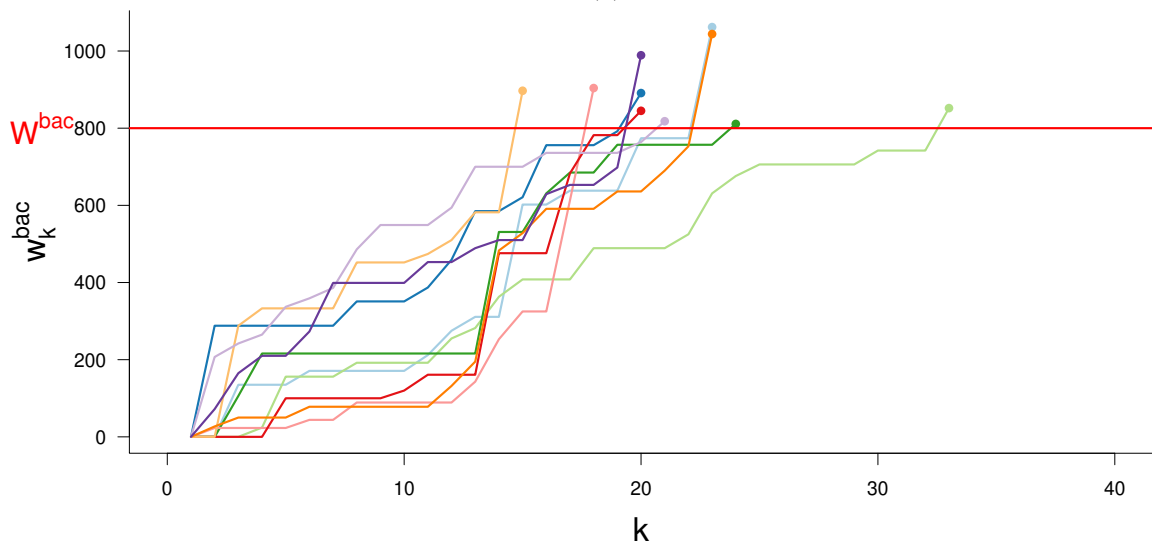
Dans la suite de la section, nous présentons les résultats des 10 premières simulations de remplissage par les trois approches. L'évolution de $Bel_k(A_\alpha)$ est tracée en fonction de la progression du tri dans les sous-figures 5.11a, 5.12a et 5.13a, respectivement pour les trois approches. De même, l'évolution de W_k^{bac} est tracée dans les sous-figures 5.11b, 5.12b et 5.13b. Chaque courbe représente une simulation et s'arrête avec une puce marquant la fin du remplissage du bac. Lorsque le remplissage requiert plus que 40 itérations, la fin du tri n'est pas affichée et une flèche l'indique à la place de la puce. Notons que chaque simulation a une durée différente en raison du taux de tris positifs, c'est-à-dire que plus le taux de fragments rejetés est élevé, plus la durée du tri est longue (pour atteindre W). Pour mieux comparer les résultats des approches, nous avons utilisé les mêmes simulations dans les trois procédures. La moyenne de toutes les simulations pour $Bel_{final}(A_\alpha)$ et pour le *taux de tris positifs final* sont présentés dans la TABLE 5.4. Notons que dans toutes les simulations, la contrainte sur la quantité d'impuretés est respectée, mais cela ne serait pas toujours le cas avec des expérimentations plus longues ou si nous avions plus de fragments difficiles à classer. Nous évaluons donc les procédures en fonction de la condition sur la composition du bac à la fin du tri (36).

La procédure naïve a l'avantage de remplir rapidement le bac (voir TABLE 5.4). Pourtant, comme nous pouvons l'observer sur la sous-figure 5.11a, cette procédure ne garantit pas que la contrainte (36) sur les impuretés du bac final sera respectée. En fait, nous pourrions vérifier (36), en étant beaucoup plus exigeants concernant chaque fragment ajouté dans le bac. Cependant, cela ne serait pas possible pour le tri industriel où la capacité des bacs W^{bac} est importante. Commentons les résultats obtenus avec l'approche prudente. Comme mentionné dans la sous-section 5.2.2, cette procédure impose que pour toutes les étapes k du remplissage du bac la contrainte $Bel_k(A_\alpha) \geq \eta$ soit vérifiée. L'inconvénient de cette procédure est qu'elle se concentre uniquement sur la composition courante du bac sans anticiper sur la fin du tri. Par conséquent, dans une situation favorable, c'est-à-dire lorsque $Bel_k(A_\alpha)$ est bien au-delà de η , certains mauvais fragments pourront être ajoutés au bac sans dommages. Cependant, cette action répétée trop souvent fait diminuer $Bel_k(A_\alpha)$. Lorsque $Bel_k(A_\alpha)$ se rapproche de η , et que le bac est proche d'être plein, il devient de plus en plus difficile d'accepter un nouveau fragment qui puisse maintenir $Bel_k(A_\alpha)$ supérieur à η . Finalement, le tri peut être ralenti à cause des premières décisions. Nous pouvons le voir sur la Fig. 5.12a, pour 4 simulations, $Bel_k(A_\alpha)$ descend très près de η et 40 itérations ne suffisent pas pour remplir le bac. Dans la procédure flexible, nous pouvons avoir $Bel_k(A_\alpha) < \eta$ pendant le tri, mais comme nous pouvons le voir dans la sous-figure 5.13a, la condition sur $Bel_k(A_\alpha)$ est toujours satisfaite à la fin du tri. De plus le taux de tris positifs final est meilleur que pour la procédure prudente (voir TABLE 5.4). Cela est dû à la sélection des fragments qui est plus appropriée avec cette procédure au double objectif, de respecter la contrainte de (36) et de minimiser la quantité de fragments rejetés.

Nous avons présenté les résultats de trois stratégies de tri sur des simulations de flux de plastiques à partir de données réelles. Malgré l'attrait que peut susciter l'approche naïve en terme de complexité, nous constatons que la procédure naïve n'est pas suffisante pour respecter la condition sur la croyance dans le respect des contraintes à la fin du remplissage. Les approches

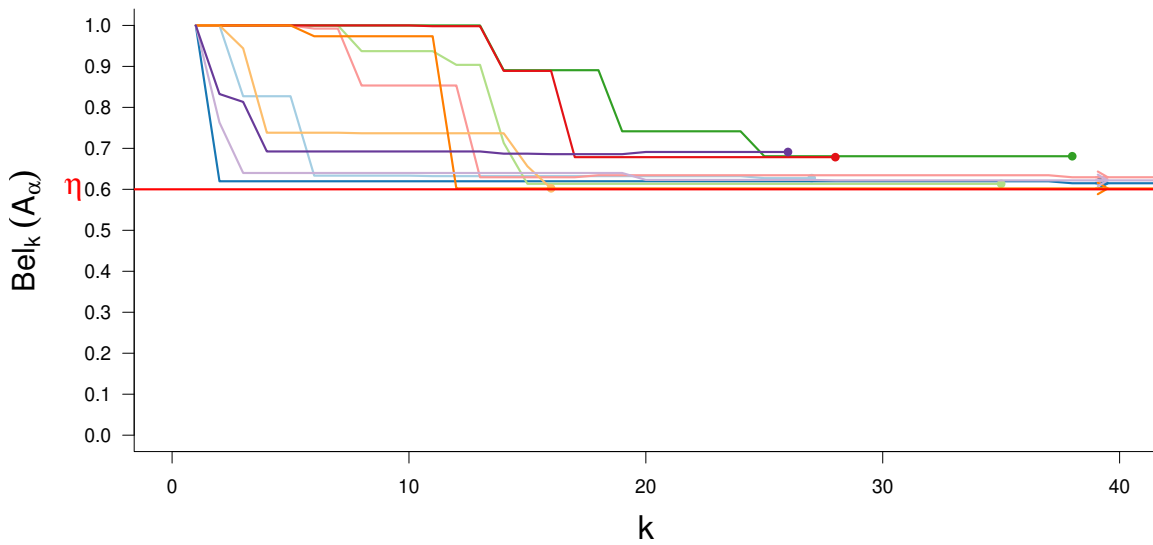


(a)

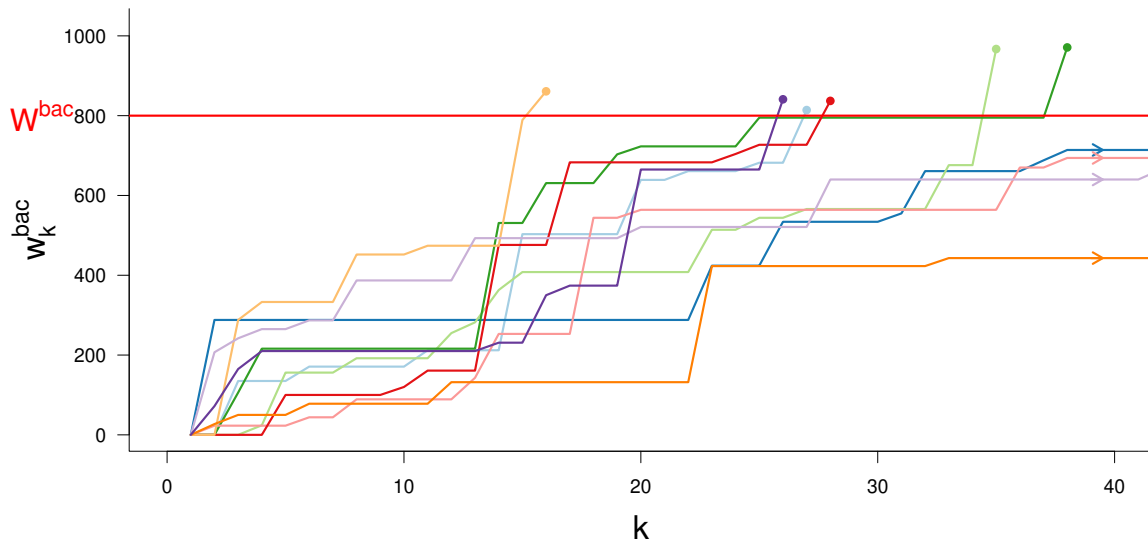


(b)

FIG. 5.11 Suivi au cours du remplissage du bac dédié aux plastiques styréniques avec la stratégie *naïve*, 5.11a de la croyance que la contrainte est respectée dans le bac, 5.11b du poids du contenu du bac W_k^{bac}

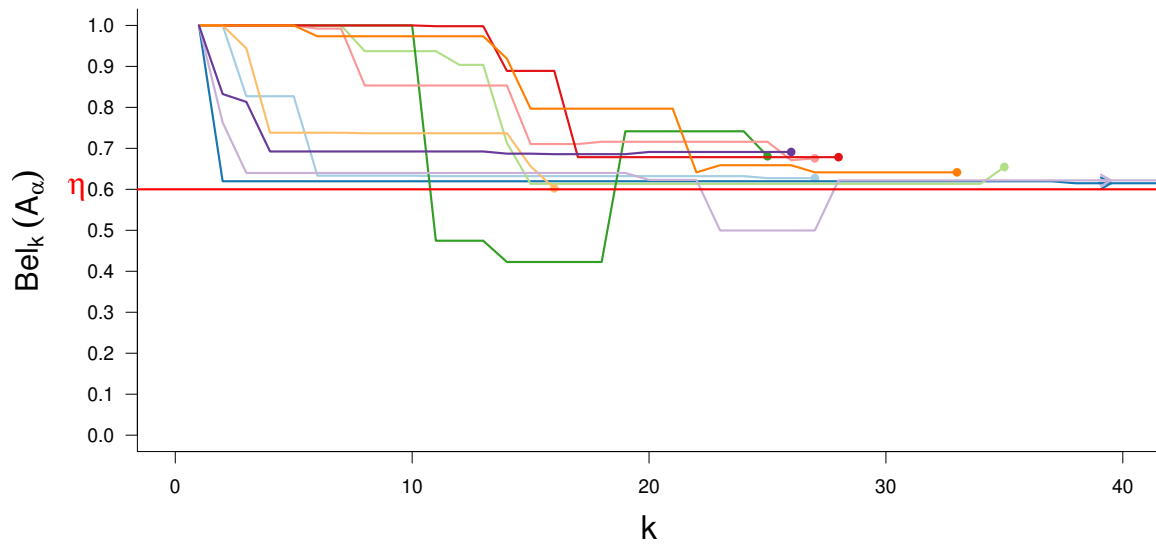


(a)

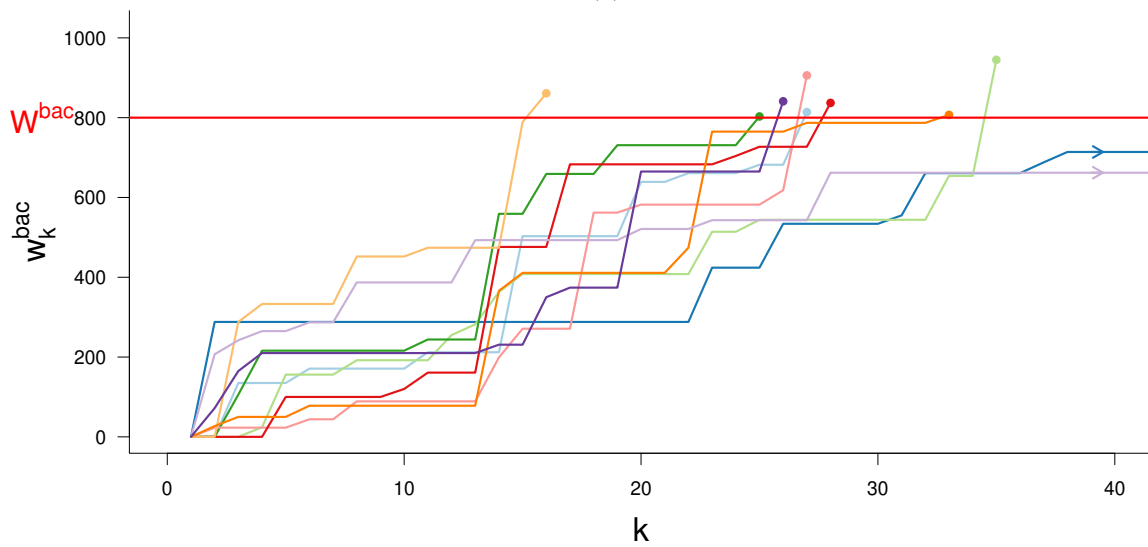


(b)

FIG. 5.12 Suivi au cours du remplissage du bac dédié aux plastiques styréniques avec la stratégie **prudente**, 5.12a de la croyance que la contrainte est respectée dans le bac, 5.12b du poids du contenu du bac W_k^{bac}



(a)



(b)

FIG. 5.13 Suivi au cours du remplissage du bac dédié aux plastiques styréniques avec la stratégie **flexible**, 5.13a de la croyance que la contrainte est respectée dans le bac, 5.13b du poids du contenu du bac W_k^{bac}

prudente et flexible qui mettent en oeuvre une décision plus sophistiquée, sont capables de maîtriser l'évolution de la croyance dans le respect des contraintes. L'approche flexible est préférable à l'approche prudente car elle permet un remplissage plus rapide du bac.

5.4 Discussion

Dans cette section, nous discutons les possibilités de généraliser l'approche proposée au cas de fragments qui ne sont pas monomatières et au cas de plusieurs bacs.

5.4.1 Généralisation au cas de fragments qui ne sont pas monomatières

Sans perte de généralité, les stratégies de tri présentées dans ce chapitre pourraient être utilisées pour traiter un problème de tri impliquant des fragments composées de plusieurs types de plastiques, i.e, fragments qui ne sont pas monomatière. Nous donnons ci-dessous quelques considérations concernant une possibilité pour obtenir des informations concernant des proportions imprécises dans ce cas. Nous pourrions faire appel à un classifieur multi-labels fournissant l'ensemble des types de matériaux présents dans le fragment fr . Supposons que ce classifieur multi-labels fournisse également une probabilité a posteriori $p(\cdot|fr)$ de la présence d'une combinaison d'éléments de Θ dans le fragment fr . Dans ce qui suit, nous adoptons la notation $\theta_i \wedge \theta_j$ pour exprimer que le fragment est composé des types de matériaux θ_i et θ_j . Pour simplifier, nous considérons uniquement les cas où le classifieur multi-labels hésite entre deux compositions.

Exemple 11. Soit $\theta = \{\theta_1, \theta_2, \theta_3\}$. Considérons que le classifieur multi-labels hésite entre deux compositions, θ_1 et $\theta_1 \wedge \theta_3$, nous avons $p(\theta_1|fr) + p(\theta_1 \wedge \theta_3|fr) = 1$. Dans ce cas, le classifieur hésite à considérer fr comme un matériau monomatière de type θ_1 ou bien comme une combinaison des deux types de matériaux θ_1 et θ_3 . Considérons qu'il fournisse les probabilités suivantes : $p(\theta_1|fr) = 0,4$ et $p(\theta_1 \wedge \theta_3|fr) = 0,6$. Notons que l'information " $p(\theta_1 \wedge \theta_3|fr) = 0,6$ " ne nous renseigne pas sur les proportions des matériaux θ_1 et θ_3 dans fr mais seulement sur leur éventuelle présence dans le fragment. En d'autres termes, nous savons seulement que les proportions θ_1 et θ_3 sont dans $]0, 1[$, que nous transformerons en l'intervalle fermé $[0, 1]$ sans perte d'informations pertinentes pour notre étude (voir Remarque 1 (suite)). Enfin, à partir de ces informations sur la présence de matériaux dans le fragment fr , nous pouvons exprimer une fonction de masse $m(\cdot|fr)$ sur les proportions de ces matériaux. Les éléments focaux correspondants \mathcal{I}^{θ_1} et $\mathcal{I}^{\theta_1 \wedge \theta_3}$ sont :

$$\begin{aligned}\mathcal{I}^{\theta_1} &= ([1, 1], [0, 0], [0, 0]), \\ \mathcal{I}^{\theta_1 \wedge \theta_3} &= ([0, 1], [0, 0], [0, 1]),\end{aligned}$$

où \mathcal{I}^{θ_1} est une information provenant de la sortie du classifieur : $p(\theta_1|x) = 0,4$ et $\mathcal{I}^{\theta_1 \wedge \theta_3}$ de $p(\theta_1 \wedge \theta_3|x) = 0,6$. Les masses de probabilité sont transférées à \mathcal{I}^{θ_1} et $\mathcal{I}^{\theta_1 \wedge \theta_3}$:

$$m(\mathcal{I}^{\theta_1}|fr) = 0,4, \quad m(\mathcal{I}^{\theta_1 \wedge \theta_3}|fr) = 0,6.$$

Plus généralement, considérons un fragment fr pour lequel le classifieur fournit une probabilité non nulle pour la composition D des matériaux. Le vecteur de proportions imprécises correspondant est alors $\mathcal{I}^D = ([l_1, u_1], \dots, [l_n, u_n])$ où pour $i \in \{1 \dots n\}$:

$$[l_i, u_i] = \begin{cases} [0, 0] & \text{si } \theta_i \notin D, \\ [1, 1] & \text{si } D = \theta_i, \\ [0, 1] & \text{si } \theta_i \in D \text{ et } |D| > 1. \end{cases}$$

En assignant des probabilités à ces vecteurs de proportions imprécises, nous obtenons des vecteurs de proportions imprécises et incertaines tels que considérés dans la définition 2. Ainsi, nous pouvons appliquer la somme traduisant le mélange de plusieurs fragments ainsi que les outils d'inférences présentés dans la section 5.1.

5.4.2 Généralisation au cas de plusieurs bacs

Nous donnons à présent des éléments de résolution du problème de tri généralisé au cas de plusieurs bacs soumis à des contraintes sur leurs proportions (Jacquin et al., 2020a). Pour la discussion qui suit, nous revenons au cas de fragments de plastiques purs. Considérons \mathcal{B} bacs. A chaque étape, nous devons traiter un nouveau fragment et choisir dans quel bac il doit être soufflé. Pour éviter de compliquer les notations, nous omettons la référence au temps (indice k), dans le problème d'optimisation.

Dans un cadre déterministe où la nature de la pièce à traiter fr est donnée précisément par un classifieur standard δ , $\delta(fr) \in \Theta$, le problème d'optimisation sous contraintes peut être défini comme suit :

$$\max_{b \in \{1, \dots, \mathcal{B}\}} g_b(\delta(fr)) \quad (37a)$$

$$\text{tel que } h_b(\tau_1^b, \dots, \tau_n^b) \leq 0, \quad b = 1, \dots, \mathcal{B}, \quad (37b)$$

où :

- la fonction objectif (37a) est $g : \Theta \rightarrow \mathbb{R}^+$, avec $g_b(\theta)$ le gain obtenu si un matériau de type $\theta \in \Theta$ est ajouté au bac b .
- τ_i^b est la proportion du matériau de type θ_i dans le bac b après lui avoir ajouté le fragment fr .
- Les contraintes (37b) pour chaque bac b sont exprimées par une fonction $h_b : [0, 1]^n \rightarrow [-1, 1]$. Elles sont de la forme $h_{b,A}(\tau_1, \dots, \tau_n) = \sum_{i \in A} \tau_i - \alpha_b \leq 0$ avec $A \subseteq \Theta$, signifiant que la proportion de matériau de type A ne devrait pas excéder le seuil α_b dans le bac b .

Notons que cette formalisation du problème de tri où la décision est prise à chaque étape est une simplification du problème réel qui lui devrait reposer sur un critère intégral à la fin du processus de tri. Cependant en incluant ces considérations, nous serions confrontés à un problème complexe à plusieurs étapes qui n'est pas réalisable dans la pratique. Dans ce qui suit, nous suggérons une adaptation du problème d'optimisation sous contraintes (37a)-(37b) lorsque l'information sur les proportions est incertaine et/ou imprécise. L'outil classique pour effectuer une optimisation sous incertitude est l'optimisation stochastique. Nous étendrons ce dernier cas au cadre des fonctions de croyance, en considérant l'espérance basée sur l'intégrale de Choquet au lieu de l'espérance classique comme fonction objectif.

Lorsque la composition du fragment fr est incertaine, par exemple donnée par une fonction de masse de probabilité (pmf) $p(\cdot|fr)$ sur Θ . L'optimisation devient alors stochastique, et (37a) est remplacée par

$$\max_{b \in \{1, \dots, \mathcal{B}\}} \mathbb{E}_{p(\cdot|fr)}[g_b]$$

où $\mathbb{E}_{p(\cdot|fr)}$ est l'espérance par rapport à $p(\cdot|fr)$. Remarquons que $p(\cdot|fr)$ peut être converti en probabilité discrète sur des proportions binaires de \mathbb{U} . Afin de vérifier si les contraintes sont respectées, nous devons calculer des probabilités sur des proportions. Nous notons $p_b \textcircled{\mathbb{M}} p(\cdot|fr)$ la distribution de probabilité caractérisant le mélange du contenu du bac (décrit par p_b) et

du nouveau fragment fr (décrit par $p(\cdot|fr)$). Les contraintes (37b) sont remplacées par la probabilité de respect des contraintes :

$$\mathbb{P}_{fr,b}(h_b(\tau_1^b, \dots, \tau_n^b) \leq 0) \geq \eta, \quad b = 1, \dots, \mathcal{B}.$$

où $\mathbb{P}_{fr,b}$ est la mesure de probabilité associée à la distribution de probabilité $p_c \circledast p(\cdot|f)$, et η suffisamment grand. Finalement le problème d'optimisation stochastique sous contraintes est :

$$\max_{b \in \{1, \dots, \mathcal{B}\}} \mathbb{E}_{p(\cdot|fr)}[g_b] \tag{38a}$$

$$\text{tel que } \mathbb{P}_{fr,b}(h_b(\theta_1^b, \dots, \theta_n^b) \leq 0) \geq \eta, \quad b = 1, \dots, \mathcal{B}, \tag{38b}$$

Lorsque l'information sur la nature du fragment est entachée d'imprécision et d'incertitude, nous pouvons utiliser les fonctions de croyance à la place des probabilités. Supposons à présent que nous disposions d'une fonction de masse $m(\cdot|fr)$ portant sur $\Theta = \{\theta_1, \dots, \theta_n\}$ concernant la nature du nouveau fragment fr .

Concernant l'adaptation des contraintes (38b), nous pouvons simplement utiliser l'une des stratégies proposées dans la section 5.2. Par exemple l'adaptation de la procédure prudente consisterait à n'autoriser l'ajout d'un fragment fr dans le bac b que si :

$$Bel^{fr \circledast b}(h_b(\theta_1^b, \dots, \theta_n^b) \leq 0) > \eta,$$

où m^b désigne la fonction de masse décrivant les proportions dans le bac b .

Pour l'adaptation de la fonction objectif (38a), nous pourrions considérer l'intégrale de Choquet associée à $Bel(\cdot|fr)$ (chapitre 2, sous-section 2.1.5). En fonction du gain $g_b(\theta_i)$ obtenu si le type réel du fragment ajouté au bac b est θ_i , nous avons :

$$\max_{b \in \{1, \dots, \mathcal{B}\}} C_{Bel(\cdot|fr)}(g_b(\theta_1), \dots, g_b(\theta_n)). \tag{39}$$

La fonction objectif (39) est une valeur d'espérance basée sur l'intégrale de Choquet où les gains sont pondérés en fonction de notre croyance sur le type de matériau du nouveau fragment. Soit $x_{(1)} = \min_{i=1,n} g_b(\theta_i)$, \dots , $x_{(n)} = \max_{i=1,n} g_b(\theta_i)$ tel que $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$, alors cette valeur d'espérance garantit $x_{(1)}$ sûrement et lui ajoute les écarts $x_{(i)} - x_{(i-1)}$ pondérés par $Bel(\{\theta_{(i)}, \dots, \theta_{(n)}\}|fr)$.

Finalement, nous obtenons le problème d'optimisation suivant :

$$\max_{b \in \{1, \dots, \mathcal{B}\}} C_{Bel(\cdot|fr)}(g_b(\theta_1), \dots, g_b(\theta_n)), \tag{40a}$$

$$\text{tel que } Bel^{fr \circledast b}(h_b(\theta_1^b, \dots, \theta_n^b) \leq 0) > \eta, \quad b = 1, \dots, \mathcal{B}. \tag{40b}$$

Remarque 3. Des simulations de tri de pixels basées sur les résolutions des problèmes d'optimisation déterministe (37a)-(37b), stochastique (38a)-(38b) et évidentielle (40a)-(40b), ont montré l'intérêt de notre approche évidentielle comparée aux approches déterministe et stochastique dans (Jacquin et al., 2020a). Cependant, dans ce chapitre, avec les données dont nous disposons sur les fragments, nous n'avons pas pu illustrer l'intérêt d'une approche évidentielle par rapport aux approches probabiliste et déterministe. Dans la TABLE 5.5, nous donnons les résultats de classification des 140 fragments pour une classification standard (classification logistique standard des pixels puis vote sur les pixels), une classification probabiliste (classification logistique probabiliste des pixels, moyenne des distributions de probabilités sur tous les

pixels, puis décision avec l'espérance optimale) et une classification évidentielle (classification logistique évidentielle ECLAIR des pixels, moyenne des fonctions de masses sur tous les pixels, application de la pignistique, puis décision avec l'espérance optimale).

	classification déterministe	classification probabiliste	classification évidentielle
<i>acc</i>	0.929	0.929	0.933

TABLE 5.5 – Taux de justesse (taux de prédictions exactes) *acc* de classification de fragments

Bien que ces résultats soient en faveur du cadre évidentiel, l'intérêt dans le problème du tri est plus difficile à démontrer car de nombreuses erreurs ne sont pas dues à l'hésitation du classifieur. Dans cette situation, l'apport de notre approche évidentielle est faible. En effet, les erreurs ne sont pas évitées et l'attitude prudente empêche de compenser rapidement les erreurs. Notre approche aurait plus d'intérêt si l'incertitude et l'imprécision était plus importante comme lors de notre précédente étude ([Jacquin et al., 2020a](#)) qui traitait le problème de tri de pixels sur d'autres données industrielles. Les performances de la classification évidentielles y sont davantage mises en avant : trier sur la base d'un seul pixel augmente l'incertitude et l'imprécision de l'information que le classifieur doit traiter. Notre classifieur joue alors pleinement son rôle. La taille des fragments (83.5 pixels en moyenne dans l'expérience présentées ci-dessus) remet en question l'hypothèse des fragments monomatières utilisée dans notre approche. Plus les fragments sont petits, plus la probabilité qu'ils soient monomatières est grande et plus l'hypothèse de notre application industrielle est viable. L'étude proposée dans ([Jacquin et al., 2020a](#)) ouvre donc d'intéressantes perspectives car la tendance du tri industriel est de réduire toujours davantage les impuretés dans les plastiques recyclables pour espérer des propriétés mécaniques meilleures et garanties. Cela pourrait passer par une étape de broyage plus poussée car seuls des micro fragments monomatières, potentiellement réduits à un pixel, permettraient de garantir que les plastiques triés pourront être quasi sans impuretés.

Afin d'améliorer la fiabilité de la procédure de tri, nous pourrions envisager que des prélèvements soient effectués à la fin du tri, pour vérifier la qualité de la matière triée, i.e, le respect des contraintes sur les proportions. Des actions pourraient être menées pour rattraper les possibles erreurs de tri qui empêchent de respecter les contraintes et un deuxième problème d'optimisation doit être mis en place pour optimiser ces actions. Nous serions dans ce cas confrontés à un problème d'optimisation évidentielle à plusieurs étapes. Ce type de problème a été traité dans ([Helal et al., 2018](#)). Le problème applicatif traité dans cet article est celui de l'optimisation de tournées de véhicules où les demandes à récolter sont incertaines. Dans une première étape, le problème consiste à satisfaire des demandes à des localisations différentes en minimisant le parcours des véhicules. Dans une seconde étape, les éventuels manquements de la première tournée doivent être corrigés et un problème d'optimisation est prévu pour optimiser ces corrections.

5.5 Conclusion du chapitre

Dans ce chapitre, nous avons présenté une formalisation du problème de tri de plastiques comme un mélange successif de pièces dont la composition doit être maîtrisée. Nous nous sommes d'abord intéressés à des éléments théoriques de manipulation d'information sur des vecteurs de proportions dans le cadre des fonctions de croyance. En particulier, nous avons formalisé la somme exprimant le mélange de pièces dont les compositions sont exprimées par des fonctions de masse. Ensuite, nous avons proposé des stratégies de tri naïve, prudente et flexible basées sur ces éléments théoriques. Nous avons illustré ces stratégies de tri sur des simulations construites à partir de données réelles de fragments de plastiques. Nous avons constaté que l'approche

naïve, ne tenant compte que des informations sur les fragments arrivants (et pas du mélange), ne garantit pas que le tri se terminera avec le niveau de confiance requis dans le respect des contraintes. L'approche prudente qui elle tient compte de l'état de notre connaissance sur les proportions des matériaux dans le bac, garantit que le tri se terminera avec la confiance requise dans le respect des contraintes. Cependant cette stratégie a l'inconvénient de prendre plus de temps à remplir les bacs. L'approche flexible accélère le remplissage des bacs en comparaison de l'approche prudente, en garantissant que notre confiance dans le respect des contraintes sera supérieure au niveau requis à la fin du tri.

Nous avons également abordé la question des généralisations des approches proposées aux cas de fragments qui ne seraient pas monomatières et au cas de plusieurs bacs de tri. Pour la généralisation au cas de fragments qui ne seraient pas monomatières, nous avons suggéré de faire appel à un classifieur multi-labels et nous avons expliqué comment nous pourrions convertir les informations fournies par ce classifieur en informations concernant les proportions. En ce qui concerne la généralisation au cas de plusieurs bacs, nous l'avons formalisée en un problème d'optimisation sous contraintes.

Nous envisageons les perspectives suivantes. D'abord, concernant la partie théorique, nous pourrions tenter d'étendre les outils présentés au cas d'éléments focaux plus généraux, i.e., non convexes. D'autre part, les outils proposés génèrent trop d'éléments focaux ce qui pose un problème pour l'application. Pour rendre compatibles notre modèle théorique et les contraintes de l'application, nous nous intéressons à des méthodes d'approximation des fonctions de croyances ([Tedjini et al., 2020](#)). Concernant la fusion de l'information portée par les pixels des fragments, nous devons étudier plus précisément le choix de l'opérateur de combinaison. Nous pourrions par exemple améliorer cette phase de fusion des informations des pixels en considérant des affaiblissements pour les pixels situés sur les bords des fragments qui sont ceux qui introduisent le plus de bruit dans la décision. En effet, dans l'image hyperspectrale, ces pixels sont à la frontière entre le fragment et le revêtement du convoyeur. Enfin, concernant la stratégie de tri flexible, nous avons peu discuté le choix des paramètres γ et ν utilisés pour estimer nos hypothèses sur les prochains fragments. Ces paramètres devraient fixer le niveau d'optimisme adéquat pour remplir le bac rapidement. Nous pourrions envisager une phase d'apprentissage pour déterminer ces paramètres.

Cette thèse s'intègre dans projet de collaboration entre l'IMT Mines Ales, SUEZ et la société Pellenc st, qui conçoit et commercialise des machines de tri optique pour les déchets ménagers et industriels. La collaboration a pour objectif la conception d'une machine de tri en ligne pour les plastiques issus de la filière Déchets d'Equipements Electriques et Electroniques (DEEE). La contribution de ce travail de recherche concerne l'algorithme de tri qui doit être intégré à la machine pour automatiser la décision du tri en temps réel.

Dans ce travail de thèse, nous nous sommes intéressés au traitement de données incertaines. En effet, les données hyperspectrales qui sont classiquement utilisées par les algorithmes de tri automatique sont entachées d'incertitudes sur la nature du matériau représentée par les spectres (le spectre correspond-t-il à la matière principale du fragment ou à une impureté ? à quel point le vieillissement a modifié le spectre ? quelle est la part des perturbations atmosphériques ? etc.). Nous avons aussi des imprécisions sur la nature du matériau liées en particulier au vieillissement des plastiques à recycler, ce qui rend difficile l'identification précise des plastiques monnayables dans l'économie du recyclage. La prise en compte de ces imperfections de la donnée dans le tri automatique remet en question la pertinence des algorithmes aujourd'hui exploités pour le tri automatique de plastiques dans l'industrie. C'est le challenge que nous avons choisi de relever dans le cadre de cette thèse.

Le problème applicatif décrit dans le chapitre 1 doit répondre au double objectif suivant. D'une part il est nécessaire de respecter des contraintes sur les proportions d'impuretés dans les bacs de tri afin de garantir que les fragments pourront être recyclés en tant que matière à haute valeur marchande. La procédure de tri doit donc détecter au mieux tous les plastiques cibles afin de maximiser la quantité de plastiques soufflés dans les bacs de tri les plus valorisables. Les procédures classiquement utilisées pour résoudre ce problème de tri faisant appel à la classification standard ne garantissent pas que les contraintes sur les taux d'impuretés seront respectées à la fin du tri si les données à traiter sont trop fortement entachées d'incertitude et d'imprécision.

Ce constat nous a donc amenés à considérer des représentations de l'incertitude et de l'imprécision afin d'exploiter toute l'information disponible dans les données hyperspectrales de fragments de plastiques usés.

Dans le chapitre 2, nous avons introduit les cadres existants permettant de manipuler des données incertaines. Nous avons expliqué comment des classifieurs alternatifs, que nous appelons *classifieurs prudents*, peuvent exploiter ces cadres pour remettre en question la forme classique des sorties des classifieurs de l'industrie qui se résume systématiquement à une classe précise. Nous avons proposé des sorties alternatives exprimant les doutes/hésitations du classifieur par

le biais d'un sous-ensemble de classes potentielles lorsque prendre une décision plus précise sur la base du contenu informationnel des données exploitées nécessiterait une prise de risque non explicable. Dans le chapitre 3, nous nous sommes donc concentrés sur une catégorie de classifieurs prudents, appelés *classifieurs imprécis* capables de fournir des ensembles de classes candidates lorsque l'information est insuffisante pour se décider sur une classe précise. Ce type de sorties peut être directement exploité dans le problème du tri de plastiques en ajoutant des bacs de tri dédiés à des familles de plastiques : à défaut d'identifier précisément un plastique, le classifieur peut identifier sa famille, la valeur de recyclage associée à ce bac moins sélectif sera moindre, mais toujours supérieure à celle d'un bac de rejets où tous les plastiques non identifiés seraient envoyés indépendamment de leur nature. Cette prudence permet de minimiser les erreurs de tri dans les bacs dédiés à des plastiques précis, puisqu'en cas de doute, un fragment peut être soufflé dans un bac dédié à un ensemble de plastiques. En s'autorisant une sortie imprécise, on accroît la certitude. Par ailleurs, les contenus des bacs dédiés à des familles de plastiques peuvent également être valorisés en tant que tels pour des applications de recyclage moins exigeantes. Dans ce chapitre, nous avons proposé un état de l'art sur les classifieurs imprécis existants et nous avons énoncé cinq critères pour le choix d'un classifieur imprécis : la prise en compte de l'information imparfaite ; la complexité algorithmique ; la flexibilité ; la généralité ; la performance.

Dans le chapitre 4, nous avons proposé un classifieur imprécis évidentiel, flexible car la décision est paramétrée par une grandeur qui permet de gérer le degré de prudence et générique puisque notre approche peut se construire à partir de n'importe quel classifieur probabiliste. Le choix du cadre des fonctions de croyances pour notre approche se justifie par la présence simultanée d'imprécisions et d'incertitude et la nécessité de fusionner des informations disponibles au niveau des pixels à l'échelle du fragment. Notre approche baptisée ECLAIR pour Evidential CLAssification of Incomplete data via Imprecise Relabelling a d'abord été présentée dans l'article ([Jacquin et al., 2019](#)). Nous avons ensuite mis en relief ses avantages par rapport à une approche de tri classique dans le cadre de notre application de tri des plastiques sombres ([Jacquin et al., 2020d](#)). Notre idée pour cette approche était de remettre en cause et de transformer l'espace de discernement considéré par un classifieur probabiliste : au lieu de ne considérer que les classes précises initiales, nous l'entraînons sur des sous-ensembles de classes obtenus via un ré-étiquetage imprécis explicitant l'imprécision présente dans les caractéristiques dans les zones de chevauchements en particulier. Lorsque ce classifieur est appliqué à de nouveaux exemples, il renvoie alors une fonction de croyance. Enfin, nous obtenons le classifieur imprécis en déterminant la *sortie imprécise* i.e., l'ensemble de classes candidates qui minimise une fonction de risque. Cette fonction de risque tient compte de l'erreur de spécialisation injustifiée (i.e., le classifieur a pris un pari sur une classe précise sans raison sémantique), autrement dit elle pénalise des sorties plus précises que ce que l'information disponible ne devrait le permettre. Les expérimentations de ce chapitre ont montré des performances compétitives avec deux approches de l'état de l'art pour 8 jeux de données ayant peu de classes.

Le dernier chapitre de cette thèse est consacré à l'intégration des spécificités du problème de tri, i.e., la prise en compte de l'aspect séquentiel des décisions et contraintes spécifiques sur les proportions d'impuretés tolérées dans les bacs de tri. Dans ce chapitre, nous avons alors présenté des outils pour représenter des informations concernant des proportions imprécises de composants dans un mélange dans le cadre des fonctions de croyances. Ces outils nous ont permis de proposer des stratégies de tri, i.e., remplissage des bacs avec des proportions de différentes matières, que nous avons illustrées sur des simulations de tri à partir de données acquises industriellement. Les éléments théoriques concernant nos propositions de stratégies de tri ont fait l'objet d'une publication ([Jacquin et al., 2020c](#)) avec un problème de tri de pièces qui ne sont pas monomatériaux simulés. Dans les articles suivants ([Jacquin et al., 2020a,b](#)), nous avons

montré l'intérêt de notre approche de tri évidentielle comparée à des approches probabiliste et déterministe sur nos premières données acquises industriellement. Dans ces premiers travaux, nous n'avons pas effectué la phase de fusion et nous avons considéré que les fragments de plastiques étaient tous constitués d'un unique pixel. Ainsi, l'hypothèse de fragment monomatière utilisée dans notre application industrielle était plus facilement vérifiée lorsque le fragment était petit, réduit au pixel dans nos études préliminaires. Les données utilisées par la suite dans notre travail de recherche souffrent moins de l'imperfection des mesures pour deux raisons majeures. Il y a eu d'abord des améliorations techniques significatives de la machine réalisées par l'entreprise Pellenc ST. Ensuite, la taille conséquente des fragments (en moyenne les informations de 83.5 pixels sont fusionnées pour un fragment) permet la mise en œuvre de stratégie de votes majoritaires simples pour déterminer la nature d'un fragment réduisant ainsi les risques d'erreur. En raison de ces améliorations techniques et de la taille des fragments, l'approche déterministe finit par être préférable aux approches probabiliste ou évidentielle sur notre nouveau jeu de données. Cependant, il faut avoir en tête que l'hypothèse de fragment monomatière ne saurait être toujours vérifiée sur des fragments de cette taille, et qu'ainsi, même si un simple vote des pixels du fragment permet d'affirmer que celui-ci est très majoritairement de type A, cette décision n'empêchera pas en pratique qu'une quantité infinitésimale d'impuretés ne soit soufflée dans le bac de type A. Le problème n'est plus l'imprécision des données, mais les résidus d'impuretés qu'on laisse mécaniquement passer avec le fragment pourtant très majoritairement de type A. La seule alternative pour passer outre l'envoi de résidus d'impuretés dans un bac de tri ne relève alors plus de la théorie de l'information, et consiste à broyer plus finement les fragments pour être certain qu'au final, ils soient si petits que l'hypothèse monomatière redevienne légitime et que si la décision de tri est fiable alors le bac contiendra effectivement un plastique quasi pur. Cette remarque remet ainsi en perspective nos travaux initiaux au niveau du pixel où nous avons montré les apports de notre approche lorsque les données sont critiquement imprécises et incertaines. L'avenir du recyclage de plastiques de seconde vie réside vraisemblablement sur une étape de broyage fine qui permettra de tendre vers un tri sans impuretés, la taille des fragments réduite au pixel redonnera alors toute son importance à des approches comme celle que nous avons proposée où le classifieur doit prendre une décision de tri sur la seule information d'un simple pixel. Notre proposition de classification prudente reprendra alors tout son sens.

Dans la suite, nous donnons nos principales perspectives d'amélioration des travaux de recherche menés durant cette thèse.

Le principal axe de nos perspectives reste la gestion de la complexité de nos algorithmes. La gestion de la complexité est d'autant plus critique que l'application visée est en temps réel. Concernant notre approche de classification imprécise, en l'absence d'algorithmes efficaces pour la recherche de la solution dans la minimisation du risque proposé, nous devons parcourir tous les sous-ensembles de classes candidates envisageables. Une perspective d'amélioration serait d'étudier des algorithmes efficaces pour la minimisation du risque proposée à l'instar de (Mortier *et al.*, 2019) dans le cadre probabiliste. En ce qui concerne les stratégies de tri proposées, le problème de la complexité y est encore plus prégnant. En effet, nous avons vu que ces procédures étaient exponentielles en fonction du nombre d'itérations. Nous envisageons de réduire le nombre d'éléments focaux utilisés pour décrire notre connaissance sur les proportions des différents plastiques dans un bac en effectuant des approximations des fonctions de croyances compatibles avec les contraintes de notre application, par exemple en nous basant sur les travaux (Tedjini *et al.*, 2020).

Le second grand axe de perspectives est l'optimisation des approches de classification imprécise (ECLAIR) et la stratégie "flexible" de tri proposée. Dans l'approche ECLAIR, nous avons introduit plusieurs possibilités pour effectuer le ré-étiquetage, nous devons étudier de façon

encore plus approfondie les avantages et inconvénients des méthodes de ré-étiquetage imprécis proposées. Concernant la stratégie de tri flexible, nous avons peu discuté le choix des paramètres γ et ν utilisés pour estimer nos hypothèses sur les prochains fragments. Ces paramètres devraient fixer le niveau d'optimisme adéquat pour remplir le bac rapidement. Nous pourrions envisager une phase d'apprentissage pour déterminer des paramètres optimaux.

Articles de conférence

- JACQUIN, L., IMOSSATEN, A., TROUSSET, F., MONTMAIN, J. et PERRIN, D. (2019). Evidential classification of incomplete data via imprecise relabelling : Application to plastic sorting. In Ben Amor, N., Quost, B. et Theobald, M., éditeurs : Scalable Uncertainty Management, pages 122–135, Cham. Springer International Publishing. [86](#), [146](#)
- JACQUIN, L., IMOSSATEN, A., DESTERCKE, S., TROUSSET, F., MONTMAIN, J. et PERRIN, D. (2020c). Manipulating focal sets on the unit simplex : application to plastic sorting. In FUZZ 2020-IEEE International Conference on Fuzzy Systems. [146](#)
- JACQUIN, L., IMOSSATEN, A. et DESTERCKE, S. (2020a). Handling mixture optimisation problem using cautious predictions and belief functions. In International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, pages 394–407. Springer. [140](#), [141](#), [142](#), [146](#)
- JACQUIN, L., IMOSSATEN, A. et DESTERCKE, S. (2020b). Traitement du problème d’optimisation de mélange dans le cas de données incertaines via les fonctions de croyance. In LFA’2020-29èmes Rencontres Francophones sur la Logique Floue et ses Applications. [146](#)

Article de journal

- JACQUIN, L., IMOSSATEN, A., TROUSSET, F., PERRIN, D. et MONTMAIN, J. (2020d). Control of waste fragment sorting process based on mir imaging coupled with cautious classification. Resources, Conservation and Recycling, page 105258. [146](#)

Jeux de données

Les jeux de données utilisés dans cette thèse proviennent du site *UCI Machine Learning Repository*. Ci-dessus nous présentons brièvement ces jeux de données. Notons que nous avons ajouté un jeu de données *plastics* qui correspond aux spectres d'apprentissage présentés dans le chapitre 1. Dans la première colonne nous donnons la lettre utilisée dans le chapitre 4.

	nom	# exemples	# caractéristiques	# classes
	ecoli	336	7	8
a	forestTypes	198	27	4
b	glass	214	9	6
c	iris	150	4	3
	optdigit	3823	64	10
d	plastics	23389	154	4
e	seeds	210	7	3
	segmentation	210	19	7
f	vehicle	846	18	4
	vowel	990	10	11
g	wallFollowing	5456	24	4
h	wine	178	13	3
	yeast	1484	8	10

Détails concernant les expérimentations

Les expérimentations ont été menées sur R version 4.0.2 sur un iMac13,1 Intel Core i5 avec une vitesse du processeur 2,7 GHz, 4 coeurs et 8 Go de Mémoire.

Principales *routines* (et packages) R utilisés par chapitre

Chapitre 2

- Classifieur naïf bayésien : *naiveBayes* (e1071),
- *k*-nn : *knn3* (caret),
- CART : *rpart* (rpart),
- Forêt d'arbres décisionnels : *randomForest* (randomForest),
- SVM : *svm* (e1071),
- Réseau de neurones artificiels : *nnet* (nnet),
- *EkNNfit* et *EkNNval* (evclass),
- Régression logistique : *multinom* (nnet),
- PCA : *princomp* (MASS).

Chapitre 3

- Classifieur naïf bayésien : *naiveBayes* (e1071).

Chapitre 4

- Discrétisation utilisant le principe de longueur de description minimale : *mdlp* (discretization),
- SVM : *svm* (e1071),

- Classifieur naïf bayésien : *naiveBayes* (e1071),
- k -nn : *knn3* (caret).

Chapitre 5

- Régression logistique : *multinom* (nnet).

Loi multinomiale et loi de Dirichlet

Dans la suite nous présentons deux lois que nous rencontrons au cours de ce manuscrit.

Définition 1. Loi multinomiale Considérons r expériences indépendantes ayant v sorties possibles, chacune associée à une probabilité p_i , $i \in \{1, \dots, v\}$ telle que $\sum_{i=1}^v p_i = 1$. Notons R_i la variable aléatoire correspondant au nombre de fois que la i -ème sortie est obtenue lors des r expériences. Nous avons $\sum_{i=1}^v R_i = r$. Alors le vecteur (R_1, \dots, R_v) est caractérisé par la distribution de probabilité suivante :

$$P(R_1 = r_1, \dots, R_v = r_v) = \frac{r!}{r_1! \dots r_v!} p_1^{r_1} \dots p_v^{r_v}.$$

Cette distribution de probabilité est appelée loi multinomiale.

Exemple 1. *Donnons un exemple d'évènement dont la probabilité se calcule par la loi multinomiale. Considérons 10 expériences ($r = 10$) de jet d'un dé à 6 faces ($v = 6$) non truqué ($p_1 = \dots = p_6 = 1/6$). Alors la probabilité de l'évènement "Obtenir 5 fois la face 1, 3 fois la face 4 et 2 fois la face 5" ($P(R_1 = 5, R_2 = 0, R_3 = 0, R_4 = 3, R_5 = 2, R_6 = 0)$) se calcule par $\frac{10!}{5!3!2!} 1/6^{10} \simeq 4.167619 \times 10^{-5}$.*

Définition 2. Loi de Dirichlet La loi de Dirichlet s'applique à un vecteur de variables aléatoires continues $\Lambda_1, \dots, \Lambda_d$ chacune à valeur dans $[0, 1]$ et telles que $\sum_{i=1}^d \Lambda_i = 1$. Cette loi est paramétrée par un vecteur $\alpha = (\alpha_1, \dots, \alpha_d)$ pour lequel chaque $\alpha_i > 0$. La densité de Dirichlet est telle que pour $(\lambda_1, \dots, \lambda_d) \in (\Lambda_1, \dots, \Lambda_d)$:

$$f(\lambda_1, \dots, \lambda_d; \alpha) \propto \prod_{i=1}^d \lambda_i^{\alpha_i - 1},$$

où \propto signifie "est proportionnel à".

 Classifieur crédal naïf

Le classifieur *crédal naïf* ou *Naive credal classifier* (NCC) en anglais, est une extension à la classification imprécise, dans le cadre des probabilités imprécises du classifieur standard bayésien naïf ou naive bayes classifier (NBC) proposé dans (Zaffalon, 1999). Le classifieur NCC repose sur la même condition d'indépendance conditionnellement aux classes des attributs que pour NBC (voir le chapitre 2). Considérons X_i , $i = 1, \dots, d$ les d attributs représentant les objets à classer et C la variable aléatoire représentant la classe associée à un objet \mathbf{x} . La condition d'indépendance conditionnellement aux classes permet de déterminer la probabilité que le vecteur aléatoire $X = (X_1, \dots, X_d)$ soit égal à $\mathbf{x} = (x_1, \dots, x_d)$ conditionnellement à une des classes en utilisant les probabilités que chaque variable d'attribut soit égale à la valeur correspondante dans le vecteur \mathbf{x} conditionnellement à cette classe comme suit

$$P(X = \mathbf{x} | C = \theta) = \prod_{i=1}^d P(X_i = x_i | C = \theta). \quad (41)$$

Dans le cas du classifieur NBC une seule pmf $p : \Theta \rightarrow [0, 1]$ est déterminée pour représenter les chances de chaque classe d'être la vraie classe de \mathbf{x} . La nouveauté apportée par le classifieur NCC est de considérer un ensemble de pmf sur Θ , i.e., un ensemble credal \mathcal{P} .

Dans la suite, nous allons présenter la version du NCC proposée dans (Zaffalon, 2001), pour laquelle l'ensemble crédal est inféré grâce au modèle de Dirichlet imprécis (version imprécise de la densité de Dirichlet introduite dans l'annexe C).

Notations et expression de la vraisemblance Notons C la variable de la classe prenant ses valeurs sur Θ et la variable des d attributs $X = (X_1, \dots, X_d)$ à valeurs dans $\mathcal{X} = (\mathcal{X}_1, \dots, \mathcal{X}_d)$ où chaque \mathcal{X}_i , $i = 1, \dots, d$ est un ensemble fini. Considérons L exemples observés dont les valeurs des attributs et des classes sont connues :

$$(\mathbf{x}^l, \theta^l)_{1 \leq l \leq L},$$

telles que $\forall l \in \{1, \dots, L\}$ $\mathbf{x}^l \in \mathcal{X} = (\mathcal{X}_1, \dots, \mathcal{X}_d)$ et $\theta^l \in \Theta$. Nous avons besoin d'introduire les notations suivantes afin de faciliter l'exposé :

- $\lambda_\theta = P(C = \theta)$,
- $\lambda_{\theta, \mathbf{x}} = P(C = \theta \text{ et } X = \mathbf{x})$,

- $\lambda_{x_i|\theta} = P(X_i = x_i|C = \theta)$,
- $\lambda_{\mathbf{x}|\theta} = P(X = \mathbf{x}|C = \theta)$.

Notons qu'avec ces notations, l'équation (41) se ré-écrit $\lambda_{\mathbf{x}|\theta} = \prod_{i=1}^d \lambda_{x_i|\theta}$. Nous en déduisons :

$$\lambda_{\theta, \mathbf{x}} = \lambda_{\theta} \prod_{i=1}^d \lambda_{x_i|\theta}.$$

Les notations suivantes concernent des fréquences d'observations parmi les L exemples observés.

- $\text{occ}(\theta)$: la fréquence d'observation de la classe θ .
- $\text{occ}(x_i|\theta)$: la fréquence d'observation de la valeur x_i de la caractéristique \mathcal{X}_i pour la classe θ .

Notons que pour tout $\theta \in \Theta$, $i \in \{1, \dots, d\}$ et $x_i \in \mathcal{X}_i$, avons :

$$\sum_{\theta \in \Theta} \text{occ}(\theta) = L \quad \text{et} \quad \sum_{x_i \in \mathcal{X}_i} \text{occ}(x_i|\theta) = \text{occ}(\theta).$$

Enfin nous adopterons également les notations suivantes :

- \mathbf{occ} : le vecteur de toutes les fréquences des exemples obtenus.
- $\boldsymbol{\lambda}$: le vecteur dont les éléments sont les probabilités $\lambda_{\theta, \mathbf{x}}$ pour chaque $(\theta, \mathbf{x}) \in \Theta \times \mathcal{X}_1 \times \dots \times \mathcal{X}_d$.

Après avoir observé le vecteur \mathbf{occ} , la fonction de vraisemblance pour le problème s'exprime comme :

$$L(\boldsymbol{\lambda}|\mathbf{occ}) \propto \prod_{\theta \in \Theta} \left[\lambda_{\theta}^{\text{occ}(\theta)} \prod_{i=1}^d \prod_{x_i \in \mathcal{X}_i} \lambda_{x_i|\theta}^{\text{occ}(x_i|\theta)} \right] \quad (42)$$

L'auteur propose une version imprécise de cette vraisemblance en la remplaçant par une vraisemblance paramétrée généralisant la distribution de Dirichlet appelée modèle de Dirichlet imprécis.

Définition 1. Modèle de Dirichlet imprécis (MDI) Nous donnons ici la définition du modèle de Dirichlet imprécis (MDI) proposée dans (Zaffalon, 2001). Une densité a priori dans le modèle de Dirichlet imprécis s'exprime de façon similaire à la formule (42) en remplaçant les fréquences observées $\mathbf{occ}(\cdot)$ par le paramétrage $\mathbf{st}(\cdot) - 1$, avec $s > 0$ fixé et $\mathbf{t}(\cdot)$ un hyper-paramètre :

$$\pi(\boldsymbol{\lambda}|\mathbf{t}, s) \propto \prod_{\theta \in \Theta} \left[\lambda_{\theta}^{\mathbf{st}(\theta)-1} \prod_{i=1}^d \prod_{x_i \in \mathcal{X}_i} \lambda_{x_i|\theta}^{\mathbf{st}(x_i|\theta)-1} \right]$$

Nous reconnaissons un produit de densité de Dirichlet (voir annexe C). Le MDI est défini, pour un s fixé, comme l'ensemble des densités a priori $\pi(\boldsymbol{\lambda}|\mathbf{t}, s)$ telles que \mathbf{t} satisfait des contraintes suivantes pour tout i, x_i et θ :

$$\sum_{\theta \in \Theta} \mathbf{t}(\theta) = 1, \quad \sum_{x_i \in \mathcal{X}_i} \mathbf{t}(x_i|\theta) = \mathbf{t}(\theta), \quad \mathbf{t}(x_i|\theta) > 0.$$

En multipliant la densité a priori et la vraisemblance, nous obtenons une densité a posteriori

sous la même forme que la densité a priori en remplaçant $st(\cdot)$ par $st(\cdot) + occ(\cdot)$. Ensuite nous pouvons montrer que l'espérance de $\lambda_{\theta, \mathbf{x}}$ s'exprime comme

$$E[\lambda_{\theta, \mathbf{x}} | \mathbf{occ}, \mathbf{t}] = P(\theta, \mathbf{x} | \mathbf{occ}, \mathbf{t}) = P(\theta | \mathbf{occ}, \mathbf{t}) \prod_{i=1}^d P(x_i | \theta, \mathbf{occ}, \mathbf{t}), \quad (43)$$

où

$$P(\theta | \mathbf{occ}, \mathbf{t}) = E[\lambda_{\theta} | \mathbf{occ}, \mathbf{t}] = \frac{occ(\theta) + st(\theta)}{L + s}$$

et

$$P(x_i | \theta, \mathbf{occ}, \mathbf{t}) = E[\lambda_{x_i | \theta} | \mathbf{occ}, \mathbf{t}] = \frac{occ(x_i | \theta) + st(x_i | \theta)}{occ(\theta) + st(\theta)}.$$

Pour un exemple \mathbf{x} à classifier, les contraintes sur \mathbf{t} se réduisent à $0 < t(x_i | \theta) < t(\theta)$ pour tout $i = 1, \dots, d$ et $\theta \in \Theta$, et $\sum_{\theta} t(\theta) = 1$. Notons $E[u(\theta) | \mathbf{x}, \mathbf{occ}, \mathbf{t}]$ l'espérance de gain de choisir la classe θ sachant le vecteur d'attributs \mathbf{x} , les données précédemment observées \mathbf{occ} et le vecteur de d'hyperparamètres \mathbf{t} . En choisissant la fonction de gain $0/1$ ¹, nous avons

$$E[u(\theta) | \mathbf{x}, \mathbf{occ}, \mathbf{t}] = P(\theta | \mathbf{x}, \mathbf{occ}, \mathbf{t}) \quad (44)$$

L'auteur infère ensuite l'ordre partiel sur les classes en considérant la relation de préférence appelée *dominance crédale*.

Définition 2. Dominance crédale La classe θ_i est préférée à la classe θ_j pour le critère de la dominance crédale si et seulement si

$$E[U(\theta_i) | \mathbf{x}, \mathbf{occ}, \mathbf{t}] > E[U(\theta_j) | \mathbf{x}, \mathbf{occ}, \mathbf{t}],$$

pour toutes les valeurs de \mathbf{t} respectant des contraintes de MDI.

D'après l'équation (44), cela revient à dire que θ_i est préférée à la classe θ_j si et seulement si

$$P(\theta_i | \mathbf{x}, \mathbf{occ}, \mathbf{t}) > P(\theta_j | \mathbf{x}, \mathbf{occ}, \mathbf{t}),$$

pour les mêmes \mathbf{t} . Ce qui est équivalent à $P(\theta_i, \mathbf{x} | \mathbf{occ}, \mathbf{t}) > P(\theta_j, \mathbf{x} | \mathbf{occ}, \mathbf{t})$ pour les mêmes \mathbf{t} . Afin de vérifier si θ_i est préférée à θ_j , il suffit donc de résoudre le problème d'optimisation suivant

$$\begin{aligned} \inf & \frac{P(\theta_i, \mathbf{x} | \mathbf{occ}, \mathbf{t})}{P(\theta_j, \mathbf{x} | \mathbf{occ}, \mathbf{t})}, \\ & \sum_{\theta} t(\theta) = 1, \\ & 0 < t(x_i | \theta) < t(\theta) \quad \forall i = 1, \dots, d, \theta \in \Theta. \end{aligned}$$

Ensuite si le résultat est strictement supérieur à 1, avons θ_i préférée à θ_j . En utilisant l'équation (43), l'auteur reformule ce problème d'optimisation et propose une résolution exacte. Finalement la complexité de classification d'un exemple $\mathbf{x} = (x_1, \dots, x_d)$ (impliquant le calcul des préférences des classes deux à deux) est $\mathcal{O}(dn^2)$ avec $n = |\Theta|$.

1. Cette fonction de gain est la soustraction à 1 de la fonction de perte $0/1$, c'est à dire $u(\theta_j, \theta_i) = 0$, si $i \neq j$, et $u(\theta_j, \theta_j) = 1$

Détails des résultats du chapitre 4

	NCC	NONDET	ECLAIR
forestTypes	1.00	0.95	0.96
glass	0.63	0.51	0.62
iris	1.00	0.97	0.97
plastics	1.00	0.49	0.49
seeds	1.00	0.91	0.91
vehicle	1.00	0.58	0.68
wallFollowing	1.00	0.58	0.73
wine	1.00	0.98	0.98

TABLE E.1 – Récompenses acc NONDET et ECLAIR basés sur Naive Bayes avec $\beta = 1$ et NCC avec $s = 1$ - résultats moyens obtenus sur 10 expériences de validations croisées à 10 blocs

	NCC	NONDET	ECLAIR
forestTypes	0.37	0.95	0.95
glass	0.19	0.45	0.49
iris	0.47	0.96	0.96
plastics	0.36	0.48	0.48
seeds	0.47	0.90	0.90
vehicle	0.36	0.48	0.53
wallFollowing	0.36	0.54	0.59
wine	0.47	0.97	0.97

TABLE E.2 – Récompenses u_{65} NONDET et ECLAIR basés sur Naive Bayes avec $\beta = 1$ et NCC avec $s = 1$ - résultats moyens obtenus sur 10 expériences de validations croisées à 10 blocs

	NCC	NONDET	ECLAIR
forestTypes	0.48	0.95	0.95
glass	0.24	0.48	0.54
iris	0.60	0.96	0.96
plastics	0.48	0.48	0.48
seeds	0.60	0.91	0.91
vehicle	0.48	0.52	0.59
wallFollowing	0.48	0.56	0.64
wine	0.60	0.98	0.98

TABLE E.3 – Récompenses u_{80} NONDET et ECLAIR basés sur Naive Bbayes avec $\beta = 1$ et NCC avec $s = 1$ - résultats moyens obtenus sur 10 expériences de validations croisées à 10 blocs

	NCC	NONDET	ECLAIR
forestTypes	0.25	0.95	0.95
glass	0.13	0.42	0.44
iris	0.34	0.95	0.95
plastics	0.25	0.48	0.48
seeds	0.33	0.90	0.90
vehicle	0.25	0.44	0.47
wallFollowing	0.25	0.53	0.54
wine	0.33	0.97	0.97

TABLE E.4 – Récompenses $acc_{discounted}$ NONDET et ECLAIR basés sur Naive Bayes avec $\beta = 1$ et NCC avec $s = 1$ - résultats moyens obtenus sur 10 expériences de validations croisées à 10 blocs

- ABELLAN, J. et MASEGOSA, A. R. (2012). Imprecise classification with credal decision trees. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 20(05):763–787. 76, 84
- ABELLÁN, J. et MORAL, S. (2005). Upper entropy of credal sets. applications to credal classification. International Journal of Approximate Reasoning, 39(2-3):235–255. 86, 89
- ADEME, E. F. (2019). Rapport annuel du registre des déchets d'équipements électriques et électroniques. Groupement Deloitte Développement Durable, Alice DEPROUW (In Extenso Innovation Croissance), Marion JOVER (IEIC), Mathilde BORIE (IEIC). 2019., page 112. xiii, 6, 7, 8
- ALLAIS, M. (1953). Le comportement de l'homme rationnel devant le risque : critique des postulats et axiomes de l'école américaine. Econometrica : Journal of the Econometric Society, pages 503–546. 24
- ANTONUCCI, A., BRÜHLMANN, R., PIATTI, A. et ZAFFALON, M. (2009). Credal networks for military identification problems. International Journal of Approximate Reasoning, 50(4):666–679. 25
- ANTONUCCI, A., CORANI, G. et BERNASCHINA, S. (2012). Active learning by the naive credal classifier. 64
- APPRIOU, A. (1998). Uncertain data aggregation in classification and tracking processes. In Aggregation and Fusion of imperfect information, pages 231–260. Springer. 51
- BAE, J.-S., OH, S.-K., PEDRYCZ, W. et FU, Z. (2019). Design of fuzzy radial basis function neural network classifier based on information data preprocessing for recycling black plastic wastes : comparative studies of ATR FT-IR and raman spectroscopy. Applied Intelligence, 49(3):929–949. 11
- BARBIER, S., PERRIER, S., FREYERMUTH, P., PERRIN, D., GALLARD, B. et GILON, N. (2013). Plastic identification based on molecular and elemental information from laser induced breakdown spectra : a comparison of plasma conditions in view of efficient sorting. Spectrochimica Acta Part B : Atomic Spectroscopy, 88:167–173. 11
- BERGER, J. O. (1990). Robust bayesian analysis : sensitivity to the prior. Journal of statistical planning and inference, 25(3):303–328. 25

- BERKSON, J. (1944). Application of the logistic function to bio-assay. Journal of the American statistical association, 39(227):357–365. 46
- Blanch-Perez-del NOTARIO, C., SAEYSB, W. et LAMBRECHTSC, A. (2019). Hyperspectral imaging for textile sorting in the visible–near infrared range. Journal of Spectral Imaging, 8. 17
- BOSER, B. E., GUYON, I. M. et VAPNIK, V. N. (1992). A training algorithm for optimal margin classifiers. In Proceedings of the fifth annual workshop on Computational learning theory, pages 144–152. 44
- BREIMAN, L., FRIEDMAN, J., STONE, C. et OLSHEN, R. (1984). Classification and Regression Trees. The Wadsworth and Brooks-Cole statistics-probability series. Taylor & Francis. 44
- BUCKLEY, J. J. et HAYASHI, Y. (1994). Fuzzy neural networks : A survey. Fuzzy sets and systems, 66(1):1–13. 53
- CAMERER, C. et WEBER, M. (1992). Recent developments in modeling preferences : Uncertainty and ambiguity. Journal of risk and uncertainty, 5(4):325–370. 26
- CHAWLA, N. V., BOWYER, K. W., HALL, L. O. et KEGELMEYER, W. P. (2002). Smote : synthetic minority over-sampling technique. Journal of artificial intelligence research, 16:321–357. 46, 96
- CHEN, P. et SUN, Z. (1991). A review of non-destructive methods for quality evaluation and sorting of agricultural products. Journal of Agricultural Engineering Research, 49:85–98. 9
- CHOW, C. (1970). On optimum recognition error and reject tradeoff. IEEE Transactions on information theory, 16(1):41–46. 61, 70, 71
- CHOW, C.-K. (1957). An optimum character recognition system using decision functions. IRE Transactions on Electronic Computers, (4):247–254. 61, 70, 71
- CORANI, G. et ZAFFALON, M. (2008). Credal model averaging : an extension of bayesian model averaging to imprecise probabilities. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pages 257–271. Springer. 86, 89
- CORANI, G. et ZAFFALON, M. (2009). Lazy naive credal classifier. In Proceedings of the 1st ACM SIGKDD Workshop on Knowledge Discovery from Uncertain Data, pages 30–37. ACM. 86, 89
- CORTES, C. et VAPNIK, V. (1995). Support-vector networks. Machine learning, 20(3):273–297. 44
- COVER, T. et HART, P. (1967). Nearest neighbor pattern classification. IEEE transactions on information theory, 13(1):21–27. 43
- COZ, J. J. d., DÍEZ, J. et BAHAMONDE, A. (2009). Learning nondeterministic classifiers. Journal of Machine Learning Research, 10(Oct):2273–2293. 72, 78, 86, 89, 97, 106
- CUCCHIELLA, F., D’ADAMO, I., KOH, S. L. et ROSA, P. (2015). Recycling of weees : An economic assessment of present and future e-waste streams. Renewable and Sustainable Energy Reviews, 51:263–272. 6

- DE CAMPOS, L. M., HUETE, J. F. et MORAL, S. (1994). Probability intervals : a tool for uncertain reasoning. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2(02):167–196. 31, 73, 117, 125
- DEMPSTER, A. P. (2008). Upper and lower probabilities induced by a multivalued mapping. In Classic Works of the Dempster-Shafer Theory of Belief Functions, pages 57–72. Springer. 74
- DENOEU, T. (1995). A k-nearest neighbor classification rule based on dempster-shafer theory. IEEE transactions on systems, man, and cybernetics, 25(5):804–813. 51, 60, 99
- DENOEU, T. (2000). A neural network classifier based on dempster-shafer theory. IEEE Transactions on Systems, Man, and Cybernetics-Part A : Systems and Humans, 30(2):131–150. 51, 60
- DENOEU, T. (2008). Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. Artificial Intelligence, 172(2-3):234–264. 35
- DENOEU, T. (2019). Decision-making with belief functions : A review. International Journal of Approximate Reasoning, 109:87–110. 52, 74, 100
- DER KIUREGHIAN, A. et DITLEVSEN, O. (2009). Aleatory or epistemic? does it matter? Structural safety, 31(2):105–112. 26
- DESTERCKE, S. (2008). Représentation et combinaison d’informations incertaines : nouveaux résultats avec applications aux études de sûreté. Thèse de doctorat, Université de Toulouse, Université Toulouse III-Paul Sabatier. 25
- DUBOIS, D. et PRADE, H. (1988). Representation and combination of uncertainty with belief functions and possibility measures. Computational intelligence, 4(3):244–264. 36
- DUBOIS, D. et PRADE, H. (2006). Représentations formelles de l’incertain et de l’imprécis. Concepts et méthodes pour l’aide à la décision, 1:11–165. 24, 26, 27
- DUBOIS, D. et PRADE, H. (2012). Possibility theory : an approach to computerized processing of uncertainty. Springer Science & Business Media. 25
- DUBUISSON, B. et MASSON, M. (1993). A statistical decision rule with incomplete knowledge about classes. Pattern recognition, 26(1):155–165. 48, 71
- DUDA, R. O., HART, P. E. et al. (1973). Pattern classification and scene analysis, volume 3. Wiley New York. 43, 60
- ELLSBERG, D. (1961). Risk, ambiguity, and the savage axioms. The quarterly journal of economics, pages 643–669. 24
- FAYYAD, U. et IRANI, K. (1993). Multi-interval discretization of continuous-valued attributes for classification learning. 106
- FISHER, R. A. (1935). The fiducial argument in statistical inference. Annals of eugenics, 6(4):391–398. 62
- FISHER, R. A. (1936). The use of multiple measurements in taxonomic problems. Annals of eugenics, 7(2):179–188. 44, 47
- FLOREA, M. C., JOUSSELME, A.-L., BOSSÉ, É. et GRENIER, D. (2009). Robust combination rules for evidence theory. Information Fusion, 10(2):183–197. 36

- FÜRNRKRAZ, J. et HÜLLERMEIER, E. (2010). Preference learning and ranking by pairwise comparison. In Preference learning, pages 65–82. Springer. 60
- GEBEL, M. (2009). Multivariate calibration of classifier scores into the probability space. Thèse de doctorat, Citeseer. 50
- GILBOA, I. et SCHMEIDLER, D. (1994). Additive representations of non-additive measures and the choquet integral. Annals of Operations Research, 52(1):43–65. 52
- GUPTA, S. S. (1965). On some multiple decision (selection and ranking) rules. Technometrics, 7(2):225–245. 86
- HA, T. M. (1997). The optimum class-selective rejection rule. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(6):608–615. xiv, 62, 70, 71, 86, 89
- HARISPE, S., IMOUSSATEN, A., TROUSSET, F. et MONTMAIN, J. (2015). On the consideration of a bring-to-mind model for computing the information content of concepts defined into ontologies. In Fuzzy Systems (FUZZ-IEEE), 2015 IEEE International Conference on, pages 1–8. IEEE. 69
- HELAL, N., PICHON, F., PORUMBEL, D., MERCIER, D. et LEFÈVRE, É. (2018). The capacitated vehicle routing problem with evidential demands. International Journal of Approximate Reasoning, 95:124–151. 142
- HÉRAULT, J. et ANS, B. (1984). Réseau de neurones à synapses modifiables : Décodage de messages sensoriels composites par apprentissage non supervisé et permanent. Comptes rendus des séances de l'Académie des sciences. Série 3, Sciences de la vie, 299(13):525–528. 47
- HÉRAULT, J., JUTTEN, C. et ANS, B. (1985). Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé. In 10 Colloque sur le traitement du signal et des images, FRA, 1985. GRETSI, Groupe d'Etudes du Traitement du Signal et des Images. 47
- HO, T. K. (1995). Random decision forests. In Proceedings of 3rd international conference on document analysis and recognition, volume 1, pages 278–282. IEEE. 44
- HOTELLING, H. (1933). Analysis of a complex of statistical variables into principal components. Journal of educational psychology, 24(6):417. 47
- HÜLLERMEIER, E. et BRINKER, K. (2008). Learning valued preference structures for solving classification problems. Fuzzy Sets and Systems, 159(18):2337–2352. xiii, 48, 49, 60, 75, 86
- HÜLLERMEIER, E., DESTERCCKE, S. et COUSO, I. (2019). Learning from imprecise data : adjustments of optimistic and pessimistic variants. In International Conference on Scalable Uncertainty Management, pages 266–279. Springer. 99
- ISHIBUCHI, H. et NII, M. (2000). Neural networks for soft decision making. Fuzzy Sets and Systems, 115(1):121–140. 75, 86
- ISHIBUCHI, H., NOZAKI, K., YAMAMOTO, N. et TANAKA, H. (1995). Selecting fuzzy if-then rules for classification problems using genetic algorithms. IEEE Transactions on fuzzy systems, 3(3):260–270. 54

- JACQUIN, L., IMOSSATEN, A. et DESTERCKE, S. (2020a). Handling mixture optimisation problem using cautious predictions and belief functions. In International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, pages 394–407. Springer. 140, 141, 142, 146
- JACQUIN, L., IMOSSATEN, A. et DESTERCKE, S. (2020b). Traitement du problème d'optimisation de mélange dans le cas de données incertaines via les fonctions de croyance. In LFA'2020-29èmes Rencontres Francophones sur la Logique Floue et ses Applications. 146
- JACQUIN, L., IMOSSATEN, A., DESTERCKE, S., TROUSSET, F., MONTMAIN, J. et PERRIN, D. (2020c). Manipulating focal sets on the unit simplex : application to plastic sorting. In FUZZ 2020-IEEE International Conference on Fuzzy Systems. 146
- JACQUIN, L., IMOSSATEN, A., TROUSSET, F., MONTMAIN, J. et PERRIN, D. (2019). Evidential classification of incomplete data via imprecise relabelling : Application to plastic sorting. In BEN AMOR, N., QUOST, B. et THEOBALD, M., éditeurs : Scalable Uncertainty Management, pages 122–135, Cham. Springer International Publishing. 86, 146
- JACQUIN, L., IMOSSATEN, A., TROUSSET, F., PERRIN, D. et MONTMAIN, J. (2020d). Control of waste fragment sorting process based on mir imaging coupled with cautious classification. Resources, Conservation and Recycling, page 105258. 146
- JAFFRAY, J.-Y. (1988). Application of linear utility theory to belief functions. In International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, pages 1–8. Springer. 53
- JEFFREYS, H. (1946). An invariant form for the prior probability in estimation problems. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 186(1007):453–461. 29
- JIAO, L., PAN, Q., DENŒUX, T., LIANG, Y. et FENG, X. (2015). Belief rule-based classification system : Extension of frbcs in belief functions framework. Information Sciences, 309:26–49. 75
- JIAO, L., PAN, Q., FENG, X. et YANG, F. (2013). An evidential k-nearest neighbor classification method with weighted attributes. In Proceedings of the 16th International Conference on Information Fusion, pages 145–150. IEEE. 51
- JOHN, G. H. (1995). Robust decision trees : Removing outliers from databases. In KDD, volume 95, pages 174–179. 46
- KASSOUF, A., MAALOULY, J., RUTLEDGE, D. N., CHEBIB, H. et DUCRUET, V. (2014). Rapid discrimination of plastic packaging materials using mir spectroscopy coupled with independent components analysis (ica). Waste management, 34(11):2131–2138. 11
- KELLER, J. M., GRAY, M. R. et GIVENS, J. A. (1985). A fuzzy k-nearest neighbor algorithm. IEEE transactions on systems, man, and cybernetics, (4):580–585. 53, 60
- KRENKER, A., BEŠTER, J. et KOS, A. (2011). Introduction to the artificial neural networks. Artificial Neural Networks : Methodological Advances and Biomedical Applications. InTech, pages 1–18. 45
- KUNCHEVA, L. (2000). Fuzzy classifier design, volume 49. Springer Science & Business Media. 25, 53

- LABREUCHE, C. et GRABISCH, M. (2003). The choquet integral for the aggregation of interval scales in multicriteria decision making. Fuzzy Sets and Systems, 137(1):11–26. 37
- LACHAIZE, M., LE HÉGARAT-MASCLE, S., ALDEA, E., MAITROT, A. et REYNAUD, R. (2018). Evidential split-and-merge : Application to object-based image analysis. International Journal of Approximate Reasoning, 103:303–319. 17, 22
- LALLICH, S., MUHLENBACH, F. et ZIGHED, D. A. (2002). Improving classification by removing or relabeling mislabeled instances. In International Symposium on Methodologies for Intelligent Systems, pages 5–15. Springer. 46, 99
- LETTVIN, J. Y., MATURANA, H. R., MCCULLOCH, W. S. et PITTS, W. H. (1959). What the frog’s eye tells the frog’s brain. Proceedings of the IRE, 47(11):1940–1951. 45
- LIAN, C., RUAN, S. et DENŒUX, T. (2015). An evidential classifier based on feature selection and two-step classification strategy. Pattern Recognition, 48(7):2318–2327. 51
- LIU, Z.-g., PAN, Q. et DEZERT, J. (2014). Classification of uncertain and imprecise data based on evidence theory. Neurocomputing, 133:459–470. 74, 86
- LOEFFEL, P.-X., MARSALA, C. et DETYNIĘCKI, M. (2015). Classification with a reject option under concept drift : The droplets algorithm. In 2015 IEEE international conference on data science and advanced analytics (DSAA), pages 1–9. IEEE. 70
- LORENTE, D., ALEIXOS, N., GÓMEZ-SANCHIS, J., CUBERO, S., GARCÍA-NAVARRETE, O. L. et BLASCO, J. (2012). Recent advances and applications of hyperspectral imaging for fruit and vegetable quality assessment. Food and Bioprocess Technology, 5(4):1121–1142. 17
- MACQUEEN, J. et al. (1967). Some methods for classification and analysis of multivariate observations. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, volume 1, pages 281–297. Oakland, CA, USA. 24
- MAMDANI, E. H. (1977). Application of fuzzy logic to approximate reasoning using linguistic synthesis. IEEE transactions on computers, (12):1182–1191. 25
- MASSEN, R. (1998). Method and device for sorting materials. US Patent 5,794,788. 9
- MATARAZZO, A., TUCCIO, G., TEODORO, G., FAILLA, F. et GIUFFRIDA, V. A. (2019). Mass balance as green economic and sustainable management in weee sector. Energy Procedia, 157:1377–1384. 7
- MERCER, J. (1909). Xvi. functions of positive and negative type, and their connection the theory of integral equations. Philosophical transactions of the royal society of London. Series A, containing papers of a mathematical or physical character, 209(441-458):415–446. 45
- MORAL, S., MANTAS, C. J., CASTELLANO, J. G. et ABELLÁN, J. (2020). Imprecise classification with non-parametric predictive inference. In International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, pages 53–66. Springer. 86, 89
- MORTIER, T., WYDMUCH, M., DEMBCZYŃSKI, K., HÜLLERMEIER, E. et WAEGEMAN, W. (2019). Efficient set-valued prediction in multi-class classification. arXiv preprint arXiv :1906.08129. 62, 72, 86, 89, 111, 147

- MUHLENBACH, F., LALLICH, S. et ZIGHED, D. A. (2004). Identifying and handling mislabelled instances. Journal of Intelligent Information Systems, 22(1):89–109. 93
- MUROFUSHI, T. et SUGENO, M. (1989). An interpretation of fuzzy measures and the choquet integral as an integral with respect to a fuzzy measure. Fuzzy sets and Systems, 29(2):201–227. 37
- NAUCK, D., KLAWONN, F. et KRUSE, R. (1997). Foundations of neuro-fuzzy systems. John Wiley & Sons, Inc. 54
- NEYMAN, J. (1937). X-outline of a theory of statistical estimation based on the classical theory of probability. Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences, 236(767):333–380. 62
- NGUYEN, V.-L., DESTERCKE, S., MASSON, M.-H. et HÜLLERMEIER, E. (2018). Reliable multi-class classification based on pairwise epistemic and aleatoric uncertainty. In International Joint Conference on Artificial Intelligence. 74, 75, 86
- NGUYEN, V.-L. et HÜLLERMEIER, E. (2019). Reliable multi-label classification : Prediction with partial abstention. arXiv preprint arXiv :1904.09235. 68
- ONGONDO, F. O., WILLIAMS, I. D. et CHERRETT, T. J. (2011). How are WEEE doing? a global review of the management of electrical and electronic wastes. Waste Management, 31(4):714–730. 8
- PACLIK, P. et DUIN, R. P. (2005). Designing multi-modal classifiers of spectra : a study on industrial sorting application. In 2nd Spectral Imaging workshop, Villach, Austria. 17
- PACLIK, P., LEITNER, R. et DUIN, R. P. (2006). A study on design of object sorting algorithms in the industrial application using hyperspectral imaging. Journal of Real-Time Image Processing, 1(2):101–108. 17
- PAPADOPOULOS, H. (2008). Inductive conformal prediction : Theory and application to neural networks. In Tools in artificial intelligence. Citeseer. 73
- PAZZANI, M., MERZ, C., MURPHY, P., ALI, K., HUME, T. et BRUNK, C. (1994). Reducing misclassification costs. In Machine Learning Proceedings 1994, pages 217–225. Elsevier. 46
- PELLICANÒ, N., LE HÉGARAT-MASCLE, S. et ALDEA, E. (2018). 2cobel : A scalable belief function representation for 2d discernment frames. International Journal of Approximate Reasoning, 103:320–342. 114
- PERRIN, D., MANTAUX, O., IENNY, P., LÉGER, R., DUMON, M. et LOPEZ-CUESTA, J.-M. (2016). Influence of impurities on the performances of hips recycled from waste electric and electronic equipment (WEEE). Waste Management, 56:438–445. 17, 63
- PILLAI, I., FUMERA, G. et ROLI, F. (2013). Multi-label classification with a reject option. Pattern Recognition, 46(8):2256–2266. 68
- PLATT, J. et al. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Advances in large margin classifiers, 10(3):61–74. 50
- QUOST, B., MASSON, M.-H. et DESTERCKE, S. (2020a). Dealing with atypical instances in evidential decision-making. In International Conference on Scalable Uncertainty Management, pages 217–225. Springer. 74

- QUOST, B., MASSON, M.-H. et DESTERCKE, S. (2020b). Traitement évidentiel d'exemples non-conformes en discrimination. In *Vingt-neuvièmes Rencontres Francophones sur la Logique Floue et ses Applications*. 48
- RAMONI, M. et SEBASTIANI, P. (2001). Robust bayes classifiers. *Artificial Intelligence*, 125(1-2):209–226. 25
- RAO, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrics*, 14(1):1–17. 48
- RIZA, L. S., BERGMEIR, C. N., HERRERA TRIGUERO, F., BENÍTEZ SÁNCHEZ, J. M. et al. (2015). frbs : Fuzzy rule-based systems for classification and regression in r. American Statistical Association. 54, 60
- ROH, S.-B., OH, S.-K., PARK, E.-K. et CHOI, W. Z. (2017). Identification of black plastics realized with the aid of raman spectroscopy and fuzzy radial basis function neural networks classifier. *Journal of Material Cycles and Waste Management*, 19(3):1093–1105. 17, 22
- ROKACH, L. et MAIMON, O. (2005). Clustering methods. In *Data mining and knowledge discovery handbook*, pages 321–352. Springer. 24
- ROUBOS, J. A., SETNES, M. et ABONYI, J. (2003). Learning fuzzy classification rules from labeled data. *Information Sciences*, 150(1-2):77–93. 54
- ROZENSTEIN, O., PUCKRIN, E. et ADAMOWSKI, J. (2017). Development of a new approach based on midwave infrared spectroscopy for post-consumer black plastic waste sorting in the recycling industry. *Waste Management*, 68:38–44. 11
- SHAFFER, G. (1976). *A mathematical theory of evidence*, volume 42. Princeton university press. 25, 32, 51
- SHAFFER, G. (1981). Constructive probability. *Synthese*, 48(1):1–60. 74
- SHAFFER, G. et VOVK, V. (2008). A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9(Mar):371–421. 72
- SHILANE, P., MIN, P., KAZHDAN, M. et FUNKHOUSER, T. (2004). The princeton shape benchmark. In *Proceedings Shape Modeling Applications, 2004.*, pages 167–178. IEEE. 82
- SIGNORET, C. (2019). *Valorisation de Matières Premières Secondaires (MPS) thermoplastiques en mélange issues de tri spectroscopique en ligne - Projet Mélanie*. Thèse de doctorat. Thèse de doctorat dirigée par Perrin, Didier et Ienny, Patrick Chimie et Physico-Chimie des Matériaux IMT Mines Alès 2019. 9
- SIGNORET, C., CARO-BRETELLE, A.-S., LOPEZ-CUESTA, J.-M., IENNY, P. et PERRIN, D. (2019a). MIR spectral characterization of plastic to enable discrimination in an industrial recycling context : I. specific case of styrenic polymers. *Waste Management*, 95:513–525. 14
- SIGNORET, C., CARO-BRETELLE, A.-S., LOPEZ-CUESTA, J.-M., IENNY, P. et PERRIN, D. (2019b). MIR spectral characterization of plastic to enable discrimination in an industrial recycling context : II. specific case of polyolefins. *Waste Management*, 98:160–172. 14
- SIGNORET, C., CARO-BRETELLE, A.-S., LOPEZ-CUESTA, J.-M., IENNY, P. et PERRIN, D. (2019c). Mir spectral characterization of plastic to enable discrimination in an industrial recycling context : Ii. specific case of polyolefins. *Waste Management*, 98:160–172. 8, 126

- SIGNORET, C., CARO-BRETELLE, A.-S., LOPEZ-CUESTA, J.-M., IENNY, P. et PERRIN, D. (2020). Alterations of plastics spectra in mir and the potential impacts on identification towards recycling. Resources, Conservation and Recycling, 161:104980. 11
- SILLA, C. N. et FREITAS, A. A. (2011). A survey of hierarchical classification across different application domains. Data Mining and Knowledge Discovery, 22(1-2):31–72. 69, 82, 83
- SMETS, P. (1988). Non-Standard Logics for Automated Reasoning. Academic Press. 32, 33
- SMETS, P. (1990). The combination of evidence in the transferable belief model. IEEE Transactions on pattern analysis and machine intelligence, 12(5):447–458. 36
- SMETS, P. et KENNES, R. (1994). The transferable belief model. Artificial intelligence, 66(2):191–234. 25, 32
- SUN, Y., WONG, A. K. et KAMEL, M. S. (2009). Classification of imbalanced data : A review. International journal of pattern recognition and artificial intelligence, 23(04):687–719. 46
- SUTTON-CHARANI, N., DESTERCCKE, S. et DENOËUX, T. (2012). Classification trees based on belief functions. In Belief Functions : Theory and Applications, pages 77–84. Springer. 51
- SUTTON-CHARANI, N., DESTERCCKE, S. et DENOËUX, T. (2013). Learning decision trees from uncertain data with an evidential em approach. In 2013 12th International Conference on Machine Learning and Applications, volume 1, pages 111–116. IEEE. 51
- TEDJINI, T., PICHON, F. et LEFEVRE, E. (2020). Vers une généralisation de l’approximation des fonctions de croyance. In LFA’2020-29èmes Rencontres Francophones sur la Logique Floue et ses Applications. 143, 147
- TSOUMAKAS, G. et VLAHAVAS, I. (2007). Random k-labelsets : An ensemble method for multilabel classification. In European conference on machine learning, pages 406–417. Springer. 77
- TUCKER, L. R. (1958). Determination of parameters of a functional relation by factor analysis. Psychometrika, 23(1):19–23. 48
- TURING, A. et HAUGELAND, J. (1950). Computing machinery and intelligence. MIT Press Cambridge, MA. 1
- UE (2012). Article 11 recovery targets, directive 2012/19/eu of the european parliament and of the council of 4 july 2012 on waste electrical and electronic equipment (WEEE). Official Journal of the European Union. 7
- VANNOORENBERGHE, P. et DENOËUX, T. (2001). Likelihood-based vs. distance-based evidential classifiers. In 10th IEEE International Conference on Fuzzy Systems.(Cat. No. 01CH37297), volume 1, pages 320–323. IEEE. 51
- VENABLES, W. N. et RIPLEY, B. D. (2013). Modern applied statistics with S-PLUS. Springer Science & Business Media. 46
- WALLEY, P. (1991). Statistical Reasoning with Imprecise Probabilities. Chapman & Hall. 25
- WALLEY, P. et PETER, W. (1991). Statistical Reasoning with Imprecise Probabilities. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis. 30, 31, 73, 106

- WANG, L.-X. et MENDEL, J. M. (1992). Generating fuzzy rules by learning from examples. IEEE Transactions on systems, man, and cybernetics, 22(6):1414–1427. 39
- YAGER, R. R. (1987). On the dempster-shafer framework and new combination rules. Information sciences, 41(2):93–137. 36
- YAGER, R. R. (1988). On ordered weighted averaging aggregation operators in multicriteria decision making. IEEE Transactions on systems, Man, and Cybernetics, 18(1):183–190. 53
- YANG, G. (2016). Cautious models in supervised machine learning. Theses, Université de Technologie de Compiègne. 63, 77
- YANG, G., DESTERCKE, S. et MASSON, M.-H. (2016). The costs of indeterminacy : How to determine them ? IEEE transactions on cybernetics, 47(12):4316–4327. 62, 64, 76, 84
- YUAN, Y. et SHAW, M. J. (1995). Induction of fuzzy decision trees. Fuzzy Sets and systems, 69(2):125–139. 53
- ZADEH, L. A. (1965). Fuzzy sets. Information and control, 8(3):338–353. 25
- ZADROZNY, B. et ELKAN, C. (2002). Transforming classifier scores into accurate multiclass probability estimates. In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 694–699. 50
- ZAFFALON, M. (1999). A credal approach to naive classification. In ISIPTA, volume 99, pages 405–414. 25, 73, 76, 86, 89, 157
- ZAFFALON, M. (2001). Statistical inference of the naive credal classifier. In ISIPTA, volume 1, pages 384–393. 73, 106, 157, 158
- ZAFFALON, M., CORANI, G. et MAUÁ, D. (2012). Evaluating credal classifiers by utility-discounted predictive accuracy. International Journal of Approximate Reasoning, 53(8):1282–1301. xiv, 72, 77, 80, 81, 106, 107
- ZAFFALON, M., WESNES, K. et PETRINI, O. (2003). Reliable diagnoses of dementia by the naive credal classifier inferred from incomplete cognitive data. Artificial intelligence in medicine, 29(1-2):61–79. 63
- ZHANG, C., WANG, W. et QIAO, X. (2017). On reject and refine options in multicategory classification. arXiv preprint arXiv :1701.02265. 75, 86
- ZHANG, J., SUBASINGHA, S., PREMARATNE, K., SHYU, M.-L., KUBAT, M. et HEWAWASAM, K. (2004). A novel belief theoretic association rule mining based classifier for handling class label ambiguities. In Proc. Workshop Foundations of Data Mining (FDM'04), Int'l Conf. Data Mining (ICDM'04). 99
- ZIMMERMANN, H.-J. (2011). Fuzzy set theory—and its applications. Springer Science & Business Media. 25
- ZOUHAL, L. M. et DENOEU, T. (1998). An evidence-theoretic k-nn rule with parameter optimization. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 28(2):263–271. 74
- ZOUHAL, L. M. et DENOEU, T. (1998). An evidence-theoretic k-nn rule with parameter optimization. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 28(2):263–271. 99