



HAL
open science

Conception d'un système de diagnostic adaptatif et prédictif basé sur la méthode Fuzzy Pattern Matching pour la surveillance en ligne des systèmes évolutifs

Moamar Sayed Mouchaweh

► **To cite this version:**

Moamar Sayed Mouchaweh. Conception d'un système de diagnostic adaptatif et prédictif basé sur la méthode Fuzzy Pattern Matching pour la surveillance en ligne des systèmes évolutifs. Autre [cs.OH]. Université de Reims - Champagne Ardenne, 2002. Français. NNT : . tel-00002637

HAL Id: tel-00002637

<https://theses.hal.science/tel-00002637>

Submitted on 27 Mar 2003

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.

UNIVERSITE DE REIMS CHAMPAGNE-ARDENNE

THESE

présentée à

L'UFR SCIENCES EXACTES ET NATURELLES

pour l'obtention du titre de

**DOCTEUR DE L'UNIVERSITE DE
REIMS CHAMPAGNE-ARDENNE**

en

**GENIE INFORMATIQUE, AUTOMATIQUE
ET TRAITEMENT DU SIGNAL**

par

Moamar SAYED MOUCHAWEH

**Conception d'un système de diagnostic adaptatif et prédictif
basé sur la méthode Fuzzy Pattern Matching pour la surveillance
en ligne des systèmes évolutifs
Application à la supervision et au diagnostic d'une ligne de
peinture au trempé**

Soutenue le 11 décembre 2002 devant le jury composé de :

Sylviane GENTIL

Professeur à l'Institut Nationale Polytechnique de
Grenoble (Rapporteur)

Patrick MILLOT

Professeur à l'Université de Valenciennes (Rapporteur)

Didier MAQUIN

Maître de Conférence HDR à l'Institut National
Polytechnique de Lorraine

Janan ZAYTOON

Professeur à l'Université de Reims Champagne-Ardenne

Patrice BILLAUDEL

Professeur à l'Université de Reims Champagne-Ardenne
(directeur de thèse)

Remerciements

Ce travail de thèse à été réalisé au sein du groupe Automatique du Laboratoire d'Automatique et de Micro-électronique dirigé par le Professeur Janan ZAYTOON à l'Université de Reims Champagne Ardenne. Je le remercie pour m'avoir accueilli dans son laboratoire.

J'adresse mes remerciements sincères à M. Patrice BILLAUDEL Professeur à l'Université de Reims Champagne Ardenne pour avoir accepté de m'encadrer pendant les trois ans de thèse. Ses grandes compétences et appréciables conseils m'ont guidé de manière forte tout au long de mes travaux de thèse.

Je tiens à adresser tous mes remerciements à M. Patrick MILLOT Professeur à l'Université de Valenciennes et à Mme. Sylviane GENTIL Professeur à l'ENSIEG/INPG de Grenoble pour avoir accepté d'être les rapporteurs de cette thèse, je les en remercie vivement.

Je remercie sincèrement M. Didier MAQUIN Maître de Conférence HDR à l'Institut National Polytechnique de Lorraine d'avoir accepté d'être examinateur dans le jury.

Je tiens à témoigner ma reconnaissance à M. Ivon GARDAN Professeur à l'Université de Reims Champagne Ardenne, Directeur de l'Institut de Formation Technique Supérieure pour le soutien et l'encouragement qu'il m'a donné.

Je remercie toute l'équipe de l'Institut de Formation Technique Supérieure pour leur gentillesse et soutien, ce dont je leur suis reconnaissant.

Table des matières

INTRODUCTION GENERALE.....	9
1. DIAGNOSTIC PAR RECONNAISSANCE FLOUE DES FORMES	12
1.1. DIAGNOSTIC ET SUPERVISION	12
1.2. METHODES DE DIAGNOSTIC	13
1.3. RECONNAISSANCE DES FORMES	14
1.3.1. Phase d'analyse	15
1.3.1.1. Détermination de l'espace de représentation,	15
1.3.1.2. Réduction de l'espace de représentation.....	15
1.3.1.3. Détermination de l'espace de décision	16
1.3.2. Phase de choix d'une méthode de décision.....	16
1.3.3. Phase d'exploitation	17
1.4. NOTION DE REJET	18
1.5. RECONNAISSANCE FLOUE DES FORMES	19
1.5.1. Principe.....	19
1.5.2. Application de la logique floue en décision	20
1.5.2.1. Règle floue avec rejet en ambiguïté.....	21
1.5.2.2. Règle floue avec rejet en appartenance.....	22
1.5.3. Méthodes d'apprentissage flou non-supervisé.....	23
1.5.4. Méthodes floues de classification supervisée.....	27
1.6. CONCLUSION	29
2. THEORIE DES POSSIBILITES	31
2.1. INTRODUCTION	31
2.2. LA THEORIE DES ENSEMBLES FLOUS	31
2.3. MESURES DE CONFIANCE.....	32
2.3.1. Théorie des probabilités	33
2.3.2. Théorie des possibilités.....	34
2.3.3. Théorie des croyances de Dempster-Shafer.....	36
2.3.4. Choix de la théorie de représentation de nos données.....	38
2.4. TRANSFORMATIONS PROBABILITE-POSSIBILITE	38
2.4.1. Principe de cohérence de Zadeh.....	39
2.4.2. Principe de cohérence de Klir	40
2.4.3. Principe de cohérence de Dubois et Prade.....	41
2.4.4. Transformation Variable (TV)	42
2.5. COMPARAISON ENTRE LES DIFFERENTES TRANSFORMATIONS.....	45
2.5.1. Critère de normalisation.....	45

2.5.2. Critère de maximum de spécificité.....	46
2.5.3. Critère de conservation de l'incertitude	52
2.5.4. Critère de conservation de la préférence.....	54
2.5.5. Critère de conservation de la forme	55
2.5.6. Critère de conservation de l'ignorance	56
2.6. CONCLUSION	57
3. RECONNAISSANCE POSSIBILISTE DES FORMES	59
3.1. INTRODUCTION	59
3.2. METHODE POSSIBILISTE DES C MOYENNES	59
3.3. FUZZY PATTERN MATCHING (FPM)	61
3.3.1. Phase d'apprentissage	62
3.3.1.1. Construction des histogrammes de données	62
3.3.1.2. Histogrammes de probabilité.....	63
3.3.1.3. Histogrammes de possibilité.....	63
3.3.1.4. Densités de possibilité	63
3.3.2. Phase de classification.....	64
3.4. OPERATEUR D'AGREGATION.....	64
3.5. AGREGATION ET FPM	65
3.6. PRESENTATION DES DONNEES TESTS.....	66
3.6.1. Données Matières Plastiques.....	66
3.6.2. Données Codes Métalliques.....	67
3.6.3. Données Machine à Laver	68
3.6.4. Données Injections Plastiques	68
3.6.5. Données Iris.....	69
3.7. APPRENTISSAGE CONTINU	69
3.7.1. Apprentissage continu dans FPM.....	70
3.7.2. Apprentissage incrémental dans FPM.....	73
3.7.3. Résultats expérimentaux	74
3.7.3.1. Temps de mise à jour et taille de mémoire	74
3.7.3.2. Nombre de points rejetés en appartenance.....	80
3.8. ADAPTATION DE FPM POUR DES CLASSES DE FORME COMPLEXE	84
3.8.1. Application de FPM aux classes de forme simple.....	84
3.8.2. Application de FPM aux classes de forme complexe	85
3.8.3. Fuzzy Pattern Matching Multiprofil (FPMM)	88
3.8.3.1. Décomposition des classes en sous-classes et calcul des densités de possibilité	88
3.8.3.2. Utilisation des opérateurs d'agrégation	89
3.8.3.3. Fusion des possibilités d'appartenance.....	89
3.8.3.4. Performances de l'algorithme	90
3.8.3.5. Apprentissage incrémental pour la méthode FPMM.....	92
3.8.4. Fuzzy Pattern Matching avec une fonction Exponentielle (FPME).....	94
3.8.5. Fuzzy Pattern Matching Corrélée (FPMC)	95

3.8.5.1. Phase d'apprentissage.....	96
3.8.5.2. Phase de classification.....	97
3.8.6. Performances de FPMC.....	100
3.9. CONCLUSION.....	101
4. OPTIMISATION DES PARAMETRES DE LA METHODE FPM.....	103
4.1. INTRODUCTION.....	103
4.2. PARAMETRES D'UN HISTOGRAMME.....	103
4.2.1. Nombre de barres.....	103
4.2.2. Bornes inférieure et supérieure.....	106
4.3. ESTIMATION DES PERFORMANCES DE FPM.....	106
4.4. MESURES DE SEPARABILITE DES CLASSES.....	108
4.5. MESURE DE SEPARABILITE ADAPTEE A FPM.....	110
4.6. ZONES DE REJET DANS LES CLASSES.....	115
4.7. EXEMPLES.....	117
4.7.1. Données Matières Plastiques.....	117
4.7.2. Données Machine à Laver.....	118
4.7.3. Données Codes Métalliques.....	119
4.7.4. Données Injection Plastique.....	120
4.7.5. Données Iris.....	121
4.8. INFLUENCE DES BORNES DES HISTOGRAMMES.....	122
4.9. DETERMINATION DES PARAMETRES DES HISTOGRAMMES.....	124
4.10. METHODE INCREMENTALE POUR LA MISE A JOUR DE LA MESURE DE SEPARABILITE.....	125
4.11. AMELIORATION DE LA MESURE DE SEPARABILITE.....	128
4.11.1. FPM avec des barres de largeur Adaptative (FPMA).....	128
4.11.2. Résultats expérimentaux.....	129
4.11.2.1. Données Matières Plastiques.....	129
4.11.2.2. Données Machine à Laver.....	130
4.11.2.3. Données Codes Métalliques.....	131
4.11.2.4. Données injection plastique.....	131
4.11.2.5. Données Iris.....	132
4.12. METHODE INCREMENTALE POUR LA MISE A JOUR DE LA MESURE DE SEPARABILITE POUR FPMA....	132
4.13. CONCLUSION.....	134
5. ELIMINATION DU BRUIT ET DETECTION DES ZONES DE HAUTE DENSITE.....	136
5.1. INTRODUCTION.....	136
5.2. PARAMETRE K ET DETECTION DES ZONES DE HAUTE DENSITE.....	136
5.2.1. Nécessité d'un seuil.....	136
5.2.2. Rôle du paramètre k dans la Transformation Variable.....	137
5.2.3. Choix du seuil de probabilité pour déterminer k.....	138
5.2.4. Relation entre spécificité et seuil.....	138
5.3. DETERMINATION AUTOMATIQUE DE K.....	140

5.4. EXEMPLES	141
5.4.1. <i>Elimination du bruit</i>	141
5.4.2. <i>Détection des zones de haute densité</i>	143
5.4.2.1. Cas d'une seule zone de haute densité.....	143
5.4.2.2. Cas de plusieurs zones de haute densité.....	146
5.5. APPLICATION INDUSTRIELLE	148
5.5.1. <i>Fonctionnement de la ligne de peinture au trempé</i>	149
5.5.2. <i>Paramètres de contrôle de fonctionnement de la ligne</i>	150
5.5.3. <i>Qualité du produit</i>	151
5.5.4. <i>Paramètres de qualité des produits peints</i>	152
5.5.4.1. Viscosité	153
5.5.4.2. pH.....	153
5.5.4.3. Température de la peinture :	153
5.5.5. <i>Construction du module de diagnostic de qualité des produits peints</i>	153
5.5.5.1. Détermination et réduction de l'espace de représentation.....	153
5.5.5.2. Détermination de l'espace de décision	154
5.5.5.3. Application de FPM avec la TV	155
5.6. SEUIL ADAPTATIF DE LA TV	156
5.7. CHOIX DE <i>H</i>	158
5.8. CONCLUSION	161
6. DIAGNOSTIC ADAPTATIF ET PREDICTIF EN UTILISANT FPM.....	162
6.1. INTRODUCTION	162
6.2. DIAGNOSTIC PREDICTIF	162
6.3. DIAGNOSTIC ADAPTATIF.....	164
6.4. INTEGRATION DE FPM DANS UN MODULE DE DIAGNOSTIC PREDICTIF ET ADAPTATIF	165
6.4.1. <i>Prédiction de l'état futur</i>	166
6.4.1.1. Fonction d'évolution.....	166
6.4.1.2. Matrice et vecteur d'évolution.....	169
6.4.1.3. Vitesse d'évolution.....	171
6.4.1.4. Points de bruit dans la séquence d'évolution	173
6.4.2. <i>Détection de nouveaux modes de fonctionnement</i>	174
6.4.2.1. Détection d'un seul mode de fonctionnement.....	174
6.4.2.2. Détection de deux modes de fonctionnement	176
6.4.3. <i>Architecture d'un module de diagnostic prédictif et adaptatif en ligne</i>	177
6.5. EXPERIMENTATION ET VALIDATION.....	178
6.5.1. <i>Exemple de simulation</i>	178
6.5.1.1. Cas d'évolution.....	178
6.5.1.2. Cas d'évolution accompagnée de bruit.....	179
6.5.1.3. Cas d'apparition d'une nouvelle classe.....	180
6.5.1.4. Cas d'apparition de deux nouvelles classes	181
6.5.1.5. Cas d'évolution vers une nouvelle classe	182
6.5.2. <i>Exemple des données Matières Plastiques</i>	182

6.5.2.1. Cas d'apparition de deux nouvelles classes	185
6.5.2.2. Cas des points aberrants	186
6.5.2.3. Cas d'évolution linéaire.....	187
6.5.2.4. Cas d'évolution non linéaire.....	188
6.5.3. <i>Exemple des données Machine à Laver</i>	189
6.5.3.1. Cas d'apparition d'une nouvelle classe.....	189
6.5.3.2. Cas d'apparition de deux nouvelles classes	190
6.5.4. <i>Exemple des données Injection Plastique</i>	191
6.5.4.1. Cas d'apparition d'une nouvelle classe.....	191
6.5.4.2. Cas d'apparition de deux nouvelles classes	192
6.5.4.3. Cas d'apparition de quatre nouvelles classes.....	193
6.5.5. <i>Exemple des données Iris</i>	194
6.5.5.1. Cas d'apparition d'une seule classe	194
6.5.5.2. Cas d'apparition de deux classes	194
6.6. CONCLUSION	195
CONCLUSION GENERALE.....	197
BIBLIOGRAPHIE	199

Introduction générale

Ce travail a été réalisé au sein du groupe Automatique du Laboratoire d'Automatique et de Microélectronique de l'Université de Reims Champagne Ardenne. Les objectifs de notre groupe sont la mise au point de modules de diagnostic et de supervision de processus industriels en utilisant les techniques de reconnaissance floue des formes.

La supervision des processus industriels permet d'accroître la productivité et de diminuer le coût d'entretien par l'amélioration de la disponibilité des systèmes de production. Ces processus sont susceptibles d'évoluer sous deux types de modes de fonctionnement : des modes normaux et des modes anormaux. Un mode de fonctionnement anormal correspond à un dysfonctionnement du système entraînant une réalisation partiellement ou complètement non conforme de la tâche pour laquelle ces processus sont conçus.

Le diagnostic est une composante principale d'un module de supervision. Il consiste à déterminer à chaque instant le mode de fonctionnement dans lequel le système se trouve. Il s'appuie sur une connaissance a priori des modes de fonctionnement et sur une connaissance instantanée matérialisée par une nouvelle observation de l'état du système. Il existe plusieurs approches pour réaliser le diagnostic, le choix d'une approche est lié au mode de représentation de la connaissance. Lorsque cette connaissance est de type numérique, issue de capteurs, et lorsqu'on ne peut pas construire un modèle analytique du système, l'approche par Reconnaissance des Formes (RdF) est particulièrement adaptée.

La RdF se base sur la classification des formes en les comparant à des formes types. Une forme est l'observation du fonctionnement du processus, représentée par un point dans un espace à a paramètres appelé l'espace de représentation. Les formes types sont les modes de fonctionnement représentés par des ensembles de points occupant des zones restreintes de l'espace de représentation appelées classes.

L'application de la RdF se heurte souvent aux problèmes suivants :

- dans une base de connaissance incomplète, tous les modes de fonctionnement ne sont pas représentés. En conséquence, un module de diagnostic doit être adaptatif afin d'inclure à sa base de connaissance les nouveaux modes dès qu'ils apparaissent,
- lorsque le système évolue vers un mode anormal, il est nécessaire d'anticiper cette évolution plutôt que d'attendre d'arriver à ce mode afin d'éviter ses conséquences surtout s'il est dangereux. En conséquence le module de diagnostic doit être prédictif,
- dans le cas d'un système évolutif, la base de connaissance doit être enrichie grâce à

l'information apportée par les nouvelles observations. Cet enrichissement doit être réalisé en temps réel.

Le diagnostic par RdF est réalisé en utilisant une méthode de classification qui associe une nouvelle observation à une classe correspondant à un mode de fonctionnement. Nous nous sommes intéressés aux méthodes de classification non-paramétrique, plus adaptées aux cas industriels, pour lesquelles la densité de probabilité est complètement inconnue. Nous avons choisi la méthode de classification Fuzzy Pattern Matching (FPM) pour les raisons suivantes :

- elle est simple à appliquer,
- le temps de calcul pour la classification d'une nouvelle observation est faible et indépendant de la taille de l'ensemble d'apprentissage,
- elle est capable de traiter des données qui sont à la fois incertaines et imprécises.

Cependant cette méthode, de part son principe, n'est ni prédictive ni adaptative et nécessite la détermination, par l'utilisateur, de plusieurs paramètres qui influencent fortement ses performances. De plus, ses performances se dégradent dans le cas du bruit représenté par des points aberrants et dans le cas des classes de forme non sphérique. L'objectif principal de cette thèse consiste à remédier à ces inconvénients afin que FPM soit apte à être utilisée dans un module de diagnostic en temps réel adaptatif et prédictif pour des systèmes évolutifs.

Le mémoire est découpé en 6 chapitres. Dans le premier, nous présentons les différentes approches utilisées pour la conception d'un module de diagnostic, puis les approches utilisant la RdF et leurs limites.

Le deuxième chapitre est consacré à la présentation de la théorie des possibilités et des transformations probabilité-possibilité les plus courantes. Les méthodes utilisant cette théorie sont les plus adaptées à des cas où l'on dispose de peu d'information, celle-ci étant à la fois incertaine et imprécise. Nous proposons une transformation probabilité-possibilité qui possède des caractéristiques intéressantes pour améliorer les performances de FPM.

Le troisième chapitre présente deux méthodes de RdF utilisant la théorie des possibilités : Fuzzy C Means et FPM. Nous développons cette dernière méthode pour qu'elle soit opérante dans le cas de données provenant de systèmes évolutifs ou présentant des classes non sphériques.

Dans le quatrième chapitre, nous proposons une approche automatique pour optimiser les paramètres de la méthode FPM afin d'obtenir les meilleures performances.

Le cinquième chapitre est consacré à l'utilisation de notre transformation probabilité-

possibilité pour détecter les sous-classes, représentées par des zones de haute-densité, et pour éliminer les points aberrants. Nous appliquons FPM en utilisant cette transformation sur une ligne de peinture pour maîtriser la qualité des produits.

Dans le dernier chapitre, nous proposons une approche pour que la méthode FPM soit apte à être intégrée dans un module de diagnostic en temps réel adaptatif et évolutif.

1. Diagnostic par reconnaissance floue des formes

1.1. Diagnostic et supervision

Un système physique, ou un processus technologique, est conçu pour réaliser une tâche. Cette tâche peut être totalement ou partiellement remplie avec un certain niveau de qualité. Ce système, ou processus, peut évoluer sous différents modes de fonctionnement, parmi lesquels on peut citer :

- les modes normaux sous lesquels la mission est totalement remplie, l'un d'eux est le mode nominal correspondant à une qualité parfaite de la mission,
- les modes anormaux sous lesquels la mission est partiellement ou totalement non remplie, incluant les modes défaillants, les modes interdits, ...,
- les modes évolutifs qui caractérisent des modes, souvent transitoires, de passage d'un mode de fonctionnement à un autre.

Le diagnostic d'un système est l'identification du mode de fonctionnement, à chaque instant, par ses manifestations extérieures. Son principe général consiste à confronter les données relevées au cours du fonctionnement réel du système avec la connaissance que l'on a de son fonctionnement normal ou défaillant. Si le mode de fonctionnement identifié est un mode défaillant, le système de diagnostic peut localiser sa cause.

La supervision est définie comme étant la surveillance et le contrôle de l'exécution d'une opération ou d'un travail accompli par un homme ou par une machine. En présence de défaillance, la supervision prendra toutes les démarches nécessaires pour le retour du système vers un mode de fonctionnement normal. Un système de supervision traitant l'apparition d'un défaut peut être défini par les opérations suivantes :

- l'acquisition de mesures ou de compte rendus permettant de vérifier que les consignes envoyées vers le processus produisent exactement les effets attendus, cela est réalisé par des capteurs,
- la perception qui fournit des données pré-traitées par des techniques de traitement du signal, extraction de caractéristiques ...,
- le diagnostic qui consiste à déterminer dans quel état de fonctionnement se trouve le

système. Il établit un lien de cause à effet entre un symptôme observé et la défaillance qui est survenue,

- éventuellement, le diagnostic prédictif qui permet une intervention préventive sur le système pour éviter qu'un mode non désiré ne se produise,
- enfin, la commande d'une séquence d'actions correctives destinée à rendre au système son état de fonctionnement normal.

La Figure 1.1 montre les différentes composantes d'un système de supervision.

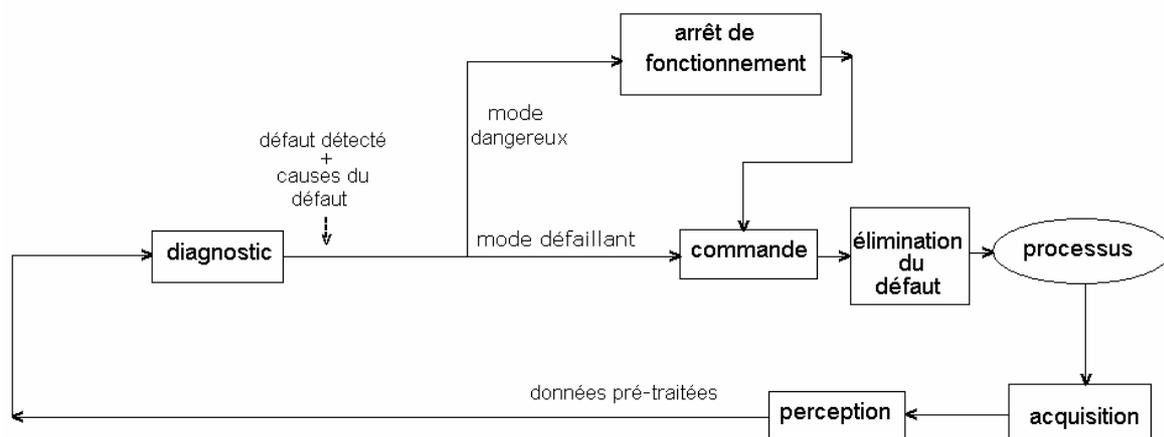


Figure 1.1. Système de supervision traitant l'apparition d'un défaut.

Le diagnostic et la supervision continus d'un système industriel permettent d'améliorer la productivité par l'augmentation de la disponibilité.

1.2. Méthodes de diagnostic

Effectuer un diagnostic consiste à comparer l'information instantanée issue du système à la connaissance a priori disponible sur ce système. Les méthodes de diagnostic varient selon le type de représentation des modes de fonctionnement, elles peuvent être classées comme suit :

- les méthodes basées sur des modèles analytiques ou mathématiques,
- les méthodes symboliques ou les systèmes experts,
- les méthodes basées sur la Reconnaissance des Formes (RdF).

La connaissance a priori et la connaissance instantanée des modes de fonctionnement sont matérialisées par un ensemble d'observations. Le type de représentation de ces observations détermine le choix de la méthode de diagnostic.

Lorsque les observations sont de type symbolique, des effets, les systèmes experts s'imposent. Chaque système expert contient une base de connaissance, une base de faits et de règles, et un moteur d'inférence. Ce dernier choisit les règles en fonction des faits observés.

Quand les observations sont de type numérique, issues de capteurs, et que l'on dispose d'un modèle mathématique du système, les méthodes basées sur des modèles analytiques sont privilégiées. En revanche si les observations sont de type numérique et qu'on ne peut pas construire de modèle mathématique correspondant aux divers modes de fonctionnement du système ou si sa complexité le rend inexploitable, les méthodes basées sur la RdF sont les plus adaptées. La section suivante montre comment l'approche par RdF sert à la conception d'un système de diagnostic et fait ressortir l'avantage essentiel qu'elle présente par rapport aux autres méthodes, à savoir : procurer au système de diagnostic les moyens d'appréhender la connaissance a priori incomplète des modes de fonctionnement.

1.3. Reconnaissance des formes

Le diagnostic par RdF peut être présenté comme une solution alternative à l'approche par modèle puisque les modes de fonctionnement sont modélisés, pas de manière analytique, mais en utilisant uniquement un ensemble des mesures de ces modes [BOU96].

La RdF est la science qui se base sur la définition d'algorithmes permettant de classer des objets ou des formes en les comparant à des formes types [BOU97]. Ses applications interviennent dans de nombreux domaines tels que la reconnaissance vocale, la reconnaissance de caractères, l'automatisation industrielle, le diagnostic médical, et la classification de documents. De manière générale, on distingue les types de RdF suivants :

- la RdF structurelle qui se base sur une représentation des formes à l'aide de grammaires,
 - la RdF statistique qui s'appuie sur une représentation purement numérique des formes.
- C'est cette approche qui sera développée dans la suite.

Pour la RdF statistique, une forme est un ensemble de a paramètres ou caractères. Quand les paramètres sont des nombres réels, une forme peut être définie par un point ou un vecteur $\underline{x} = (x^1, x^2, \dots, x^a)$ dans un espace de dimension a défini par ces paramètres. Cet espace est appelé l'espace de représentation.

Les prototypes sont des points représentatifs des formes types dans l'espace de représentation, dans notre cas les formes types sont des modes de fonctionnement. Puisque

dans un cas réel, les points sont bruités, une nouvelle observation sera rarement rigoureusement confondue avec l'un des prototypes. Les nouvelles observations forment donc des zones autour de chaque prototype, ces zones sont appelées classes. Le problème de la RdF est de savoir décider à quelle classe, parmi c classes (w_1, w_2, \dots, w_c) connues, associer une nouvelle forme \underline{x} observée.

Le fonctionnement d'un système de diagnostic par RdF se déroule en trois phases [DUB90] : une phase d'analyse, une phase de choix d'une méthode de décision et une phase d'exploitation.

1.3.1. Phase d'analyse

1.3.1.1. Détermination de l'espace de représentation,

Il s'agit de construire le vecteur forme \underline{x} [BOU96]. Elle consiste à déterminer les paramètres et les techniques de pré-traitement nécessaires pour permettre une meilleure reconnaissance ou classification. Elle est le fruit de l'expérience acquise sur le processus et conditionne le succès du système de diagnostic par RdF. Un ensemble de N observations : $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_N$, recueillies sur le processus à l'aide des capteurs, constitue l'ensemble d'apprentissage.

1.3.1.2. Réduction de l'espace de représentation

Il s'agit de réduire la dimension de l'espace de représentation afin d'une part de diminuer le temps de classification d'une nouvelle observation, et d'autre part d'éviter la dégradation des performances à cause d'une dimension trop élevée de l'espace de représentation [MIL93]. Cette opération de réduction peut être effectuée par des méthodes d'extraction ou de sélection de paramètres. L'extraction consiste à combiner les paramètres initiaux, non ou peu corrélés, pour aboutir à un nombre inférieur de nouveaux paramètres. Si la corrélation entre ces paramètres est linéaire, on peut utiliser des méthodes d'extraction linéaire telle que l'Analyse en Composantes Principales (ACP). En revanche si cette corrélation est non linéaire, il faut utiliser des méthodes d'extraction non-linéaire telle que le réseau Vector Quantization and Projection (VQP) [DEM93]. La sélection consiste à ne conserver que les paramètres les plus pertinents, ou les plus informatifs, au sens d'un critère de séparabilité entre les classes et d'un critère de condensation des points d'une même classe, comme par exemple le rapport de Fisher [FIS36] et son extension [DOC81]. Ce dernier est efficace seulement dans le cas d'une structure sphérique des classes.

1.3.1.3. Détermination de l'espace de décision

Il s'agit de l'ensemble de toutes les classes possibles. Si on connaît la classe d'origine de chaque observation de l'ensemble d'apprentissage, l'espace de décision est parfaitement connu et l'apprentissage des classes peut être fait en mode dit supervisé. Dans le cas contraire, l'apprentissage des classes est fait en mode non supervisé et il passe par deux étapes : une phase de partitionnement automatique de l'ensemble des formes en classes et une phase de validation de ces classes soit par un expert, soit par des critères de validité de la partition. Il existe de nombreuses méthodes de classification non supervisée nommées méthodes de coalescence, par exemple : les méthodes hiérarchiques [LEB95] et les méthodes des nuées dynamiques dont l'algorithme des Centres Mobiles [BAL67, DID71]. Le principe général de ces méthodes est de détecter dans l'ensemble d'apprentissage une structure de classes disjointes au sens d'un critère de similarité, de telle façon que les observations d'une même classe soient semblables et que celles de classes différentes soient dissemblables.

1.3.2. Phase de choix d'une méthode de décision

Les méthodes de décision définissent une règle de décision pour la classification des nouvelles observations aux différentes classes de l'ensemble d'apprentissage. La performance de ces méthodes est estimée en utilisant un indice de performance, en général on choisit la probabilité de mauvaise classification.

La règle de décision peut être construite en utilisant une approche statistique ou une approche analytique. Dans le premier cas, on associe à chaque vecteur forme une densité de probabilité conditionnelle $f(\underline{x} | w_j)$ par rapport à chaque classe w_j . Dans le cas paramétrique, et si la connaissance est complète, la loi de cette densité est totalement connue, y compris ses paramètres. Si la connaissance sur $f(\underline{x} | w_j)$ est incomplète mais si sa forme est connue, on estime ses paramètres.

Chaque classe w_j est caractérisée par une probabilité a priori $p(w_j)$, ces probabilités vérifient la condition d'orthogonalité définie par :

$$\sum_{j=1}^c p(w_j) = 1 \quad (1.1)$$

La règle de décision la plus courante est la règle de Bayes définie par [DUB90] :

$$\underline{x} \text{ est classée dans } w_j \text{ si } p(w_j | \underline{x}) = \max_{i=1..c} p(w_i | \underline{x}) \quad (1.2)$$

où $p(w_j | \underline{x})$ désigne la probabilité a posteriori pour que la forme \underline{x} provienne de la classe w_j , et elle est définie par la formule de Bayes [DUB90] :

$$p(w_j | \underline{x}) = \frac{f(\underline{x} | w_j) \cdot p(w_j)}{\sum_{i=1}^c f(\underline{x} | w_i) \cdot p(w_i)} \quad (1.3)$$

La quantité au dénominateur est la densité de probabilité non conditionnelle de \underline{x} appelée aussi densité mélange $f(\underline{x})$. Si la forme de $f(\underline{x} | w_j)$ est inconnue, on utilise des méthodes non paramétriques soit pour l'estimation de $f(\underline{x} | w_j)$ telles que les noyaux de Parzen [DUB90] ou pour l'estimation des probabilités a posteriori $p(w_j | \underline{x})$ en utilisant par exemple la méthode des k plus proches voisins (k -ppv) [COV67] et ses extensions, ou les réseaux de neurones [PEL93].

Il existe d'autres méthodes de classification qui divisent directement l'espace de représentation en des zones de décision en utilisant des indicateurs, des sorties binaires, telles que les réseaux de neurones basés sur la notion de distance [REI82] ou les réseaux Learning Vector Quantization [KOH90] ou encore les classifieurs linéaires tels que les réseaux perceptron.

Dans l'approche analytique, une fonction discriminante est déterminée en estimant les paramètres d'une fonction mathématique, de manière à séparer au mieux les classes, c'est-à-dire à minimiser la probabilité de mauvaise classification. Le choix de la fonction mathématique dépend de la complexité de la frontière de décision. En pratique, on commence par des fonctions linéaires puis si la probabilité de mauvaise classification n'est pas satisfaite, on passe à des fonctions mathématiques plus complexes comme par exemple des fonctions quadratiques.

1.3.3. Phase d'exploitation

Dans cette phase, le système de diagnostic par RdF peut être mis en service. Il permet de classer chaque nouvelle observation recueillie sur le système dans une des classes connues, en appliquant la règle de décision élaborée dans la phase d'analyse. La détermination de cette classe permet de connaître le mode de fonctionnement actuel du système.

1.4. Notion de rejet

Son intégration au sein d'un système de diagnostic par RdF permet de diminuer l'erreur de classification en utilisant deux types de rejet : le rejet en ambiguïté et le rejet en distance. Le rejet en ambiguïté, introduit par Chow [CHO57], affecte une observation, se situant approximativement à égale distance entre plusieurs classes, dans une nouvelle classe appelée la classe de rejet en ambiguïté w_o . Cette classe rassemble les observations correspondant au cas d'indécision.

Lorsqu'une observation à classer se trouve dans une zone de l'espace éloignée de toute classe, elle n'est représentée par aucune classe connue. Pour cela, Dubuisson [DUB90] introduit une règle de rejet en distance. Les points situés loin des classes connues sont affectés dans une classe appelée la classe de rejet en distance w_d . La règle de décision incluant à la fois le rejet en ambiguïté et le rejet en distance dans le cas paramétrique bayésien est [DUB90] :

$$\underline{x} \text{ est classée dans } \begin{cases} w_d & \text{si } f(\underline{x}) < C_d \\ w_j & \text{si } f(\underline{x}) \geq C_d \text{ et } p(w_j | \underline{x}) = \max_{i=1..c} p(w_m | \underline{x}) \geq 1 - C_r \\ w_0 & \text{si } f(\underline{x}) \geq C_d \text{ et } \max_{i=1..c} p(w_i | \underline{x}) < 1 - C_r \end{cases} \quad (1.4)$$

où C_r désigne le coût de rejet en ambiguïté et C_d désigne le seuil de rejet en distance. La Figure 1.2 montre le découpage de l'espace de représentation de dimension a en $c + 2$ classes.

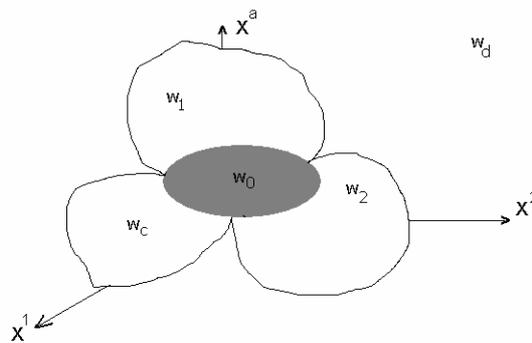


Figure 1.2. Partition de l'espace de représentation en $c + 2$ classes.

La notion de rejet est la base de toute procédure de décision adaptative [BOU97]. Les rejets en ambiguïté peuvent amener à fusionner deux classes et les rejets en distance peuvent engendrer la création de nouvelles classes sous le contrôle du module d'adaptation. La Figure 1.3 montre la structure d'un système de diagnostic par RdF.

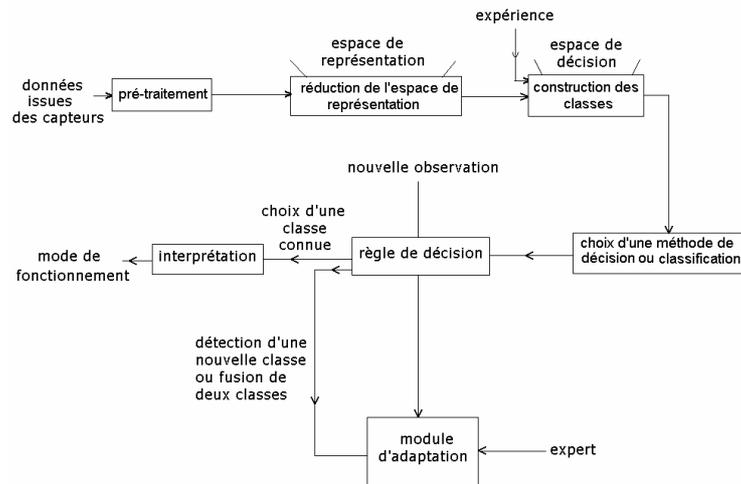


Figure 1.3. Structure d'un système de diagnostic par RdF.

1.5. Reconnaissance floue des formes

1.5.1. Principe

Dans la théorie classique, l'appartenance d'une observation \underline{x} à une classe w_j est définie à l'aide d'une fonction binaire abrupte appelée fonction caractéristique $\mu_j(\underline{x})$:

$$\mu_j(\underline{x}) = \begin{cases} 1 & \text{si } \underline{x} \in w_j \\ 0 & \text{si } \underline{x} \notin w_j \end{cases} \quad (1.5)$$

La nature imprécise des informations issues d'un système réel limite les performances de l'approche classique. En effet, l'étiquetage des observations est fait par un expert humain. Cet étiquetage est une source d'incertitude puisque la connaissance de l'expert sur le système n'est jamais parfaite. De plus, une observation située au centre d'une classe et une observation située en bordure de cette même classe n'ont pas la même représentativité de la classe, pourtant elles sont affectées à la même classe avec la même valeur d'appartenance, cela crée une autre source d'incertitude. Le bruit, dû à l'utilisation des capteurs ou au système lui-même, constitue une troisième source d'incertitude.

La théorie des ensembles flous, introduite par Zadeh [ZAD65], est l'une des solutions pour gérer les différents types d'incertitudes intervenant dans l'utilisation du système de diagnostic par RdF. Elle substitue la notion de fonction caractéristique par une fonction d'appartenance. Celle-ci est une fonction continue sur $[0, 1]$ qui permet de qualifier le degré de compatibilité entre une observation \underline{x} et une classe w_j . \underline{x} est d'autant plus compatible avec w_j que son

degré d'appartenance est proche de 1. Cet aspect graduel ou continu de la fonction d'appartenance, et non plus discontinu, permet de nuancer le degré de ressemblance d'une observation à une classe. La continuité de la fonction d'appartenance permet de suivre l'évolution des nouvelles observations dans l'espace de décision flou ce qui permet de faire un pronostic sur l'évolution future du système [FRE92]. La Figure 1.4 montre la différence de concept entre une fonction caractéristique et une fonction d'appartenance.

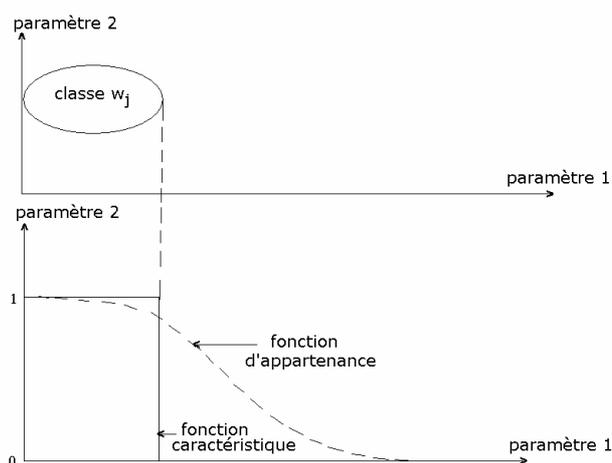


Figure 1.4. Illustration du concept d'appartenance au sens classique et au sens flou.

1.5.2. Application de la logique floue en décision

En RdF floue, une observation \underline{x} n'appartient pas totalement à une classe mais partiellement à toutes les classes. Ce concept entraîne la présence d'un vecteur d'appartenance $\mu_j(\underline{x})$, de dimension $c \times 1$, pour chaque observation. Chaque élément du vecteur d'appartenance représente le coefficient d'appartenance de l'observation à chaque classe. La juxtaposition des vecteurs d'appartenance, de toutes les observations de l'ensemble d'apprentissage, conduit à la matrice d'appartenance U de dimension $c \times N$. L'élément u_{jn} représente le coefficient d'appartenance de l'observation \underline{x}_n à la classe w_j . Cette matrice respecte les propriétés suivantes :

$$\forall j = 1 \dots c, \forall n = 1 \dots N : u_{jn} \in [0, 1] \quad (1.6)$$

$$\forall j = 1 \dots c : 0 < \sum_{n=1}^N u_{jn} < N \quad (1.7)$$

$$\forall n = 1 \dots N : \sum_{j=1}^c u_{jn} = 1 \quad (1.8)$$

(1.6) traduit la condition de normalisation et (1.7) assure que le nombre de classes est plus grand que 1 et impose qu'aucune classe ne soit vide. (1.8) traduit la condition d'orthogonalité ou le concept de totale d'appartenance, elle est appelée l'hypothèse probabiliste.

Une règle de décision floue est limitée au calcul du vecteur d'appartenance $\mu_j(\underline{x})$ et s'exprime de la manière suivante :

$$\underline{x} \text{ est classée dans } w_j \text{ avec le degré d'appartenance } \mu_j(\underline{x}) \quad (1.9)$$

Un système de décision basé sur une telle règle est peu efficace d'un point de vue utilisateur car aucune décision n'est réellement prise. En effet, l'observation \underline{x} peut être considérée comme rejetée en ambiguïté car \underline{x} est classée dans toutes les classes. Pour pallier ce problème, la solution intuitive consiste à défuzzyfier cette décision en associant \underline{x} à la classe pour laquelle elle a la valeur maximale d'appartenance. Cette règle, appelée la règle du maximum d'appartenance, s'écrit :

$$\underline{x} \text{ est classée dans } w_j \text{ si } \mu_j(\underline{x}) = \max_{i=1 \dots c} \mu_i(\underline{x}) \quad (1.10)$$

Cette règle ne permet pas le rejet en ambiguïté, l'approche floue servant à gérer l'incertitude, liée à la classification d'une observation, n'existe donc plus. Pour remédier à ce problème, une nouvelle règle floue de décision avec option de rejet a été développée [FRE92].

1.5.2.1. Règle floue avec rejet en ambiguïté

Frélicot [FRE92] a proposé une nouvelle règle de décision floue incluant la notion de rejet en ambiguïté, basée sur l'analyse du rapport d'appartenance R défini par :

$$R = \frac{\max_{i=1 \dots c} \mu_i(\underline{x})}{\max_{j=1 \dots c; j \neq i} \mu_j(\underline{x})} = \frac{\mu_1}{\mu_2} \quad (1.11)$$

Ce rapport est compris entre 0 et 1. Si $\mu_1 \gg \mu_2$ alors, $R \rightarrow 0$ et \underline{x} peut être classifiée dans la classe w_j , pour laquelle la valeur de la fonction d'appartenance est maximale. Dans ce cas, le risque de mauvaise classification est très faible. Si $\mu_1 \approx \mu_2$ alors $R \rightarrow 1$, et le risque de mauvaise classification entre au moins deux classes, w_i et w_j , est très important. Dans ce cas, \underline{x} doit être rejetée en ambiguïté.

La nouvelle règle floue de décision, baptisée règle du rapport d'appartenance, incluant la notion de rejet en ambiguïté est donnée par [FRE92] :

$$\begin{aligned} \underline{x} \text{ est classée dans } w_j \text{ si } \mu_j(\underline{x}) = \max_{i=1..c} \mu_i(\underline{x}) \text{ et } R \leq S_{amb}, \\ \underline{x} \text{ est classée dans } w_0 \text{ si } R > S_{amb} \end{aligned} \quad (1.12)$$

S_{amb} est le seuil de rejet en ambiguïté compris entre 0 et 1.

1.5.2.2. Règle floue avec rejet en appartenance

En RdF floue, une observation \underline{x} se situant loin d'une classe devrait avoir une valeur d'appartenance faible à cette classe. Le rejet en distance devient alors un rejet en appartenance. Cette notion peut être définie de la manière suivante : une observation \underline{x} est rejetée en appartenance si la plus grande de ses valeurs d'appartenance est inférieure à un seuil S_{app} [FRE92] :

$$\underline{x} \text{ est classée dans } w_d \text{ si } \mu_j(\underline{x}) = \max_{i=1..c} \mu_i(\underline{x}) < S_{app} \quad (1.13)$$

La règle de rejet en ambiguïté est appliquée sur le sous ensemble de classes J pour lesquelles \underline{x} n'est pas rejetée en appartenance :

$$J = \{w_j : \mu_j(\underline{x}) \geq S_{app}\} \quad (1.14)$$

La règle de décision floue incluant les deux types de rejet est définie comme suit [FRE92] :

$$\begin{aligned} \underline{x} \text{ est classée dans } w_d \text{ si } J = \emptyset, \\ \underline{x} \text{ est classée dans } w_j \text{ si : } J = \{w_j\} \\ \text{ou si } |J| > 1 \text{ et } R \leq S_{amb} \text{ et } \mu_j(\underline{x}) = \max_{i:w_i \in J} \mu_i(\underline{x}), \\ \underline{x} \text{ est classée dans } w_0 \text{ si } |J| > 1 \text{ et } R > S_{amb} \end{aligned} \quad (1.15)$$

Il est possible de considérer un seuil d'appartenance $(S_{app})_j$ pour chaque classe w_j . Frélicot [FRE92] a proposé de les calculer de la manière suivante :

$$(S_{app})_j = \min_{\forall \underline{x} \in w_j} \mu_j(\underline{x}), \quad j = 1 .. c \quad (1.16)$$

A cause de la condition d'orthogonalité (1.8), un point situé loin de toutes les classes a un coefficient d'appartenance égal à $\frac{1}{c}$ pour chacune d'elles. Si le nombre de classes est faible, $\frac{1}{c}$ sera supérieur au seuil de rejet en appartenance qui est généralement petit. Le rejet en

appartenance n'est donc pas applicable.

1.5.3. Méthodes d'apprentissage flou non-supervisé

Un problème d'apprentissage non supervisé classique consiste à trouver une partition de l'ensemble d'apprentissage, contenant N individus non étiquetés, en c classes disjointes au sens d'un critère de similarité, avec $2 \leq c < N$. Ces classes sont décrites par une fonction d'appartenance définie par (1.5).

Les méthodes d'apprentissage classique non supervisé ne sont pas très satisfaisantes lorsque l'ensemble d'apprentissage n'est pas structuré en classes clairement séparable mais en classes chevauchantes. L'apprentissage flou non supervisé constitue une bonne alternative. La fonction d'appartenance est définie par (1.6). La coalescence floue peut servir de base à la coalescence classique en recréant des classes non floues à partir des maxima d'appartenance définis par (1.10).

La plus connue des méthodes floues d'apprentissage non supervisé est la méthode floue des C Moyennes (FCM). Elle est une extension directe de la méthode des Centres Mobiles étudiée essentiellement par Bezdek [BEZ81]. Elle repose sur une technique itérative minimisant le critère de type moindres carrés :

$$J_m(U, V) = \sum_{n=1}^N \sum_{j=1}^c (u_{jn})^m \cdot (d_{jn})^2 \quad (1.17)$$

où $V = (v_1, v_2, \dots, v_c)$ est l'ensemble des centroïdes des classes dans \mathbb{R}^a et U est la matrice de partition floue, donnés par les expressions suivantes :

$$v_j = \frac{\sum_{n=1}^N (u_{jn})^m \cdot x_n}{\sum_{n=1}^N (u_{jn})^m}, \quad j = 1 \dots c \quad (1.18)$$

$$u_{jn} = \frac{1}{\sum_{i=1}^c \left(\frac{\|x_n - v_j\|^2}{\|x_n - v_i\|^2} \right)^{\frac{1}{m-1}}}, \quad j = 1 \dots c \quad (1.19)$$

m est un exposant contrôlant la quantité de flou de la partition résultante. Lorsque $m = 1$, on retrouve la méthode nette des Centres Mobiles et lorsque $m \rightarrow \infty$, la partition est totalement

floue. Généralement on utilise $m = 2$. d_{jn} est la distance entre l'observation x_n et le centroïde de la classe w_j :

$$d_{jn} = (\underline{x}_n - v_j)^T \cdot A \cdot (\underline{x}_n - v_j) \quad (1.20)$$

où A est une matrice définie positive de dimension $N \times N$.

Les termes de la matrice U doivent satisfaire les conditions suivantes :

$$\forall j, n : u_{jn} \in [0, 1], \quad \forall n : \sum_{j=1}^c u_{jn} = 1, \quad \text{et} \quad \forall j : 0 < \sum_{n=1}^N u_{jn} < N$$

La méthode FCM s'arrête lorsque la partition reste stable entre 2 itérations successives c'est-à-dire lorsque le critère d'arrêt $\|U^{it} - U^{it+1}\|$ devient inférieur à un seuil ε fixé.

L'algorithme FCM se décompose alors comme suit :

Début

Fixer $c : 2 \leq c < N$ et $m : 1 < m < \infty$

Choisir la matrice A

Initialiser aléatoirement U^0 suivant (1.6)

Répéter $it = 1, 2, \dots$

Calculer les nouveaux centroïdes des classes V^{it} selon (1.18)

Mettre à jour la matrice de partition floue U^{it} selon (1.19)

Jusqu'à $\|U^{it} - U^{it+1}\| < \varepsilon$

Fin

L'algorithme converge vers un minimum local de J_m . Différents choix d'initialisation de U^0 peuvent conduire à différents minima et donc à des partitions différentes. La Figure 1.5 illustre le déplacement des centroïdes pour un nuage comportant trois classes, le seuil de convergence est fixé à 0.01.

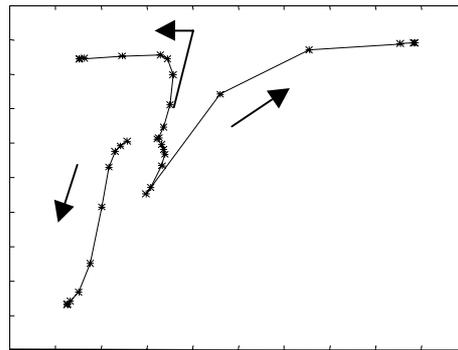


Figure 1.5. Déplacement des centroïdes pour un nuage de points à 3 classes.

Nous remarquons que les derniers déplacements sont de plus en plus petits. Ceci traduit une stabilisation de la partition. Il est prouvé que l'algorithme converge toujours mais il faut éviter les minima locaux, en ajustant judicieusement la valeur du seuil de convergence.

FCM impose la structure des classes en fixant a priori la matrice A . Si A est la matrice identité, la distance est la distance euclidienne et les classes obtenues sont de forme sphérique. Si A est l'inverse de la matrice de covariance, nous avons la distance de Mahalanobis qui donne des classes de forme elliptique. Pour remédier à ce problème, des développements de FCM ont été effectués. On cite par exemple l'algorithme de Gustafson et Kassel FCMGK qui considère que la matrice de distance A est une variable du critère à minimiser [GUS79]. Cependant, cette méthode est inefficace pour des classes de forme non-elliptique.

De plus, FCM nécessite la connaissance a priori du nombre de classes dans l'ensemble d'apprentissage. Cela pose un problème si le nombre de classes est inconnu. Le remède à ce problème est l'utilisation de critères de validation des classes. Ces critères sont en général utilisés pour savoir si l'ensemble d'apprentissage possède réellement une structure de classes puis pour vérifier si le nombre de classes obtenu est significatif de cette structure. On peut citer des critères conçus spécialement pour l'algorithmes FCM comme le coefficients de partition de Bezdek et l'entropie de classification [BEZ74]. D'autres critères peuvent être appliqués à n'importe quel algorithme de coalescence floue comme le critère de l'hypervolume flou [GAT89] et le critère de compacité et de séparabilité [XIE91]. Le Tableau 1.1 montre les formules mathématiques et les règles d'utilisation de ces critères.

Critères	Formules	Règles d'utilisation
Critère de compacité et de séparabilité	$CS(U, c) = \frac{c_o}{s_e}$	$(c^*, U^*) = \min_{c=2, n-1} [CS(U, c)]$
Coefficient de partition	$F(U, c) = \frac{1}{N} \sum_{n=1}^N \sum_{j=1}^c u_{jn}^2$	$(c^*, U^*) = \max_{c=2, N-1} [F(U, c)]$
Entropie de classification	$H(U, c) = -\frac{1}{N} \sum_{n=1}^N \sum_{j=1}^c u_{jn} \cdot \log(u_{jn})$	$(c^*, U^*) = \min_{c=2, N-1} [H(U, c)]$
Hypervolume flou	$F_{HV}(U, c) = \sum_{j=1}^c [\det(\Sigma_j)]^{(1/2)}$	$(c^*, U^*) = \min_{c=2, N-1} [F_{HV}(U, c)]$

Tableau 1.1. Critères de validité des classes

Dans ce tableau :

Σ_j est la matrice de covariance floue de la classe j :

$$\Sigma_j = \frac{S_j}{\sum_{n=1}^N (u_{jn})^m}$$

où S_j est la matrice de dispersion floue :

$$S_j = \sum_{n=1}^N (u_{jn})^m \cdot (\underline{x}_n - v_j)(\underline{x}_n - v_j)^T$$

La compacité c_o est définie par :

$$c_o = \frac{1}{N} \cdot \sum_{j=1}^c \sum_{n=1}^N [u_{jn} \cdot \|\underline{x}_n - v_j\|]^2$$

et la séparabilité s_e est déterminée par :

$$s_e = \left[\min_{j=1, c; i=1, c; j \neq i} \|v_j - v_i\| \right]^2$$

Le nombre optimum de classes c est déterminé en exécutant la méthode de coalescence pour différentes valeurs de c et en retenant la partition pour laquelle le critère de validité est optimal. Cependant ces critères présentent généralement plusieurs optima locaux et la méthode FCM reste itérative. Pour remédier aux inconvénients de la méthode FCM, [BIL98] et [BIL99] proposent les méthodes Unsupervised Graph Clustering (UGC) et Unsupervised Fuzzy Graph Clustering (UFGC). Ces méthodes sont basées sur l'approche multi-prototype pour résoudre le problème de la forme non-élliptique des classes et sur la méthode

hiérarchique pour déterminer le nombre de classes. Elles utilisent la méthode FCM pour partitionner le nuage de points en surestimant le nombre de sous-classes.

Le rejet en appartenance n'est pas applicable par la méthode FCM à cause de l'hypothèse probabiliste comme le montre la Figure 1.6. En effet le point B, se trouvant loin des 2 classes, possède des valeurs d'appartenance relativement élevées, surtout pour la première classe, par rapport à un seuil de rejet. Cela donne un autre inconvénient de la méthode FCM. Cette inconvénient est commun à toutes les méthodes d'apprentissage non supervisé qui utilisent l'approche floue.

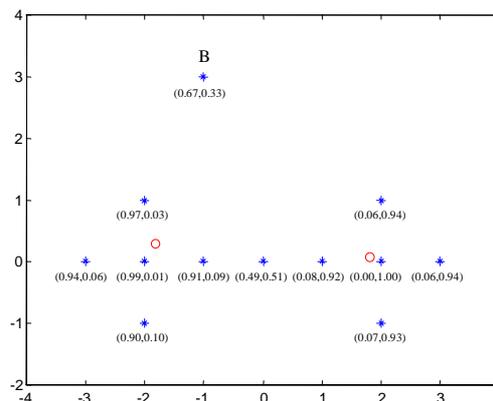


Figure 1.6. Nuage test. Les étoiles représentent les points du nuage. Les cercles correspondent aux centroïdes des 2 classes. Les valeurs entre parenthèses sont les coefficients d'appartenance à chaque classe.

1.5.4. Méthodes floues de classification supervisée

La méthode floue de classification supervisée la plus connue est la méthode floue des k Plus Proches Voisins (k-ppv). Cette méthode, très facile à mettre en œuvre, substitue l'hypothèse de la connaissance de la loi de densité conditionnelle par l'estimation de cette densité en utilisant l'ensemble d'apprentissage étiqueté : $X = \{(\underline{x}_n, w_j)\}$ avec $j = 1 \dots c$ et $n = 1 \dots N$.

La règle des k-ppv permet d'estimer la densité de probabilité conditionnelle, de l'observation \underline{x} , par rapport à la classe w_j :

$$f(\underline{x} \mid w_j) = \frac{k_j}{N.V} \quad (1.21)$$

ou sa probabilité a posteriori :

$$p(w_j | \underline{x}) = \frac{f(\underline{x} | w_j)}{\sum_{i=1}^c f(\underline{x} | w_i)} = \frac{k_j}{k} \quad (1.22)$$

V est un volume de \mathbb{R}^a centré sur \underline{x} qui contient les k -ppv de ce dernier dont k_j appartient à la classe w_j . V est suffisamment petit pour qu'on puisse considérer que f ne varie pas. Le vecteur \underline{x} est affecté à la classe w_j pour la règle 1-ppv si :

$$\underline{x} \in w_j \text{ si } \exists \underline{x}_u \in w_j, d(\underline{x}, \underline{x}_u) = \min_{\underline{x}_n \in X} d(\underline{x}, \underline{x}_n) \quad (1.23)$$

$d(\underline{x}, \underline{x}_n)$ est la distance entre l'observation \underline{x} et son plus proche voisin \underline{x}_n . Parmi les distances les plus utilisées en RdF on peut citer la distance de Minkowski, la distance de Manhattan, la distance euclidienne, la distance du max et la distance de Mahalanobis. On emploie le plus souvent la distance Euclidienne.

Une extension de la règle 1-ppv est la règle k -ppv pour laquelle le vecteur forme \underline{x} est affecté à la classe w_j si cette classe est représentée de façon majoritaire parmi les k plus proches voisins de \underline{x} :

$$\underline{x} \in w_j \text{ si } k_j = \max_{i=1..c} k_i \quad (1.24)$$

Les performances de la méthode des k -ppv dépendent de la valeur de k . Son choix est toujours délicat. Le seul moyen pour le déterminer est le critère heuristique $k = \sqrt{N}$ défini par Dubuisson [DUB90]. En pratique, pour une taille finie de l'ensemble d'apprentissage, la valeur optimale de k doit être déterminée expérimentalement par validation croisée.

Il existe plusieurs approches pour introduire la notion de flou dans la méthode des k -ppv, comme par exemple l'algorithme présenté par Béreau [BER86] ou l'algorithme présenté par Keller où le coefficient d'appartenance d'un nouveau vecteur forme \underline{x} à la classe w_j est donné par [KEL85] :

$$\mu_j(\underline{x}) = \frac{\sum_{n=1}^k \mu_{jn} \left(\frac{1}{\|\underline{x} - \underline{x}_n\|^{\frac{2}{m-1}}} \right)}{\sum_{n=1}^k \left(\frac{1}{\|\underline{x} - \underline{x}_n\|^{\frac{2}{m-1}}} \right)} \quad (1.25)$$

μ_{jn} est le coefficient d'appartenance à la classe w_j du $n^{\text{ième}}$ vecteur forme parmi les k -ppv de \underline{x} .

La règle floue des k -ppv la plus couramment utilisée est la règle du maximum d'appartenance qui consiste à affecter \underline{x} à la classe dont la fonction d'appartenance est maximale :

$$\underline{x} \in w_j \text{ si } \mu_j(\underline{x}) = \max_{i=1..c} \mu_i(\underline{x}) \quad (1.26)$$

Pour diminuer le risque de mauvaise classification, cette règle peut être modifiée par l'introduction de l'option de rejet. La condition de normalisation ou d'orthogonalité imposée par l'hypothèse probabiliste interdit l'utilisation du rejet en appartenance. Celui-ci peut être remplacé par le rejet en distance. La règle de décision avec les options de rejet en ambiguïté et en distance devient :

$$\begin{aligned} \underline{x} &\in w_d \text{ si } \bar{d}(\underline{x}, k - ppv) > T_d \\ \underline{x} &\in w_j \text{ si } \bar{d}(\underline{x}, k - ppv) \leq T_d \text{ et } k_j = \max_{i=1..c} k_i \geq k' \\ \underline{x} &\in w_0 \text{ si } \bar{d}(\underline{x}, k - ppv) \leq T_d \text{ et } k_j = \max_{i=1..c} k_i < k' \end{aligned} \quad (1.27)$$

$\bar{d}(\underline{x}, k - ppv)$ représente la distance moyenne entre \underline{x} et ses k plus proches voisins, T_d est un seuil de rejet en distance et k' est le nombre minimum, de plus proches voisins appartenant à la même classe, requis pour que le point ne soit pas rejeté en ambiguïté.

1.6. Conclusion

Le diagnostic d'un système quelconque est l'identification d'une défaillance par ses manifestations extérieures. Son principe général consiste à confronter les données relevées au cours du fonctionnement réel du système avec la connaissance que l'on a de son fonctionnement normal ou défaillant. Par contre, la supervision est définie comme étant la surveillance et le contrôle de l'exécution d'une opération ou d'un travail accompli par un homme ou par une machine. En présence de défaillance, le dispositif de supervision prendra toutes les démarches nécessaires pour le retour du système vers un mode de fonctionnement normal.

Le fonctionnement d'un système de diagnostic par Reconnaissance des Formes (RdF) se déroule en trois phases : une phase d'analyse, une phase de choix d'une méthode de décision et une phase d'exploitation.

La nature imprécise des informations issues d'un système réel limite les performances de l'approche classique. En effet, l'étiquetage des observations est fait par un expert humain. Cet étiquetage est une source d'incertitude puisque la connaissance de l'expert sur le système n'est jamais parfaite. De plus, une observation située au centre d'une classe et une observation située en bordure de cette même classe n'ont pas la même représentativité de la classe, pourtant elles sont affectées à la même classe avec la même valeur d'appartenance, cela crée une autre source d'incertitude. Le bruit, dû à l'utilisation des capteurs ou au système lui-même, constitue une troisième source d'incertitude. Pour traiter ces sources d'incertitude, l'approche floue a été introduite au diagnostic par RdF.

La notion de rejet est la base de toute procédure de décision adaptative. Son intégration au sein d'un système de diagnostic par RdF permet de diminuer l'erreur de classification en utilisant deux types de rejet : le rejet en ambiguïté et le rejet en distance ou en appartenance.

L'inconvénient de l'approche floue est la condition d'orthogonalité qui empêche le rejet en appartenance d'une observation se situant loin de toutes les classes connues. Pour remédier à cet inconvénient, d'autres méthodes de représentation et de traitement de l'incertitude ont été introduites.

Le chapitre suivant est consacré à une étude comparative des différentes représentations de l'incertitude que sont : la théorie des probabilités, la théorie des possibilités et la théorie de Dempster et Shafer. Nous nous concentrons sur la méthode possibiliste parce qu'elle est plus souple et plus efficace pour traiter des informations imparfaites et parce qu'elle est capable de remédier au problème de la condition d'orthogonalité. Nous présenterons par la suite une méthode de classification non supervisée et une méthode de classification supervisée qui utilisent l'approche possibiliste.

2. Théorie des possibilités

2.1. Introduction

Dans les systèmes de diagnostic par RdF, la précision des informations, avec lesquelles on est amené à travailler, ainsi que leur fiabilité sont limitées [ZOU97]. La justesse de la décision dépend de la prise en compte totale ou partielle de l'incertitude et de l'imprécision des informations utilisées. Cela exige des méthodes très performantes permettant la modélisation et la manipulation des imperfections sans perdre d'information.

Le problème de l'imprécision a été traité par le calcul d'erreurs, restreint aux imprécisions de caractère numérique. Au lieu d'affecter une valeur unique à une variable donnée, un intervalle des valeurs possibles lui est affecté. Au bout d'un certain nombre de mesures, les erreurs se cumulent et l'intervalle devient tellement large qu'il perd toute signification. Il manque à ce type de représentation une méthode qui permet d'ordonner les différentes valeurs de l'intervalle suivant leur degré de vraisemblance [ZOU97].

La théorie des probabilités a été introduite pour ne représenter que l'incertitude contenue dans les données. La théorie des ensembles flous, quant à elle, a été proposée par Zadeh [ZAD65] comme un outil pour ne prendre en compte que l'imprécision incluse dans les données. Il se peut cependant que des informations comprennent simultanément les deux types d'imperfection, c'est par exemple le cas avec la phrase : il faudrait probablement que le système soit refroidi quand la température atteint environ 39 degré. L'imprécision concerne le contenu de l'information, tandis que l'incertitude concerne sa véracité [DUB87]. La théorie des possibilités et la théorie de Dempster-Shafer ont été introduites pour la représentation et le traitement de l'information à la fois imprécise et incertaine.

2.2. La théorie des ensembles flous

La théorie des ensembles flous a été introduite par Zadeh [ZAD65] pour éviter la transition brusque entre deux ensembles flous et pour représenter les connaissances imprécises [ZOU91]. Un sous-ensemble flou F d'un univers de discours Ω est défini par une fonction d'appartenance $\mu_F : \Omega \rightarrow [0, 1]$, qui associe à chaque élément w de Ω un coefficient $\mu_F(w)$ indiquant son degré d'appartenance à F . Plus w s'approche de la caractérisation typique du sous-ensemble flou F , plus sa valeur d'appartenance $\mu_F(w)$ tend vers 1. Le sous ensemble flou F est défini par la connaissance des couples $(w, \mu_F(w))$.

Les opérations d'intersection, d'union et de complémentarité des sous-ensembles flous peuvent être définis à l'aide de différents opérateurs comme le minimum pour l'intersection, le maximum pour l'union et la négation pour le complémentaire.

L'intersection de deux sous-ensembles flous A et B de Ω , est un sous-ensemble flou de Ω , noté $A \cap B$, défini par :

$$\forall w \in \Omega, \mu_{A \cap B}(w) = \min(\mu_A(w), \mu_B(w)) \quad (2.1)$$

L'union de deux sous-ensembles flous A et B de Ω est un sous-ensemble flou de Ω , noté $A \cup B$, défini par :

$$\forall w \in \Omega, \mu_{A \cup B}(w) = \max(\mu_A(w), \mu_B(w)) \quad (2.2)$$

Le complémentaire \bar{A} d'un sous-ensemble flou A de Ω est un sous-ensemble flou de Ω défini par :

$$\forall w \in \Omega, \mu_{\bar{A}}(w) = 1 - \mu_A(w) \quad (2.3)$$

2.3. Mesures de confiance

Sugeno [SUG74] a proposé un cadre très général pour traiter des informations incertaines et imprécises en introduisant la notion de mesure de confiance. Considérons un ensemble Ω d'événements, la croyance ou la confiance qu'on a en la réalisation d'un événement A de Ω sera mesurée par la quantité $g(A)$. Une mesure de confiance g est une fonction définie sur l'ensemble $S(\Omega)$ des sous ensembles de Ω et ayant des valeurs dans l'intervalle $[0, 1]$ telle que :

$$g(\Phi) = 0 \text{ et } g(\Omega) = 1, \quad (2.4)$$

$$\forall A, B \subset S(\Omega), \text{ si } A \subseteq B \text{ alors } g(A) \leq g(B) \quad (2.5)$$

En conséquence immédiate de l'axiome de monotonie (2.5), les mesures de confiance de l'union et de l'intersection de deux sous-ensemble A et B de Ω sont données par [ZOU97] :

$$g(A \cup B) \geq \max(g(A), g(B)) \quad (2.6)$$

$$g(A \cap B) \leq \min(g(A), g(B)) \quad (2.7)$$

Les axiomes de Sugeno, (2.6) et (2.7), caractérisent une vaste famille de mesures de

confiance. Les plus répandues sont : les mesures de probabilité, les mesures de possibilité et les fonctions de croyance de Dempster-Shafer.

2.3.1. Théorie des probabilités

La théorie des probabilités est le modèle le plus traditionnel pour la modélisation de l'information incertaine. La probabilité de l'occurrence d'un événement A de Ω est décrite par la mesure de probabilité $P : S(\Omega) \rightarrow [0, 1]$, qui satisfait les axiomes suivants [SAN91] :

$$\forall A \subset \Omega, P(\Phi) = 0 \leq P(A) \leq P(\Omega) = 1 \quad (2.8)$$

$$\forall A, B \subset \Omega, \text{ si } A \cap B = \Phi, P(A \cup B) = P(A) + P(B) \quad (2.9)$$

L'axiome d'additivité (2.9) établit que, si deux événements sont mutuellement exclusifs, alors la probabilité qu'au moins l'un entre eux ait lieu est la somme de leur probabilités individuelles. On peut déduire de (2.8) et (2.9) la conséquence suivante : si la probabilité d'un événement A est connue alors la probabilité de l'événement contraire \bar{A} est exactement déterminée :

$$\forall A \subseteq \Omega, P(A) + P(\bar{A}) = 1 \quad (2.10)$$

On peut caractériser la probabilité des événements sur un domaine discret Ω en utilisant la distribution de probabilité $p : \Omega \rightarrow [0, 1]$ où $\sum_{w \in \Omega} p(w) = 1$. La fonction p peut être définie à partir de P par $p(w) = P(\{w\})$. De l'axiome d'additivité (2.9), on déduit :

$$\forall A \subset \Omega, P(A) = \sum_{w \in A} p(w) \quad (2.11)$$

La mesure de probabilité P a une interprétation soit fréquentiste soit subjective. Dans l'interprétation fréquentiste, utilisée dans le cas des informations statistiques, $P(A)$ représente la limite de la fréquence de l'occurrence de l'événement A dans une séquence infinie d'expériences indépendantes. Dans l'interprétation subjective, $P(A)$ représente la croyance qu'un expert accorde à l'occurrence de A .

Lorsque certains indices observés permettent d'augmenter la confiance à une hypothèse, l'axiome d'additivité (2.9) nécessite de diminuer d'autant la confiance correspondant à l'hypothèse contraire. Néanmoins, dans certains cas, des informations peuvent très bien favoriser une hypothèse sans pour autant discréditer l'hypothèse contraire. La théorie des probabilités semble donc peu adaptée à des situations où la connaissance d'un événement

comme la connaissance de son contraire sont très limitées [ZOU97]. De plus, la théorie des probabilités est trop rigide pour exprimer le cas de l'ignorance totale, en effet elle modélise ce cas par un ensemble d'événements mutuellement disjoints et équiprobables :

$$\forall w \in \Omega, p(w) = \frac{1}{|\Omega|} \quad (2.12)$$

2.3.2. Théorie des possibilités

La théorie des possibilités a été introduite par Zadeh [ZAD78] comme une dérivée de la théorie des ensembles flous pour prendre en compte l'incertitude incluse dans des objets imprécis. Une mesure de possibilité $\Pi : S(\Omega) \rightarrow [0, 1]$ est une fonction qui vérifie les axiomes suivantes [DUB87] :

$$\Pi(\Omega) = 1, \Pi(\Phi) = 0, \quad (2.13)$$

$$\forall A, B \subset \Omega, \Pi(A \cup B) = \max(\Pi(A), \Pi(B)) \quad (2.14)$$

on peut déduire des axiomes (2.13) et (2.14) que :

$$\max(\Pi(A), \Pi(\bar{A})) = 1 \quad (2.15)$$

De (2.15), on peut constater qu'un événement possible n'interdit pas pour autant à l'événement contraire de l'être également. L'axiome (2.14) montre qu'une mesure de possibilité n'exige pas que les deux événements A et B soient disjoints et qu'une mesure de possibilité n'est pas additive mais a un caractère d'ordre [DUB93].

De façon similaire à la mesure de probabilité, une mesure de possibilité peut être définie à partir d'une distribution de possibilité sur les singletons de Ω :

$$\Pi(\{w\}) = \pi(w), \Pi(A) = \sup_{w \in A} \pi(w) \text{ avec } \sup_{w \in \Omega} \pi(w) = 1 \quad (2.16)$$

Une mesure de possibilité est toujours associée à une mesure de nécessité $N : S(\Omega) \rightarrow [0, 1]$ qui vérifie les axiomes suivantes :

$$N(\Omega) = 1, N(\Phi) = 0, \quad (2.17)$$

$$\forall A, B \subset \Omega, N(A \cap B) = \min(N(A), N(B)) \quad (2.18)$$

La nécessité d'un événement A peut s'exprimer en fonction de la mesure de possibilité de

l'événement contraire \bar{A} :

$$N(A) = 1 - \Pi(\bar{A}) \quad (2.19)$$

Cette relation indique qu'un événement est d'autant plus nécessaire que son contraire n'est pas possible. Elle peut également s'exprimer en fonction de la distribution de possibilité π associée à Π :

$$N(A) = \inf_{w \in \bar{A}} (1 - \pi(w)) \quad (2.20)$$

Le modèle possibiliste permet de modéliser l'imprécision incluse dans une proposition A par un couple de valeurs $(N(A), \Pi(A))$.

De (2.14) et (2.18), les deux mesures, possibilité et nécessité, sont les deux cas limites des mesures de confiance de Sugeno (2.6) et (2.7). Il est à noter la simplicité avec laquelle s'exprime le cas de l'ignorance totale :

$$\forall w \in \Omega, \pi(w) = 1 \quad (2.21)$$

L'hypothèse que les événements sont mutuellement disjoints n'est pas nécessaire. Enonçons quelques propriétés sur la mesure de possibilité et la mesure de nécessité :

$$\forall A \subset \Omega, \Pi(A) + \Pi(\bar{A}) \geq 1, \quad (2.22)$$

$$N(A) + N(\bar{A}) \leq 1, \quad (2.23)$$

$$\min(N(A), N(\bar{A})) = 0, \quad (2.24)$$

$$\text{si } N(A) > 0 \text{ alors } \Pi(A) = 1, \quad (2.25)$$

$$\text{si } \Pi(A) < 1 \text{ alors } N(A) = 0 \quad (2.26)$$

(2.25) indique qu'un événement peu nécessaire est totalement possible.

On peut déduire l'inégalité suivante qui lie entre les mesures de nécessité, de probabilité et de possibilité :

$$N(A) \leq P(A) \leq \Pi(A) \quad (2.27)$$

Considérons maintenant la phrase " v est A ", A peut être interprété comme l'ensemble des

valeurs possibles que peut prendre la variable ν . Si certaines valeurs sont davantage possibles que d'autres, A devient un ensemble flou. Pour un élément w de A , on associe une distribution de possibilité π . $\pi_\nu(w)$ indique la possibilité que la valeur de ν soit w , en revanche $\mu_A(w)$ indique la valeur d'appartenance de w à A . La distribution de possibilité π peut être interprétée comme une fonction d'appartenance d'un ensemble flou [ZOU97, DEV93] :

$$\forall w \in A, \pi_\nu(w) = \mu_A(w) \quad (2.28)$$

Si Ω est l'ensemble des entiers positifs et F est l'ensemble flou de petits entiers décrit par les paires $(w, \mu_F(w))$ suivantes : $\{(1,1), (2,1), (3,0.8), (4,0.6), (5,0.4), (6,0.2)\}$. La proposition “ ν est un petit entier ” est associée avec la distribution de possibilité décrite en utilisant les paires $(w, \pi_\nu(w))$ suivantes : $\pi_\nu(w) = \{(1,1), (2,1), (3,0.8), (4,0.6), (5,0.4), (6,0.2)\}$. La possibilité que l'élément $w \in \Omega$ soit égale à 3, sachant que w est un petit entier, est 0.8, elle est égale au degré d'appartenance de cet élément w à l'ensemble flou F .

2.3.3. Théorie des croyances de Dempster-Shafer

La théorie des fonctions de croyance, initialement introduite par Dempster dans le cadre de l'étude des bornes inférieure et supérieure d'une famille de distributions de probabilité, fut reprise dans un cadre plus général par Shafer [ZOU97, SHA76]. Elle offre une modélisation plus souple que celle de la théorie des probabilités pour représenter des informations incertaines et imprécises issues de plusieurs sources.

Dans cette théorie, l'incertitude liée à un événement $A \subset \Omega$ est mesurée à l'aide de deux fonctions : la fonction de croyance ou de crédibilité, $Bel(A) : S(\Omega) \rightarrow [0, 1]$, et la fonction de plausibilité, $Pl(A) : S(\Omega) \rightarrow [0, 1]$:

$$Bel(A) = \sum_{B \subseteq A} m(B) \quad (2.29)$$

$$Pl(A) = \sum_{A \cap B \neq \emptyset} m(B) \quad (2.30)$$

$m(B)$ est la masse de croyance placée strictement en A et qui ne peut être allouée à aucune hypothèse plus restrictive, compte tenu de l'information disponible. $m(B)$ peut être redistribuée plus spécifiquement aux sous-ensembles de B sous réserve d'un apport d'information supplémentaire [ZOU97]. Si $m(B) \neq 0$, alors B est un élément focal de Ω .

La fonction de croyance $Bel(A)$ prend en compte toutes les hypothèses qui impliquent A , en

revanche la fonction de plausibilité $Pl(A)$ prend en compte toutes les hypothèses qui ne discréditent pas A . L'intervalle $[Bel(A), Pl(A)]$ peut être vu comme un intervalle encadrant une probabilité $P(A)$ mal connue [ZOU97] :

$$Bel(A) \leq P(A) \leq Pl(A) \quad (2.31)$$

La différence $Pl(A) - Bel(A)$ mesure l'ignorance relative à A . Les deux mesures de croyance Bel et de plausibilité Pl mettent en rapport un événement A avec l'événement contraire \bar{A} :

$$Bel(A) = 1 - Pl(\bar{A}) \quad (2.32)$$

(2.32) veut dire que plus on augmente la croyance dans une hypothèse, moins l'hypothèse contraire devient plausible.

La théorie des croyance de Dempster-Shafer présente l'avantage de permettre à la fois de représenter et de faire la distinction entre le cas de l'ignorance totale et le cas de répartition uniforme de la croyance. En théorie des possibilités on est seulement capable de représenter l'ignorance totale par (2.21). Si on veut faire la distinction entre le cas d'ignorance totale et le cas d'équi-répartition de croyance en faisant $\forall w \in \Omega, \pi(\{w\}) = 1/\Omega$, la distribution de possibilité sera inconsistante. En théorie des probabilités on est seulement capable de représenter le cas d'équi-répartition des croyance. En théorie de Dempster-Shafer, on modélise l'ignorance totale par [SAN91] :

$$\forall A \subset \Omega : Bel(\Omega) = 1 \text{ et } Bel(A) = 0 \quad (2.33)$$

L'équi-répartition des croyances est représentée par :

$$\forall w \in \Omega, Pl(\{w\}) = \text{constante} \quad (2.34)$$

Il y a une infinités de distributions de plausibilité qui vérifient (2.34), on choisit pour la simplification la distribution :

$$Pl(\{w\}) = \frac{1}{|\Omega|} \quad (2.35)$$

Cette théorie englobe les mesures de probabilité, de possibilité et de nécessité. La fonction de croyance est dite probabiliste si les éléments focaux sont des singletons. La fonction de croyance est une mesure de nécessité et la fonction de plausibilité est une mesure de possibilité si les éléments focaux sont emboîtés comme le montre Figure 2.1 [ZOU97] :

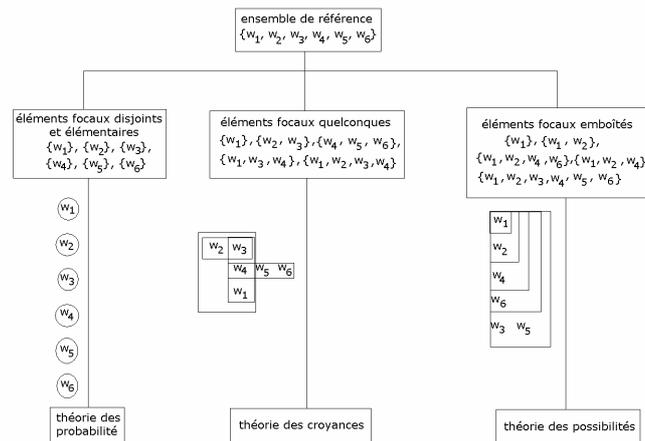


Figure 2.1. La théorie des probabilités et la théorie des possibilités dans le cadre de la théorie des croyances de Shafer

2.3.4. Choix de la théorie de représentation de nos données

Nous traitons des problèmes industriels dont les données imprécises sont modélisées par des distributions de probabilité construites d'après des histogrammes. Ces histogrammes ne sont pas exploitables par la théorie de Dempster-Shafer. En revanche, il existe de nombreuses transformations permettant de passer des distributions de probabilité à des distributions de possibilité.

2.4. Transformations probabilité-possibilité

Dans un système de diagnostic, les informations rencontrées sont à la fois imprécises et incertaines. Cela est dû aux capteurs et différents instruments de mesures, au bruit et aux experts humains.

La théorie des possibilités est bien adaptée pour la représentation et le traitement des connaissances à la fois imprécises et incertaines. La construction d'une mesure de possibilité étant intuitive, on préfère déterminer une mesure probabiliste, pauvre et imprécise par nature, et en déduire la mesure de possibilité par une transformation bijective. Une application de l'utilisation des transformations probabilité-possibilité en diagnostic par RdF floue est la construction de fonctions d'appartenance possibiliste en utilisant des données statistiques [KRI98]. Les transformations de possibilité en probabilité sont utiles dans le cas où l'expert a besoin des informations précises pour prendre une décision.

Une transformation doit garder un lien entre la mesure de probabilité et la mesure de possibilité. Ce lien est appelé le principe de cohérence dont il existe plusieurs approches dans la littérature.

2.4.1. Principe de cohérence de Zadeh

Zadeh est le premier qui proposa une définition du principe de cohérence entre une distribution de probabilité et une distribution de possibilité [ZAD78]. Il considère qu'un degré de possibilité élevé n'implique pas un degré de probabilité élevé, et un degré de probabilité faible n'implique pas un degré de possibilité faible [ZAD78]. Dans le cas extrême, si un événement est impossible, alors il est improbable mais un événement improbable peut être complètement possible. En supposant que : $\forall w_i \in \Omega, p(w_i) = p_i$ et $\pi(w_i) = \pi_i$, Zadeh a défini le degré de cohérence entre une distribution de probabilité (p_1, p_2, \dots, p_n) et une distribution de possibilité $(\pi_1, \pi_2, \dots, \pi_n)$ par :

$$c_z = \sum_{i=1}^n p_i \cdot \pi_i \quad (2.36)$$

Le degré de cohérence de Zadeh (2.36) ne définit pas une relation précise entre une distribution de probabilité et une distribution de possibilité mais il traduit simplement l'effet heuristique que la diminution de la possibilité d'un événement tend à diminuer sa probabilité mais pas l'inverse. Considérons la phrase suivante [ZAD78] : "Hans mangeait m œufs au petit déjeuner" où m appartient à l'ensemble $U = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10)$. $p_m(u \in U)$ est la probabilité que Hans a bien mangé u œufs et $\pi_m(u \in U)$ est la possibilité que Hans ait pu manger u œufs. Les valeurs de probabilité et les valeurs de possibilité correspondantes, selon le principe de cohérence de Zadeh, sont données dans le Tableau 2.1.

u œufs	1	2	3	4	5	6	7	8	9	10
$\pi_m(u \in U)$	1	1	1	1	0.8	0.6	0.4	0.2	0	0
$p_m(u \in U)$	0.1	0.8	0.1	0	0	0	0	0	0	0

Tableau 2.1. Distribution de probabilité et distribution de possibilité correspondante selon le principe de cohérence de Zadeh.

Nous remarquons à travers ce tableau qu'un événement totalement possible peut être improbable alors qu'un événement impossible est forcément improbable. Pour cet exemple, le degré de cohérence de Zadeh est égal à 1. Cela indique une cohérence maximale entre les deux distributions de probabilité et de possibilité selon le principe de Zadeh.

2.4.2. Principe de cohérence de Klir

Klir [KLI93] se base sur le principe de préservation de l'incertitude pour garder un lien entre une distribution de probabilité et une distribution de possibilité, ce principe constitue le seul lien entre deux théories non-comparables :

$$\text{incertitude}(p) = \text{incertitude}(\pi) \quad (2.37)$$

Selon Klir, l'incertitude contenue dans une distribution de probabilité p est exprimé par l'entropie de Shannon $H(p)$. En revanche, l'incertitude $Un(\pi)$ contenue dans une distribution de possibilité π est mesurée par la non-spécificité $SP(\pi)$ et la discordance $D(\pi)$. La non-spécificité $SP(\pi)$ modélise l'incertitude d'une proposition qui n'est pas spécifique, alors que la discordance modélise le degré de conflit interne de la proposition [LAS99]. Pour que le principe de la préservation de l'incertitude soit vérifiée, Klir propose de résoudre l'équation suivante :

$$H(p) = Un(\pi) = SP(\pi) + D(\pi) \quad (2.38)$$

Considérons les distributions de probabilité $p = (p_1, p_2, \dots, p_n, p_{n+1})$ et de possibilité $\pi = (\pi_1, \pi_2, \dots, \pi_n, \pi_{n+1})$ rangées en ordre descendant, avec $p_{n+1} = 0$ et $\pi_{n+1} = 0$. La même démarche sera appliquée pour toute la suite de ce chapitre. L'entropie de Shannon est calculée par :

$$H(p) = -\sum_{i=1}^n p_i \cdot \log_2(p_i) \quad (2.39)$$

et la non-spécificité et la discordance sont calculées par :

$$SP(\pi) = \sum_{i=2}^n (\pi_i - \pi_{i+1}) \cdot \log_2(i), \quad (2.40)$$

$$D(\pi) = \sum_{i=2}^n (\pi_i - \pi_{i+1}) \cdot \log_2\left(\frac{i}{\sum_{j=1}^i \pi_j}\right) \quad (2.41)$$

Klir a considéré le principe de la préservation de l'incertitude en utilisant deux échelles :

- l'échelle linéaire : les deux transformations résultantes : $p \rightarrow \pi$ et $\pi \rightarrow p$ sont respectivement définies par :

$$\pi_i = \frac{p_i}{p_1} \quad (2.42)$$

$$p_i = \frac{\pi_i}{n \cdot \sum_{i=1}^n \pi_i} \quad (2.43)$$

Ces deux transformations ne sont pas assez souple pour vérifier le principe de la préservation de l'incertitude. La transformation $p \rightarrow \pi$ définie par (2.42) est appelée la transformation normalisée.

- l'échelle logarithmique : les deux transformations définies par Klir sont :

$$\pi_i = \left(\frac{p_i}{p_1} \right)^\alpha \quad (2.44)$$

$$p_i = \frac{(\pi_i)^{1/\alpha}}{\sum_{j=1}^n (\pi_j)^{1/\alpha}} \quad (2.45)$$

Ces deux transformations vérifient le principe de la préservation de l'incertitude [KLI93]. En résolvant (2.38), on peut trouver la valeur de α qui est comprise dans l'intervalle]0, 1].

2.4.3. Principe de cohérence de Dubois et Prade

Le principe de cohérence proposé par Dubois et Prade est le suivant : soit p une distribution de probabilité dont P est la mesure, toute distribution de possibilité π dont la mesure Π satisfait :

$$\forall A \subseteq \Omega, \Pi(A) \geq P(A) \quad (2.46)$$

est cohérente avec la distribution de probabilité p .

Ce principe est basé sur le fait que la représentation possibiliste est moins informative que la représentation probabiliste. Cela est dû au fait que le modèle probabiliste donne une seule valeur pour représenter l'incertitude, tandis que le modèle possibiliste donne un intervalle pour représenter à la fois l'incertitude et l'imprécision [DUB93]. Lors d'une transformation $p \rightarrow \pi$, on perd de l'information, alors qu'une transformation $\pi \rightarrow p$ ajoute arbitrairement de l'information. On peut dire que toute distribution de possibilité est un majorant de la

distribution de probabilité. Pour une transformation $\pi \rightarrow p$, on utilise le principe de la raison insuffisante qui consiste à choisir la distribution de probabilité p qui contient le minimum d'information ou le maximum d'incertitude.

Les deux transformations $p \rightarrow \pi$ et $\pi \rightarrow p$ sont définies respectivement par :

$$\pi_i = \sum_{j=i}^n p_j \quad (2.47)$$

$$p_i = \sum_{j=i}^n \frac{(\pi_j - \pi_{j+1})}{j} \quad (2.48)$$

Les deux transformations définies par (2.47) et (2.48) ne sont pas réciproques, on appelle donc la transformation $p \rightarrow \pi$, définie par (2.47), la transformation asymétrique de Dubois et Prade. Dubois et Prade ont défini une autre transformation $p \rightarrow \pi$ symétrique satisfaisant le principe de cohérence de Dubois et Prade définie par [DUB83] :

$$\pi_i = \sum_{j=1}^n \min(p_i, p_j) \quad (2.49)$$

2.4.4. Transformation Variable (TV)

Nous proposons la transformation de probabilité en possibilité :

$$\pi_i = \left(\frac{p_i}{p_1} \right)^{k \cdot (1-p_i)} \quad (2.50)$$

que nous avons appelé Transformation Variable (TV) parce qu'elle change sa spécificité quand la valeur de k change [MOU02]. Elle vérifie la condition de cohérence de Dubois et Prade définie par (2.46) si k appartient à l'intervalle suivant :

$$0 \leq k \leq \frac{\log p_n}{(1-p_n) \cdot \log \frac{p_n}{p_1}} \quad (2.51)$$

proposition 1

La TV vérifie la condition de cohérence de Dubois et Prade définie par (2.46) si la valeur de k vérifie l'inégalité définie par (2.51).

Preuve

Soit la distribution de probabilité $p = (p_1, p_2, \dots, p_n, p_{n+1})$ et la distribution de possibilité $\pi = (\pi_1, \pi_2, \dots, \pi_n, \pi_{n+1})$ correspondante selon la TV définie par (2.50). Ces deux distributions sont ordonnées comme suit :

$$p_1 \geq p_2 \geq \dots \geq p_n \geq p_{n+1} = 0, \quad \pi_1 \geq \pi_2 \geq \dots \geq \pi_n \geq \pi_{n+1} = 0.$$

Vérifier (2.46) demande la vérification de toutes les inégalités suivantes :

$$\forall i = 1..n, \pi_i = \left(\frac{p_i}{p_1} \right)^{k(1-p_i)} \geq p_i \quad (2.52)$$

Pour $i = 1$, $\pi_1 = 1 \geq p_1$ est vérifiée. Il faut donc vérifier (2.52) pour $i = 2 \dots n$, cela nécessite que k respecte les intervalles suivants :

$$0 \leq k \leq \frac{\log p_i}{(1-p_i) \cdot \log \frac{p_i}{p_1}}, \quad \forall i = 2 \dots n \quad (2.53)$$

Nous avons : $p_i \geq p_{i+1} \Rightarrow \log p_i \geq \log p_{i+1}$, $\forall i = 1 \dots n$. Puisque $\log p_i \leq 0$, $\forall i = 1 \dots n$, nous pouvons écrire :

$$\log p_i = \log p_{i+1} + \beta, \quad \text{avec } \beta \geq 0 \quad (2.54)$$

De la même manière, nous avons :

$$\log \frac{p_i}{p_1} = \log \frac{p_{i+1}}{p_1} + \gamma, \quad \text{avec } \gamma \geq 0 \quad (2.55)$$

que nous pouvons réécrire :

$$\log p_i = \log p_{i+1} + \gamma \quad (2.56)$$

En comparant (2.54) et (2.56), nous trouvons :

$$\beta = \gamma \quad (2.57)$$

En développant l'inégalité suivante :

$$\frac{\log p_{i+1}}{\log \frac{p_{i+1}}{p_1}} \leq \frac{\log p_{i+1} + \beta}{\log \frac{p_{i+1}}{p_1} + \beta} \quad (2.58)$$

nous constatons qu'elle est vérifiée si et seulement si :

$$\log p_1 \leq 0 \quad (2.59)$$

cela est toujours vrai puisque $p_1 \leq 1$. Nous savons que $p_{i+1} \leq p_i$, ce qui donne :

$$(1 - p_{i+1}) \geq (1 - p_i) \quad (2.60)$$

De (2.54), (2.55), (2.57) et (2.58), nous aurons :

$$\frac{\log p_{i+1}}{\log \frac{p_{i+1}}{p_1}} \leq \frac{\log p_i}{\log \frac{p_i}{p_1}} \quad (2.61)$$

En utilisant (2.60), nous pouvons récrire (2.61) comme suit :

$$\frac{\log p_{i+1}}{(1 - p_{i+1}) \cdot \log \frac{p_{i+1}}{p_1}} \leq \frac{\log p_i}{(1 - p_i) \cdot \log \frac{p_i}{p_1}} \quad (2.62)$$

Nous constatons que pour que l'inégalité (2.53) soit vérifiée, il suffit qu'elle le soit pour $i = n$. k doit donc respecter l'intervalle défini par (2.51) pour que la TV vérifie la condition de cohérence de Dubois et Prade. En effet quand $k = k_{max} = \frac{\log p_n}{(1 - p_n) \cdot \log \frac{p_n}{p_1}}$, on a $\pi_n = p_n$.

La TV ajoute plus d'information pour les probabilités les plus élevées parce que la puissance $k \cdot (1 - p_i)$ est plus faible pour celles-ci.

La transformation de Klir diffère de la TV par la valeur de la puissance α qui est constante et est déterminée dans l'intervalle ouvert]0, 1] afin de préserver l'incertitude lors de la transformation probabilité-possibilité. En revanche, la puissance : $k \cdot (1 - p_i)$ de la TV est fonction de la valeur de probabilité afin de mettre en valeur les probabilités les plus élevées et en conséquence d'augmenter la spécificité.

2.5. Comparaison entre les différentes transformations

Pour comparer ces différentes transformations, nous allons utiliser les critères d'évaluations les plus courants. Ces critères définissent les propriétés et les qualités que peut posséder une transformation. Le choix de la transformation dépend du contexte du problème et donc des propriétés à vérifier.

2.5.1. Critère de normalisation

Les distributions de possibilité π ou de probabilité p sont dites normalisées si :

$$\exists w \in \Omega \text{ tel que } \pi_v(w) = 1 \quad (2.63)$$

$$\sum_{i=1}^n p_v(w_i) = 1 \quad (2.64)$$

Ces conditions signifient que l'univers Ω contient toutes les valeurs possibles de l'attribut ou de la variable v , on parle alors de monde fermé. Dans le cas des données incomplètes, l'hypothèse de monde fermé n'est pas satisfaite, on est dans le cas du monde ouvert.

Le critère de normalisation consiste à obtenir une distribution normalisée, lors de l'utilisation d'une transformation, quel que soit la distribution d'origine. Les transformations de Klir et normalisée satisfont le critère de normalisation. En revanche, les transformations symétrique et asymétrique de Dubois et Prade ne vérifient pas ce critère. En conséquence, dans le cas d'une distribution de probabilité sous-normalisée, $\sum_{i=1}^n p_v(w_i) < 1$, ces deux transformations donnent une distribution de possibilité sous normalisée.

proposition 2

La TV vérifie la condition de normalisation

Preuve

$$\text{En prenant } i = 1, \text{ nous trouvons : } \pi_1 = \pi_{max} = \left(\frac{p_1}{p_1} \right)^{k \cdot (1-p_1)} = 1.$$

Prenons l'exemple d'une distribution de probabilité sous normalisée $p = (0, 0.02, 0.03, 0.1, 0.15, 0.4, 0.1, 0.03, 0.02, 0)$ où $\sum_{i=1}^n p_i = 0.85$. La Figure 2.2 montre les distributions de

possibilité correspondantes en utilisant les différentes transformations.

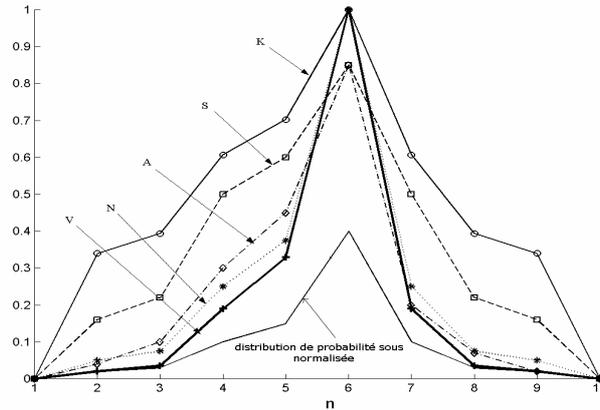


Figure 2.2. Distributions de possibilité correspondant à une distribution de probabilité sous normalisée en utilisant les transformations : Variable (V) avec $k = k_{\max}$, Normalisée (N), Asymétrique (A), Symétrique (S) et Klir (K)

Nous remarquons que seules les deux transformations symétrique et asymétrique de Dubois et Prade ne vérifient pas le critère de normalisation. On utilise une transformation vérifiant le critère de normalisation comme un opérateur de normalisation afin d'obtenir une distribution normalisée à partir d'une distribution sous-normalisée.

2.5.2. Critère de maximum de spécificité

La transformation la plus spécifique est celle pour laquelle la distribution de possibilité vérifie :

$$\exists u \in \Omega : \pi(u) = 1, \forall w \neq u \in \Omega : \pi(w) = 0 \quad (2.65)$$

La mesure de spécificité est évaluée en mesurant le degré pour lequel une distribution de possibilité π est proche de la distribution la plus spécifique définie par (2.65).

On dit qu'une théorie d'incertitude capture l'imprécision si elle permet de comparer les différentes distributions selon leur spécificité [SAN91]. La théorie des probabilités ne capture pas l'imprécision du fait que la condition de normalisation conduit à des distributions également spécifique dans le sens de (2.65). Soient π_1 et π_2 , deux distributions de possibilité, on dit que π_2 est plus spécifique que π_1 si et seulement si [DUB93, LAS99] :

$$\forall w \in \Omega, \pi_2(w) < \pi_1(w) \quad (2.66)$$

cette inégalité est appelée le principe de maximum de spécificité. On peut déduire de (2.66) que, entre deux distributions de possibilité, celle qui est la plus spécifique est celle qui est la

plus petite dans le sens de la cardinalité de l'ensemble flou, c'est à dire celle qui minimise

$\sum_{i=1}^n \pi(w_i)$. Le degré de spécificité peut donc être mesuré par :

$$S_p = \sum_{i=1}^n \pi(w_i) \quad (2.67)$$

proposition 3

Le degré de spécificité de la distribution de possibilité qui est la plus spécifique et qui vérifie le principe de cohérence de Dubois et Prade est :

$$S_{\max} = 1 + \sum_{i=2}^n p_i \quad (2.68)$$

Preuve

Pour une distribution de possibilité π vérifiant la condition de normalisation définie par (2.63), on a $\pi_1 = 1$. π vérifie la condition de cohérence de Dubois et Prade définie par (2.46) et en même temps elle est la plus spécifique dans le sens de degré de spécificité définie par (2.67) si : $p_i = \pi_i$, pour $i = 2 \dots n$, ce qui donne le degré de spécificité défini par (2.68).

Nous allons utiliser la différence entre le degré de spécificité S_p et le degré de maximum de spécificité vérifiant la condition de cohérence de Dubois et Prade, S_{\max} , pour comparer la spécificité des transformations précédentes. Prenons l'exemple d'une distribution de probabilité normalisée $p = (0, 0.02, 0.03, 0.1, 0.15, 0.4, 0.15, 0.1, 0.03, 0.02, 0)$. Le Tableau 2.2 montre les distributions de possibilité correspondantes et la différence $S_p - S_{\max}$ pour les transformations précédentes.

distribution de probabilité	Transformation				
	Variable avec $k = k_{\max}$	normalisée	asymétrique	symétrique	de Klir
0	0	0	0	0	0
0.02	0.02	0.05	0.04	0.18	0.46
0.03	0.03	0.07	0.1	0.25	0.51
0.1	0.19	0.25	0.3	0.6	0.7
0.15	0.33	0.37	0.6	0.75	0.77
0.4	1	1	1	1	1
0.15	0.33	0.37	0.45	0.75	0.77
0.1	0.19	0.25	0.2	0.6	0.7
0.03	0.03	0.07	0.07	0.25	0.51
0.02	0.02	0.05	0.02	0.18	0.46
0	0	0	0	0	0
$S_p - S_{\max}$	0.55	0.9	1.18	2.96	4.27

Tableau 2.2. Comparaison du principe du maximum de spécificité entre les transformations précédentes.

La Figure 2.3 montre ces différentes distributions de possibilité.

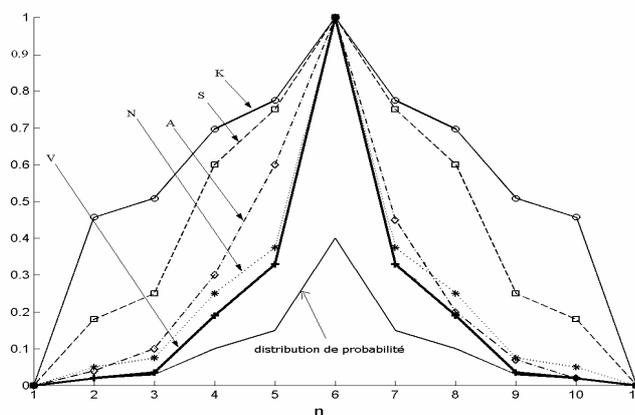


Figure 2.3. Comparaison de la spécificité entre les transformations : Variable (V) avec $k = k_{\max}$ Normalisée (N), Asymétrique (A), Symétrique (S) et Klir (K).

Nous constatons que :

- la transformation de Klir est la moins spécifique,
- la TV est la plus spécifique.

proposition 4

La transformation asymétrique est plus spécifique que la transformation normalisée si et seulement si :

$$p_1 < \frac{\sum_{i=2}^n p_i}{\sum_{j=2}^n (j-1).p_j} \quad (2.69)$$

Preuve

Si $p = (p_1, p_2, \dots, p_n)$ est une distribution de probabilité ordonnée dans le sens décroissant, alors la distribution de possibilité correspondante en utilisant la transformation asymétrique est :

$$\pi_1 = p_1 + p_2 + \dots + p_n = 1, \pi_2 = p_2 + p_3 + \dots + p_n, \pi_3 = p_3 + p_4 + \dots + p_n, \dots, \pi_n = p_n \quad (2.70)$$

Son degré de spécificité est :

$$S_p = 1 + p_2 + 2.p_3 + 3.p_4 + \dots + (n-1).p_n \quad (2.71)$$

La distribution de possibilité correspondante en utilisant la transformation normalisée est :

$$\pi'_1 = \frac{p_1}{p_1} = 1, \pi'_2 = \frac{p_2}{p_1}, \pi'_3 = \frac{p_3}{p_1}, \dots, \pi'_n = \frac{p_n}{p_1} \quad (2.72)$$

et son degré de spécificité est :

$$S'_p = 1 + \frac{p_2 + p_3 + \dots + p_n}{p_1} \quad (2.73)$$

La transformation asymétrique est plus spécifique que la transformation normalisée si l'inégalité suivante est vraie :

$$1 + p_2 + 2.p_3 + 3.p_4 + \dots + (n-1).p_n < 1 + \frac{p_2 + p_3 + \dots + p_n}{p_1} \Rightarrow$$

$$p_1 < \frac{p_2 + p_3 + \dots + p_n}{p_2 + 2.p_3 + \dots + (n-1).p_n} \Rightarrow p_1 < \frac{\sum_{i=2}^n p_i}{\sum_{j=2}^n (j-1).p_j}$$

Pour l'exemple du Tableau 2.2, $p_1 = 0.4$ et $\frac{\sum_{i=2}^n p_i}{\sum_{j=2}^n (j-1) \cdot p_j} = 0.34$, l'inégalité définie par

(2.69) n'est pas satisfaite, donc la transformation normalisée est plus spécifique que la transformation asymétrique de Dubois et Prade.

proposition 5

La TV est plus spécifique, en chaque élément de l'univers Ω , que la transformation normalisée si et seulement si :

$$k > \frac{1}{1-p_2} \quad (2.74)$$

Preuve

Si π est la distribution de possibilité calculée en utilisant la transformation normalisée et π' est la distribution de possibilité calculée en utilisant la TV, on a pour $i = 1$:

$$\pi_1 = \frac{p_1}{p_1} = 1 = \pi'_1 = \left(\frac{p_1}{p_1}\right)^{k(1-p_1)}. \text{ Pour } i = 2 \dots n, \text{ l'inégalité } \frac{p_i}{p_1} > \left(\frac{p_i}{p_1}\right)^{k(1-p_i)}$$
 est satisfaite si

$k(1-p_i) > 1$ qui donne $k > \frac{1}{1-p_i}$. Puisque p_2 est la probabilité la plus élevée pour $i = 2 \dots n$,

nous devons avoir $k > \frac{1}{1-p_2}$ afin que la TV vérifie l'inégalité de maximum de spécificité

définie par (2.66).

Pour l'exemple du Tableau 2.2, $k = k_{\max} = 1.33 > \frac{1}{1-p_2} = \frac{1}{1-0.15} = 1.18$, ce qui fait que la

TV est plus spécifique que la transformation normalisée.

proposition 6

La transformation variable est plus spécifique, en chaque élément de l'univers Ω , que la transformation asymétrique si et seulement si :

$$\frac{\log(p_n)}{(1-p_n) \cdot \log\left(\frac{p_n}{p_1}\right)} > \max_{j=2 \dots n-1} \left(\frac{\log\left(\sum_{i=j}^n p_i\right)}{(1-p_j) \cdot \log\left(\frac{p_j}{p_1}\right)} \right) \quad (2.75)$$

Preuve

Si on a une distribution de probabilité $p = (p_1, p_2, \dots, p_n)$, ordonnée dans le sens décroissant, et si $\pi = (\pi_1, \pi_2, \dots, \pi_n)$ et $\pi' = (\pi'_1, \pi'_2, \dots, \pi'_n)$ sont, respectivement, les distributions de possibilités correspondantes selon la transformation asymétrique et la TV, alors on a :

$$\pi_1 = 1, \pi_2 = \sum_{i=2}^n p_i, \pi_3 = \sum_{i=3}^n p_i, \dots, \pi_n = p_n$$

$$\pi'_1 = 1, \pi'_2 = \left(\frac{p_2}{p_1}\right)^{k \cdot (1-p_2)}, \pi'_3 = \left(\frac{p_3}{p_1}\right)^{k \cdot (1-p_3)}, \dots, \pi'_n = p_n$$

Nous avons choisi $k = k_{\max}$ calculé en utilisant (2.51) pour que la TV soit la plus spécifique possible en vérifiant toujours la condition de cohérence de Dubois et Prade. La TV est plus spécifique que la transformation asymétrique pour $i = 2 \dots n$, si :

$$\begin{aligned} \pi'_2 < \pi_2, \pi'_3 < \pi_3, \dots, \pi'_{n-1} < \pi_{n-1} &\Rightarrow \\ \left(\frac{p_2}{p_1}\right)^{k_{\max} \cdot (1-p_2)} < \sum_{i=2}^n p_i, \left(\frac{p_3}{p_1}\right)^{k_{\max} \cdot (1-p_3)} < \sum_{i=3}^n p_i, \dots, \left(\frac{p_{n-1}}{p_1}\right)^{k_{\max} \cdot (1-p_{n-1})} < p_{n-1} + p_n &\Rightarrow \\ k_{\max} > \frac{\log(\sum_{i=2}^n p_i)}{(1-p_2) \cdot \log\left(\frac{p_2}{p_1}\right)}, k_{\max} > \frac{\log(\sum_{i=3}^n p_i)}{(1-p_3) \cdot \log\left(\frac{p_3}{p_1}\right)}, \dots, k_{\max} > \frac{\log(p_{n-1} + p_n)}{(1-p_{n-1}) \cdot \log\left(\frac{p_{n-1}}{p_1}\right)} &\Rightarrow \\ \frac{\log(p_n)}{(1-p_n) \cdot \log\left(\frac{p_n}{p_1}\right)} > \max\left(\frac{\log(\sum_{i=2}^n p_i)}{(1-p_2) \cdot \log\left(\frac{p_2}{p_1}\right)}, \frac{\log(\sum_{i=3}^n p_i)}{(1-p_3) \cdot \log\left(\frac{p_3}{p_1}\right)}, \dots, \frac{\log(\sum_{i=n-1}^n p_i)}{(1-p_{n-1}) \cdot \log\left(\frac{p_{n-1}}{p_1}\right)}\right) \end{aligned}$$

Pour l'exemple du Tableau 2.2, $k_{\max} = 1.33 > \max(0.61, 0.95, 0.96, 1.29, 0.92, 1.06, 1.1)$, donc la TV est plus spécifique que la transformation asymétrique en chaque élément de l'univers Ω .

La spécificité de la TV varie en fonction de la valeur de k . Plus k augmente, plus le degré de spécificité augmente. Dans le Tableau 2.3, nous comparons le degré de cohérence de Zadeh défini par (2.36) avec la différence $S_p - S_{\max}$ pour l'exemple du Tableau 2.2. Nous constatons que :

- la transformation la plus spécifique est celle qui a le degré de cohérence de Zadeh le plus petit,
- la transformation de Klir possède le degré de cohérence de Zadeh le plus élevé et la TV possède le plus petit,
- plus le degré de cohérence de Zadeh augmente, plus la spécificité de la transformation diminue.

	Transformation				
	Variable $p \rightarrow \pi$	normalisée $p \rightarrow \pi$	asymétrique $p \rightarrow \pi$	symétrique $p \rightarrow \pi$	de Klir $p \rightarrow \pi$
$S_p - S_{\max}$	0.55	0.9	1.18	2.96	4.27
c_Z	0.54	0.57	0.61	0.77	0.82

Tableau 2.3. Comparaison entre le principe du maximum de spécificité et le degré de cohérence de Zadeh c_Z

2.5.3. Critère de conservation de l'incertitude

Une transformation probabilité-possibilité doit conserver l'information contenue dans la distribution lors de la transformation. Ce concept d'informativité est lié intuitivement au concept de l'incertitude. Cela veut dire que plus une distribution contient d'information, moins elle contient d'incertitude. Une transformation probabilité-possibilité doit être telle que les deux distributions contiennent la même quantité d'incertitude. Ce lien est appelé le principe de conservation de l'incertitude. Klir [KLI92, KLI93] s'est basé sur ce principe pour construire sa transformation définie par (2.44) et (2.45).

Nous évaluons ce critère en calculant la différence Δ entre l'incertitude $Un(\pi)$ contenue dans une distribution de possibilité π et l'incertitude $H(p)$ contenue dans une distribution de probabilité p . Le Tableau 2.4 compare Δ avec la différence $S_p - S_{\max}$ pour l'exemple du Tableau 2.2. Nous constatons que :

- la transformation de Klir est la seule transformation qui vérifie le critère de conservation de l'incertitude. Toutes les autres transformations gagnent de l'information ou autrement dit perdent de l'incertitude dans le sens $p \rightarrow \pi$,
- la transformation la plus spécifique est la transformation qui gagne le maximum d'information. C'est le cas de la TV pour cet exemple.

Transformation	Variable $p \rightarrow \pi$	normalisée $p \rightarrow \pi$	asymétrique $p \rightarrow \pi$	symétrique $p \rightarrow \pi$	de Klir $p \rightarrow \pi$
$S_p - S_{\max}$	0.55	0.9	1.18	2.96	4.27
$\Delta = \text{Un}(\pi) - H(p)$	- 1.4	- 1.28	- 0.99	- 0.30	0

Tableau 2.4. Comparaison entre le principe de maximum de spécificité et le principe de conservation de l'incertitude

Prenons maintenant l'exemple du petit déjeuner de Hans, présenté au Tableau 2.1, qui définit une distribution de possibilité obtenue pour avoir le degré de cohérence de Zadeh égal à 1. Nous comparons dans le Tableau 2.5, le principe de maximum de spécificité, le principe de conservation de l'incertitude et le degré de cohérence de Zadeh.

p	Transformation					distribution de possibilité de Zadeh
	Variable	normalisée	asymétrique	symétrique	de Klir	
0.1	0.1	0.12	0.2	0.3	0.39	1
0.8	1	1	1	1	1	1
0.1	0.1	0.12	0.1	0.3	0.39	1
0	0	0	0	0	0	1
0	0	0	0	0	0	0.8
0	0	0	0	0	0	0.6
0	0	0	0	0	0	0.4
0	0	0	0	0	0	0.2
$S_p - S_{\max}$	0	0.05	0.1	0.4	0.59	4.8
$\Delta =$ $\text{Un}(\pi) - H(p)$	- 0.63	- 0.56	- 0.47	- 0.12	0	1.12
c_Z	0.82	0.82	0.83	0.86	0.88	1

Tableau 2.5. Comparaison entre le principe du maximum de spécificité, le principe de conservation de l'incertitude et le degré de cohérence de Zadeh pour les transformations précédentes et pour l'exemple du Tableau 2.1

Nous constatons que :

- la transformation de Klir qui satisfait le principe de conservation de l'incertitude a le degré de cohérence de Zadeh $c_Z = 0.88$,
- quand c_Z passe au dessus de la valeur précédente, la transformation $p \rightarrow \pi$ perd de l'information et devient donc moins spécifique. La distribution de possibilité de Zadeh obtenue pour avoir $c_Z = 1$ est donc la moins spécifique,
- quand c_Z diminue en dessous de 0.88, la transformation gagne de l'information et devient donc plus spécifique. Nous remarquons que la TV est la transformation la plus spécifique avec un degré de spécificité S_p égal au degré maximum de spécificité S_{\max} .

Pour la transformation de possibilité en probabilité, on utilise le principe de la raison insuffisante. Dans le Tableau 2.6, nous comparons le principe de conservation de l'incertitude entre les transformations $\pi \rightarrow p$ précédentes, pour l'exemple du petit déjeuner de Hans. Nous pouvons remarquer que la distribution de probabilité de Zadeh ne satisfait pas le principe de la raison insuffisante puisqu'elle gagne de l'information. En revanche, la transformation normalisée $\pi \rightarrow p$ apporte le maximum d'incertitude par rapport aux autres transformations. En effet, la transformation normalisée dans le sens $p \rightarrow \pi$ gagne plus d'information que les transformations symétrique et asymétrique de Dubois et Prade pour l'exemple de Zadeh (cf Tableau 2.5). Dans le sens $\pi \rightarrow p$, la transformation normalisée apporte plus d'incertitude que celle de Dubois et Prade qui nous permet de constater que les transformations apportant le plus d'information dans un sens, perdent le plus d'information dans l'autre sens comme le montre le Tableau 2.6.

Transformation $\pi \rightarrow p$				
π	Klir	Dubois et Prade	normalisée	Zadeh
1	0.25	0.15	0.17	0.1
1	0.25	0.15	0.17	0.8
1	0.25	0.15	0.17	0.1
1	0.25	0.15	0.17	0
0.8	0.006	0.15	0.13	0
0.6	0	0.10	0.1	0
0.4	0	0.062	0.067	0
0.2	0	0.029	0.033	0
$\Delta = H(p) - \text{Un}(\pi)$	0	0.34	0.67	- 1.12

Tableau 2.6. Comparaison du principe de conservation de l'incertitude entre les précédentes transformations

2.5.4. Critère de conservation de la préférence

Ce critère préserve l'ordre des éléments d'une distribution lors de la transformation. Cela veut dire que lors de la transformation $p \rightarrow \pi$, l'élément le plus probable devient l'élément le plus possible et dans l'autre sens, l'élément le plus possible devient le plus probable.

Il existe deux types de conservation de la préférence : la conservation faible définie par (2.76) et la conservation forte définie par (2.77) :

$$\forall w, u \in \Omega, p(w) > p(u) \Leftrightarrow \pi(w) > \pi(u) \quad (2.76)$$

$$\left\{ \begin{array}{l} \forall w, u \in \Omega : \\ p(w) = p(u) \Leftrightarrow \pi(w) = \pi(u), \\ p(w) > p(u) \Leftrightarrow \pi(w) > \pi(u), \\ p(w) < p(u) \Leftrightarrow \pi(w) < \pi(u) \end{array} \right. \quad (2.77)$$

Les transformations symétrique, de Klir et normalisée vérifient la conservation forte de la préférence [OUS00]. En revanche, la transformation asymétrique vérifie seulement la conservation faible [OUS00]. Les transformations vérifiant le principe de cohérence de Zadeh, $c_Z = 1$, ne vérifient pas le critère de conservation faible et forte de la préférence. Cela est logique puisque le principe de cohérence de Zadeh définit une relation faible et heuristique entre les représentations probabiliste et possibiliste.

proposition 7

La TV vérifie le critère de conservation forte de la préférence

Preuve

$\forall i = 1 \dots n :$

$$p_i = p_{i+1} \Rightarrow \left(\frac{p_i}{p_1} \right)^{k \cdot (1-p_i)} = \left(\frac{p_{i+1}}{p_1} \right)^{k \cdot (1-p_{i+1})} \Rightarrow \pi_i = \pi_{i+1}$$

$$p_i < p_{i+1} \Rightarrow k \cdot (1-p_i) > k \cdot (1-p_{i+1}) \Rightarrow \left(\frac{p_i}{p_1} \right)^{k \cdot (1-p_i)} < \left(\frac{p_{i+1}}{p_1} \right)^{k \cdot (1-p_{i+1})} \Rightarrow \pi_i < \pi_{i+1}$$

$$p_i > p_{i+1} \Rightarrow k \cdot (1-p_i) < k \cdot (1-p_{i+1}) \Rightarrow \left(\frac{p_i}{p_1} \right)^{k \cdot (1-p_i)} > \left(\frac{p_{i+1}}{p_1} \right)^{k \cdot (1-p_{i+1})} \Rightarrow \pi_i > \pi_{i+1}$$

2.5.5. Critère de conservation de la forme

Ce critère permet de garantir que la distribution de possibilité obtenue a la même forme que la distribution de probabilité de départ. La vérification du critère de conservation forte de la préférence conduit à la vérification du critère de conservation de la forme. Pour cela, seule la transformation asymétrique ne garantit pas la conservation de la forme d'une distribution de probabilité, elle donne pour les éléments ayant la même valeur de probabilité, différentes valeurs de possibilité ce qui conduit à un changement de la forme entre les deux distributions.

2.5.6. Critère de conservation de l'ignorance

Ce critère permet de garantir la conservation du cas de l'ignorance totale lors de la transformation. Le cas de l'ignorance totale est indésirable parce qu'il correspond au cas où on ne peut pas prendre de décision.

Nous pouvons montrer simplement que les transformations : de Klir, normalisée, symétrique et variable conservent l'ignorance totale lors de la transformation. En revanche, la transformation asymétrique ne conserve pas l'ignorance totale puisqu'elle ne vérifie pas le critère de conservation forte de la préférence. En effet, l'ignorance totale correspond à des éléments ayant la même valeur de probabilité qui sont transformés par la transformation asymétrique en des valeurs différentes de possibilité.

Dans un cas proche de l'ignorance totale, nous souhaitons avoir une transformation qui donne une distribution de possibilité permettant de distinguer le mieux possible les éléments ayant des probabilités très voisines. Pour voir le comportement des transformations précédentes devant un cas très proche de l'ignorance totale, nous prenons l'exemple de la distribution de probabilité : $p = (0.16, 0.16, 0.16, 0.16, 0.2)$. Les deux derniers éléments ont des probabilités très proches ce qui conduit à un cas de confusion. La Figure 2.4 montre les distributions de possibilité résultantes, nous remarquons que :

- pour la transformation de Klir, on n'a pas pu trouver la valeur de α qui vérifie le critère de conservation de l'incertitude,
- la TV, avec $k = k_{max}$, fournit la distribution qui donne la meilleur distinction entre les deux éléments confondus.

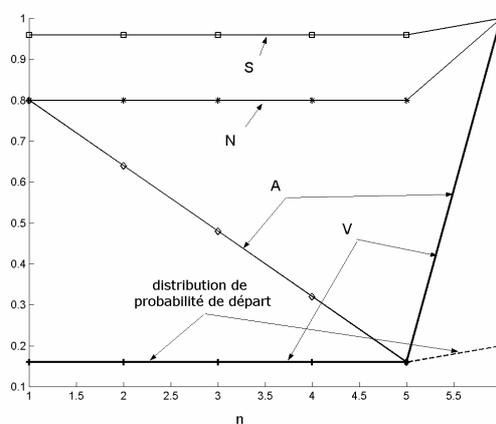


Figure 2.4. Comparaison du comportement des transformations précédentes devant un cas très proche de l'ignorance totale : Variable (V) avec $k = k_{max}$, Normalisée (N), Asymétrique (A) et Symétrique (S)

2.6. Conclusion

Dans les systèmes de diagnostic par RdF, la précision des informations, avec lesquelles on est amené à travailler, ainsi que leur fiabilité sont limitées. La justesse de la décision prise dépend de la prise en compte de l'incertitude et de l'imprécision des informations utilisées. Cela exige des méthodes très performantes permettant la modélisation et la manipulation des imperfections sans perdre d'information.

La théorie des probabilités est le modèle le plus traditionnel pour la modélisation de l'information imparfaite. Une caractéristique de la théorie des probabilités est l'axiome d'additivité qui conduit à la conséquence suivante : si la probabilité d'un événement est connue alors la probabilité de l'événement contraire est exactement déterminée. La théorie des probabilités semble donc peu adaptée à des situations où les connaissances d'un événement et de son contraire sont très limitées. De plus, la théorie des probabilités est trop rigide pour exprimer le cas de l'ignorance totale parce qu'elle modélise ce cas par un ensemble d'événements mutuellement disjoints et équiprobables.

La théorie des possibilités a été introduite par Zadeh comme un modèle plus souple et plus performant pour prendre en compte l'incertitude incluse dans des objets imprécis. Elle a un caractère d'ordre ou de préférence entre les événements. Cela signifie qu'un événement totalement possible n'interdit pas pour autant à l'événement contraire de l'être également. Cet effet montre une différence importante entre le modèle possibiliste et le modèle probabiliste qui a un caractère additif.

Dans ce chapitre, nous avons présenté les principes de cohérence, les plus connus, entre une distribution de probabilité et une distribution de possibilité. Ces principes sont : le principe de cohérence de Zadeh, le principe de cohérence de Klir, le principe de cohérence de Dubois et Prade. Puis, nous avons étudiés les transformations probabilités-possibilités les plus importantes de la littérature. Nous avons proposé une nouvelle transformation de probabilité en possibilité que nous avons appelée la Transformation Variable (TV). Les caractéristiques de cette transformation ont été comparées avec celles des autres transformations en utilisant des critères d'évaluation. Nous avons montré que la TV peut être la transformation la plus spécifique en distinguant le mieux les cas très voisins.

Dans le chapitre suivant, nous étudions une méthode de classification non supervisée et une méthode de classification supervisée qui utilisent l'approche possibiliste. Ces deux méthodes sont respectivement : la méthode Possibiliste des C Moyennes (PCM), et la méthode Fuzzy Pattern Matching (FPM). Ensuite nous nous concentrerons sur la méthode

FPM en proposant plusieurs approches afin d'améliorer ses performances.

3. Reconnaissance possibiliste des formes

3.1. Introduction

Les méthodes floues d'apprentissage non supervisé sont utilisées, dans un module de diagnostic par RdF, afin de chercher la structure des classes dans l'ensemble d'apprentissage. Le résultat est un ensemble d'apprentissage étiqueté, c'est-à-dire dont on connaît la classe de chaque observation. Cet ensemble est utilisé par une méthode floue de classification supervisée pour construire une fonction discriminante qui sert dans la phase d'exploitation à classer les nouvelles observations.

Nous avons vu, auparavant, que les méthodes floues de classification, supervisée ou non supervisée, souffrent du problème dû à l'application de la condition d'orthogonalité qui est imposée par l'hypothèse probabiliste. Dans le chapitre précédent nous avons présenté une solution à ce problème qui consiste à utiliser la théorie des possibilités. Cette dernière relaxe la condition d'orthogonalité et en conséquence elle est plus apte à modéliser les données provenant des sources d'informations faibles.

Dans ce chapitre, nous étudions une méthode d'apprentissage non supervisé : la méthode Possibiliste des C Moyennes (PCM), et une méthode supervisée : la méthode du Fuzzy Pattern Matching (FPM). Ces deux méthodes utilisent l'approche possibiliste. Nous apporterons, par la suite, des améliorations à la méthode FPM.

3.2. Méthode Possibiliste des C Moyennes

La méthode Possibiliste des C Moyennes (PCM) a été proposée par Krishnapuram en 1993 comme une solution au problème des points isolés [KRI93]. Elle substitue l'hypothèse probabiliste par une hypothèse possibiliste et en conséquence la condition d'orthogonalité n'existe plus. PCM est basée sur les hypothèses suivantes :

$$\forall j, k : u_{jk} \in [0, 1],$$

$$\max_{j=1..c} (u_{jk}) > 0,$$

$$0 < \sum_{k=1}^N u_{jk} < N$$

u_{jk} représentent les possibilités d'appartenance de l'observation k à la classe j . Le nouveau

critère à minimiser est :

$$J(U, G) = \sum_{k=1}^N \sum_{j=1}^c (u_{jk})^m \cdot (d_{jk})^2 + \sum_{j=1}^c \eta_j \cdot \sum_{k=1}^N (1 - u_{ji})^m$$

d_{jk} est la distance entre l'observation k et le centre de gravité de la classe j . La valeur de η_j fixe la distance à partir de laquelle la possibilité d'appartenance est égale à 0.5. Krishnapuram propose de prendre une valeur proportionnelle à la distance moyenne intraclasse :

$$\eta_j = \frac{\sum_{k=1}^N (u_{jk})^m \cdot (d_{jk})^2}{\sum_{k=1}^N (u_{jk})^m}$$

L'algorithme est initialisé avec la matrice de partition obtenue par la méthode FCM. Les centroïdes et les possibilités d'appartenance sont calculés de manière itérative par les expressions :

$$v_j = \frac{\sum_{k=1}^N (u_{jk})^m \cdot x_k}{\sum_{k=1}^N (u_{jk})^m},$$

$$u_{jk} = \left[1 + \left(\frac{(d_{jk})^2}{\eta_j} \right)^{\frac{1}{m-1}} \right]^{-1}$$

Le calcul se termine lorsque la différence entre les possibilités d'appartenance de deux itérations successives est inférieur à un seuil ε fixé.

La Figure 3.1 présente les valeurs de possibilités d'appartenance obtenues pour le nuage test étudié dans le premier chapitre. Les valeurs de m et ε sont respectivement fixées à 2 et 0.001. Les résultats font bien apparaître que la somme des possibilités d'appartenance d'un point n'est pas unitaire. Les points des 2 classes sont correctement classés. Le point B possède des possibilités d'appartenance faibles pour les deux classes, ce qui prouve bien qu'il en est éloigné. Un tel point peut donc être facilement rejeté selon ses valeurs de possibilité d'appartenance.

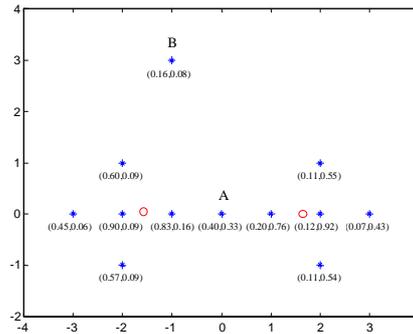


Figure 3.1. Les étoiles représentent les points du nuage. Les cercles correspondent aux centroïdes des 2 classes. Les valeurs entre parenthèses sont les possibilités d'appartenance à chaque classe.

PCM permet de résoudre le problème des points éloignés des classes mais il ne solutionne ni le problème des classes de forme complexe, ni celui de la détermination du nombre de classes dans un nuage de points. Nous n'approfondirons pas la description des méthodes d'apprentissage non supervisé car elles sortent du cadre de cette thèse.

3.3. Fuzzy Pattern Matching (FPM)

Fuzzy Pattern Matching (FPM) est une méthode de classification supervisée multi-critères [DUB88, GRA92, DEV99]. Elle utilise la théorie des ensembles flous et la théorie des possibilités pour prendre en compte à la fois l'imprécision et l'incertitude contenues dans les données [DUB88]. Cette méthode utilise c prototypes décrits par a attributs. Chaque prototype est représenté par une collection de a sous-ensembles flous qui expriment la plage des valeurs typiques de chaque attribut par rapport à chaque classe. Ces prototypes sont les densités de possibilité déduites par une linéarisation des distributions de possibilité. Ces dernières sont construites à partir des distributions de probabilité en utilisant une transformation probabilité-possibilité. Les distributions de probabilité sont calculées à partir de la répartition des points de l'ensemble d'apprentissage.

Nous supposons que la connaissance a priori, ou l'ensemble d'apprentissage, d'un système comporte N points, décrits par a attributs et distribués dans c classes. Le fonctionnement de FPM passe par deux phases : la phase d'apprentissage et la phase de classification. Nous allons expliquer ces deux phases en utilisant l'exemple de la Figure 3.2, correspondant à un ensemble d'apprentissage de 320 points répartis dans deux classes, représenté par deux attributs.

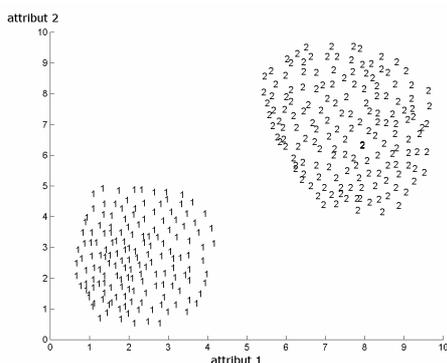


Figure 3.2. Exemple de nuage de points comportant deux classes.

3.3.1. Phase d'apprentissage

La phase d'apprentissage est nécessaire pour créer la fonction d'appartenance qui permet d'affecter les nouveaux points dans les différentes classes. Cette fonction d'appartenance est représentée par une densité de possibilité pour chaque attribut de chaque classe. L'apprentissage passe par les différentes étapes décrites ci-après.

3.3.1.1. Construction des histogrammes de données

Ces histogrammes expliquent la distribution des points de l'ensemble d'apprentissage dans les différentes classes, pour chaque attribut. Les bornes inférieure et supérieure de l'histogramme d'une classe, suivant un attribut, sont généralement les valeurs minimale et maximale des composantes des points d'apprentissage pour cet attribut. Toutefois, il arrive que ces bornes soient déterminées manuellement par l'utilisateur. L'intérêt est d'élargir la base des histogrammes de manière à agrandir le champ d'action du classifieur.

La Figure 3.3 montre les histogrammes pour chaque attribut de chaque classe. Le nombre de barres h influence les performances de FPM. Nous avons fixé expérimentalement h à 20.

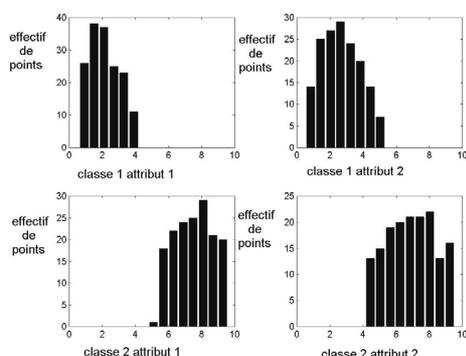


Figure 3.3. Histogrammes de données pour chaque attribut de chaque classe.

3.3.1.2. Histogrammes de probabilité

La valeur de probabilité $p_j^i(b_\alpha)$ de la barre b_α , pour l'attribut i de la classe j , est le nombre n_α de points d'apprentissage existants dans la barre, divisé par le nombre total N_j de points dans la classe :

$$p_j^i(b_\alpha) = \frac{n_\alpha}{N_j} \quad (3.1)$$

On ajoute deux barres auxquelles on associe une probabilité nulle, une au début et une à la fin de chaque histogramme. En attribuant chaque probabilité $p_j^i(b_\alpha)$ au centre de la barre α , nous obtenons une distribution de probabilité $p_j^i = (p_j^i(b_\alpha), \alpha = 1 \dots h + 2)$ pour chaque attribut i de chaque classe j .

3.3.1.3. Histogrammes de possibilité

Les histogrammes de possibilité sont déduits des histogrammes de probabilité par une transformation probabilité-possibilité. Pour l'exemple de la Figure 3.2, nous avons choisi la transformation symétrique de Dubois et Prade, pour sa simplicité et les bons résultats qu'elle donne. Les distributions de possibilité $(\pi_j^i(b_\alpha), j = 1 \dots c, i = 1 \dots a, \alpha = 1 \dots h + 2)$ sont construites de la même manière que celles de probabilité.

3.3.1.4. Densités de possibilité

Les densités de possibilité Π_j^i sont déduites des distributions de possibilité π_j^i par une linéarisation. Ces densités sont utilisées dans la phase de classification pour trouver la classe d'un nouveau point. La Figure 3.4 montre ces densités pour l'exemple de la Figure 3.2.

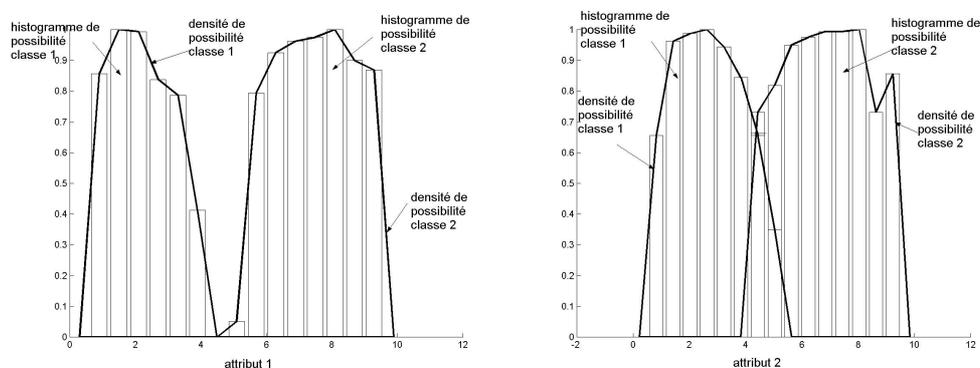


Figure 3.4. Densités de possibilité pour chaque attribut de chaque classe, pour l'exemple de la Figure 3.2.

3.3.2. Phase de classification

Le classement d'un nouveau point \underline{x} dont les valeurs sur les différents attributs sont (x^1, x^2, \dots, x^a) s'effectue en deux étapes :

- la détermination des degrés de similitude π_j^i entre x^i et chaque classe j par rapport à chaque attribut i . Ces degrés de similitude sont obtenus par interpolation linéaire sur les densités de possibilité Π_j^i ,
- la fusion de tous les degrés de similitude $(\pi_j^1, \dots, \pi_j^a)$, de la classe j , par un opérateur H . Cet opérateur peut être un minimum, un produit, une intégrale floue ...[GRA92] :

$$\mu_j(\underline{x}) = H[\pi_j^1, \dots, \pi_j^a] \quad (3.2)$$

Le résultat $\mu_j(\underline{x})$ de cette fusion représente la possibilité d'appartenance du point \underline{x} à la classe j . Le point est affecté à la classe pour laquelle il a la plus forte possibilité d'appartenance.

3.4. Opérateur d'agrégation

L'agrégation floue intervient dans le domaine de la décision multi-critère. Elle permet de combiner les données fournies par des capteurs ou des sources d'information. Cette combinaison doit être adaptée à la situation conflictuelle ou non des données. Il existe trois grands modes d'agrégation qui sont la conjonction ou t-conorme, le compromis et la disjonction ou t-norme.

Les opérateurs de type t-conormes sont des opérateurs d'union des sous ensembles flous. Ils ont un caractère compensatoire au sens où ils acceptent qu'une partie des critères soit vérifiée. Ils correspondent à un décideur ayant un comportement tolérant. La Figure 3.5 illustrant la position des t-conormes, montre que le maximum est la moins tolérante d'entre elles. Les opérateurs les plus utilisés de cette famille sont présentés dans le Tableau 3.1.

Les opérateurs de type t-norme sont des opérateurs d'intersection. Ils ont un caractère non tolérant, ceci implique que le résultat de l'agrégation dépend fortement de la plus faible des valeurs à agréger. Le minimum est la plus tolérante des t-normes comme l'illustre la Figure 3.5. Les opérateurs les plus répandus sont présentés dans le Tableau 3.1.

On appelle opérateur de compromis toute fonction d'agrégation se situant entre l'opérateur minimum et l'opérateur maximum, afin de donner un comportement intermédiaire entre la

disjonction et la conjonction. Le Tableau 3.1 montre quelques exemples.

t-normes	t-conormes	compromis
$\min(x,y)$	$\max(x,y)$	$\frac{2.x.y}{x+y}$
$x.y$	$x+y-x.y$	$\sqrt{x.y}$
$\min(0,x+y-1)$	$\min(1,x+y)$	$\frac{x+y}{2}$
$\frac{x.y}{\gamma+(1-\gamma).(x+y-xy)}$	$\frac{x+y-x.y-(1-\gamma).x.y}{1-(1-\gamma).x.y}$	
$1 - \min \left[(1-x)^p + (1-y)^p, 1 \right]$	$\min \left[(x^p + y^p)^{\frac{1}{p}}, 1 \right]$	
$\min(x,y)$ pour $x = 1$ ou $y = 1$ 0 ailleurs	$\max(x,y)$ pour $x.y = 0$ 1 ailleurs	

Tableau 3.1. Exemples de t-normes, t-conormes et opérateurs de compromis.

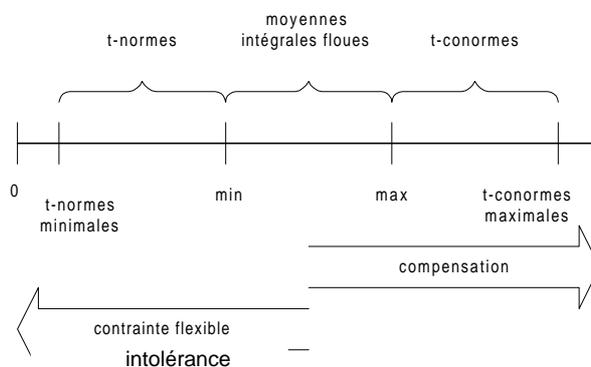


Figure 3.5. Représentation des opérateurs.

3.5. Agrégation et FPM

Compte tenu des objectifs des méthodes de classification floues, nous pouvons définir les propriétés que doit avoir l'opérateur H . Il doit respecter les propriétés mathématiques générales communes aux t-normes et aux t-conormes : commutativité, associativité, croissance et monotonie. Etant donné que la méthode FPM travaille dans l'intervalle $[0, 1]$ et qu'elle a un caractère possibiliste, l'opérateur doit avoir les propriétés suivantes :

$$H(0,0,0,0,\dots,0) = 0 \text{ et } H(1,1,1,1,\dots,1) = 1$$

Pour établir les propriétés supplémentaires de l'opérateur à utiliser, il faut revenir au principe de la méthode. Nous avons vu que la seule information que nous possédions sur les classes, à l'issue de l'apprentissage, était la projection possibiliste des points sur les attributs. En phase de classification, un point ne peut pas appartenir à une classe pour laquelle il a une valeur de possibilité d'appartenance nulle pour au moins un des attributs. Il est donc absolument nécessaire, dans ce cas, que la valeur d'appartenance globale après agrégation soit très faible, voir nulle, pour la classe considérée. Cette propriété peut être définie par l'expression mathématique suivante :

$$H(0, \theta_1, \theta_2, \theta_3, \theta_4, \dots, \theta_n) \rightarrow 0$$

En conséquence cet opérateur est de type t-norme. L'effet compensatoire des opérateurs de la famille des t-conormes n'est pas souhaité. Afin de choisir l'opérateur, ce dernier doit être applicable à un espace de dimension supérieure à 2.

Plusieurs opérateurs remplissent les conditions que nous venons de définir pour la technique de classification du FPM. Le choix de l'un entre eux consiste à choisir l'opérateur le plus performant, c'est-à-dire qui donne le meilleur taux de classification. Une étude des performances des opérateurs d'agrégation a montré que le choix du meilleur opérateur dépend du jeu de données utilisé [RIB97] et qu'en général, les opérateurs les plus simples, comme le minimum et le produit, sont les meilleurs [DEV99].

3.6. Présentation des données tests

Nous avons utilisé cinq séries de données pour tester nos développements de la méthode FPM. Les quatre premières sont issues d'études menées au sein du Laboratoire d'Automatique et de Microélectronique à l'université de Reims Champagne Ardenne.

3.6.1. Données Matières Plastiques

Les bouteilles en matière plastique sont généralement fabriquées soit en polyéthylène téréphtalate (PET), soit en polyéthylène haute densité (PEHD) soit en polychlorure de vinyle (PVC). Pour être recyclées, ces différentes catégories de bouteilles doivent être séparées.

La spectrométrie infrarouge nous permet de caractériser les différentes matières [DEV99]. L'analyse des spectres de chaque matière plastique nous a amené à sélectionner 7 longueurs d'onde discriminantes. L'étude des corrélations nous a permis de réduire ce nombre à 2. Nous possédons 90 valeurs de transmission pour les deux longueurs d'onde λ_1 et λ_2 . Ces valeurs se répartissent en trois classes de 30 points chacune comme le montre la Figure 3.6.

Les classes sont bien séparées. Elles ont toutefois des formes allongées et inclinées, ce qui pose des problèmes en classification comme nous le verrons plus loin dans ce mémoire. De plus les classes 2 et 3 sont composées de sous classes.

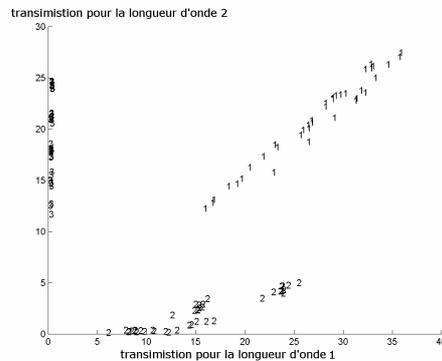


Figure 3.6. Représentation des 3 classes de matières plastiques : 1 PVC, 2 PET, 3 PEHD.

3.6.2. Données Codes Métalliques

Le Laboratoire d'Automatique et de Microélectronique a mis au point un dispositif utilisant un capteur plat à courants de Foucault pour lire des codes cachés. Ceux ci sont constitués d'une succession de fines bandes métalliques séparées par une zone isolante [GOU95]. La Figure 3.7.a représente la réponse du dispositif aux 9 codes définis par l'industriel.

Deux caractères ont été retenus pour leur pouvoir de discrimination, parmi les 52 déterminés. La Figure 3.7.b montre les 9 classes de 45 points obtenues. L'intérêt de ces données provient du nombre assez important de classes et de la faible compacité des classes 7 et 8.

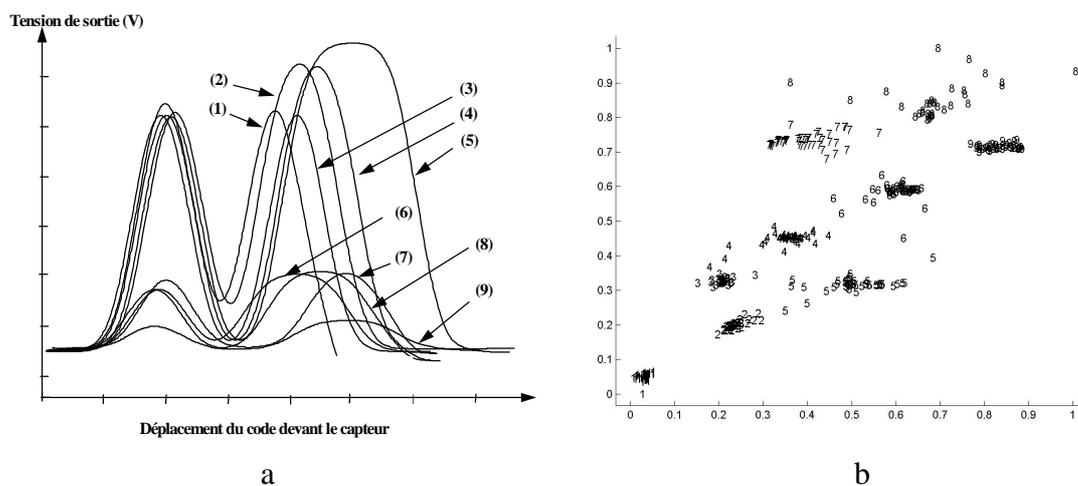


Figure 3.7. a) Réponse du dispositif aux 9 codes et b) représentation des 9 classes.

3.6.3. Données Machine à Laver

Ces données sont issues d'une étude menée au Laboratoire d'Automatique et de Microélectronique. L'objectif est de détecter les défauts de balourd, dans le tambour d'une machine à laver, en mesurant les mouvements dus aux vibrations pendant le fonctionnement. Les mesures sont réalisées par des capteurs plats à courant de Foucault [DEV99]. Nous obtenons un nuage de 204 points comportant 4 classes. Chaque classe correspond à un balourd et à sa position dans la cuve. Le nuage de points est présenté dans la Figure 3.8. Les classes ont des formes allongées et elles sont inclinées. Les classes 1 et 2 sont difficilement séparables.

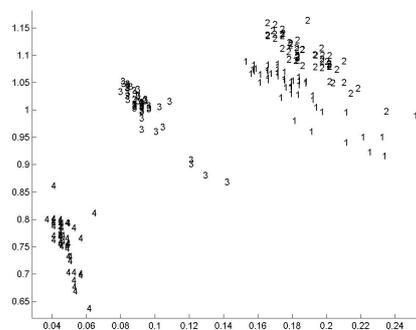


Figure 3.8. Données Machine à Laver, 1. balourd décentré à droite, 2. balourd centré, 3. balourd décentré à gauche, 4. pas de balourd.

3.6.4. Données Injections Plastiques

Ces données correspondent à l'étude d'une presse à injecter de la matière plastique. L'objectif de cette étude est de détecter les rebuts apparaissant lors de la fabrication des éprouvettes, en relevant uniquement les valeurs des paramètres de la presse, et sans contrôler directement les pièces [DEV99]. Cette étude a mené à trouver quatre classes de produits correspondant à : des pièces bonnes, des pièces incomplètes, des pièces présentant des bavures, des pièces présentant des retassures.

Ces classes sont décrites dans un espace à trois paramètres : la température de zone 3 du fourreau, le temps du maintien et la cote finale du matelas comme le montre la Figure 3.9.

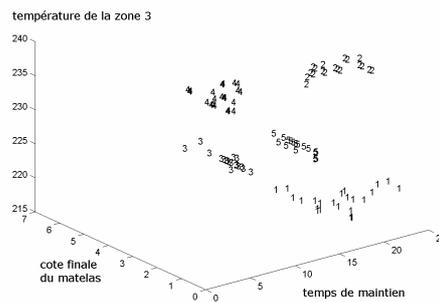


Figure 3.9. Nuage de points obtenu pour le processus d'injection plastique.

3.6.5. Données Iris

Ces données sont largement utilisées pour la comparaison des classifieurs. Elles se composent de 3 classes de fleurs : Iris Sétona, Versicolor, et Virginia. Chacune de ces classes est constituée de 50 échantillons caractérisés par 4 attributs : la longueur et la largeur des sépales, la longueur et la largeur des pétales. Deux classes parmi les trois se chevauchent en partie.

3.7. Apprentissage continu

Le diagnostic consiste à classer une nouvelle observation dans une des classes correspondant à un mode de fonctionnement connu. Le diagnostic en ligne nécessite un temps de classification faible en regard de l'application et constant quel que soit le nombre de points de l'ensemble d'apprentissage. FPM demande un temps de classification constant, pour chaque nouvelle observation, quel que soit le nombre de points de l'ensemble d'apprentissage. C'est une des raisons pour laquelle nous avons utilisé FPM comme méthode de classification supervisée dans notre module de diagnostic en ligne.

Quand le processus à diagnostiquer est évolutif ou lent, l'ensemble d'apprentissage ne contient pas suffisamment d'information pour apprendre les vraies formes des classes. L'information manquante doit être obtenue à partir des nouvelles observations classifiées.

FPM ne peut obtenir cette information qu'en faisant la mise à jour des densités de possibilité, c'est-à-dire en faisant à nouveau la phase d'apprentissage, après la classification de chaque nouvelle observation. Puisque cet apprentissage traite un volume de données élevé, il est généralement réalisé hors-ligne. Le temps nécessaire pour cet apprentissage augmente avec le nombre de points d'apprentissage, ce qui empêche la mise à jour en temps réel. Le volume mémoire nécessaire augmente de la même manière.

Pour remédier à ce problème, la solution est d'utiliser un apprentissage incrémental. En effet, l'apprentissage incrémental ajoute l'information apportée, par chaque nouvelle observation classifiée, à la connaissance qui existe déjà. Il n'a plus besoin d'utiliser l'ensemble d'apprentissage et en conséquence le temps de calcul nécessaire, pour l'ajout de la nouvelle information, ne dépend plus de la taille de celui-ci.

Il n'existe aucune méthode pour intégrer l'apprentissage incrémental dans FPM. Dans ce chapitre, nous présentons une approche incrémentale pour la mise à jour des densités de possibilité après la classification de chaque nouvelle observation, en conservant un temps et un volume mémoire constants et indépendants de la taille de l'ensemble de l'apprentissage.

3.7.1. Apprentissage continu dans FPM

L'augmentation du nombre de points, de l'ensemble d'apprentissage, améliore les performances de FPM [KLE98, DEV99]. Après un certain nombre de points classés dans une classe, les nouveaux points n'ajoutent plus d'information à la base de connaissance et ils deviennent donc redondants. La Figure 3.10 montre cette amélioration, en terme de taux de classification, pour plusieurs valeurs du nombre de barres h des histogrammes.

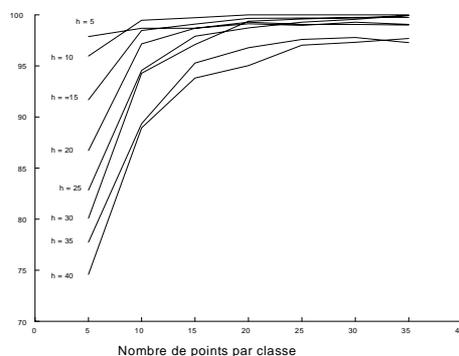


Figure 3.10. Relation entre le nombre de points d'apprentissage par classe et le taux de classification, pour différentes valeurs du nombre de barres h . pour les données Matières Plastiques

Pour montrer l'intérêt de l'apprentissage continu lors de l'utilisation de FPM, nous prenons l'exemple des deux classes circulaires de la Figure 3.11.

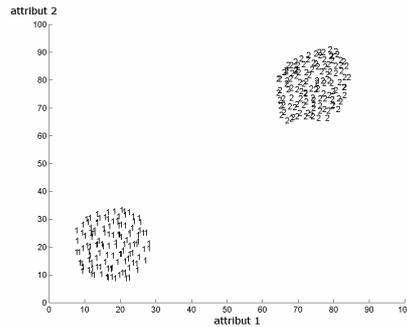


Figure 3.11. Exemple de deux classes.

La Figure 3.12.a montre les densités de possibilité calculées avec la moitié des points de l'ensemble d'apprentissage pour chaque classe. La Figure 3.12.b montre les densités calculées après apprentissage continu sur l'autre moitié. Nous voyons apparaître des différences entre ces deux cas. L'apprentissage continu enrichit les densités de possibilité et en conséquence il améliore les performances de FPM.

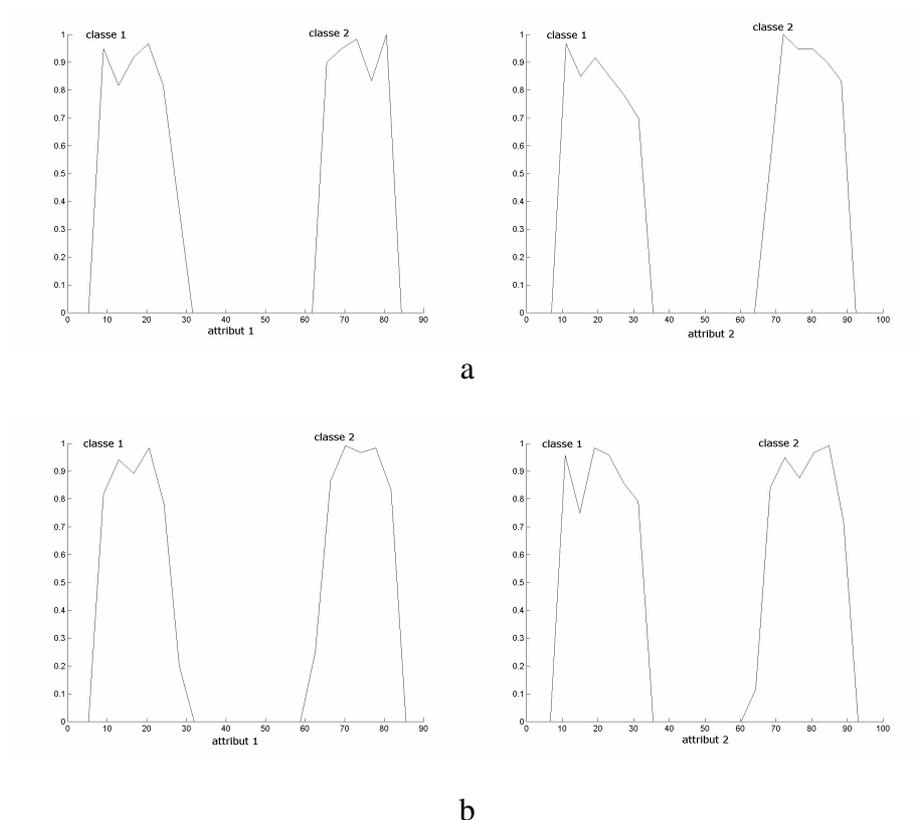


Figure 3.12. Densités de possibilité a) calculées avec la moitié de l'ensemble d'apprentissage, b) avec apprentissage continu sur l'autre moitié.

Les nouvelles observations qui sont rejetées ne participent pas à l'enrichissement des densités de possibilités. L'apprentissage continu tient donc compte de l'évolution des classes dans le cas simple c'est-à-dire lorsque l'évolution consiste en un allongement des classes par

des observations voisines et successives. Cela donne un autre intérêt de l'apprentissage continu. La Figure 3.13 montre un exemple où le nombre de points d'apprentissage, pour la classe 1, n'est pas suffisant pour donner la forme réelle de cette classe.

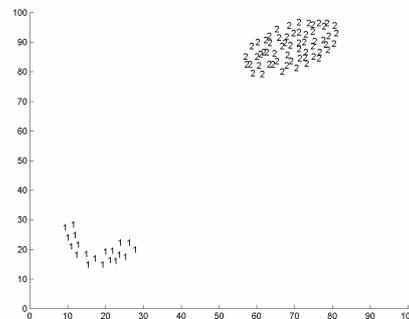


Figure 3.13. Ensemble d'apprentissage initial.

La Figure 3.14 montre le résultat graphique de la classification successive, sans apprentissage continu, de nouvelles observations qui devraient appartenir à la classe 1. Nous remarquons que FPM rejette presque la totalité de ces nouvelles observations.

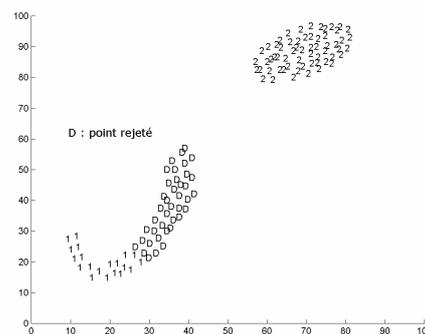


Figure 3.14. Classification de nouvelles observations sans utiliser l'apprentissage continu.

En revanche, en utilisant l'apprentissage continu, FPM classe toutes les nouvelles observations dans la classe 1 et restitue donc la forme réelle de cette classe, comme le montre la Figure 3.15. Cependant il faut faire à nouveau l'apprentissage de toutes les densités de possibilité après la classification de chaque nouvelle observation en utilisant tout l'ensemble d'apprentissage. En conséquence l'apprentissage continu est réalisé hors-ligne.

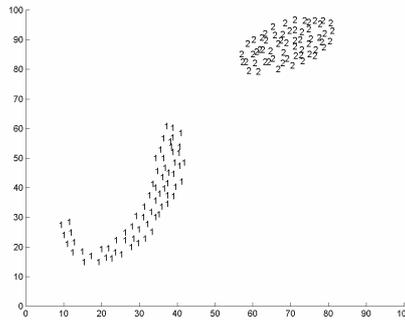


Figure 3.15. Classification de nouvelles observations avec l'apprentissage continu.

3.7.2. Apprentissage incrémental dans FPM

Nous proposons une approche pour intégrer l'apprentissage incrémental dans la méthode FPM [MOU02a]. Nous appliquons notre approche sur l'histogramme de la classe j suivant l'attribut i , cette démonstration est identique pour tous les autres histogrammes. Nous avons vu auparavant que la probabilité p_k associée à la barre k est égale au nombre de points n_k , de la classe j appartenant à cette barre, divisé par le nombre total de points d'apprentissage N_j de la classe j . Quand un nouveau point est classifié dans la classe j , le nombre de points appartenant à cette classe devient $N_j + 1$, donc si ce point se trouve dans la barre k , la probabilité de chaque barre devient :

$$p'_1 = \frac{n_1}{N_j + 1}, \quad p'_2 = \frac{n_2}{N_j + 1}, \quad \dots, \quad p'_k = \frac{n_k + 1}{N_j + 1}, \quad p'_{k+1} = \frac{n_{k+1}}{N_j + 1}, \quad \dots, \quad p'_h = \frac{n_h}{N_j + 1} \quad (3.3)$$

nous changeons la proposition (3.3) pour qu'elle soit :

$$\left\{ \begin{array}{l} p'_1 = \frac{n_1}{N_j} \cdot \frac{N_j}{N_j + 1} = p_1 \cdot \frac{N_j}{N_j + 1}, \\ p'_2 = \frac{n_2}{N_j} \cdot \frac{N_j}{N_j + 1} = p_2 \cdot \frac{N_j}{N_j + 1}, \quad \dots, \\ p'_k = \frac{n_k}{N_j} \cdot \frac{N_j}{N_j + 1} + \frac{1}{N_j + 1} = p_k \cdot \frac{N_j}{N_j + 1} + \frac{1}{N_j + 1}, \\ p'_{k+1} = \frac{n_{k+1}}{N_j} \cdot \frac{N_j}{N_j + 1} = p_{k+1} \cdot \frac{N_j}{N_j + 1}, \quad \dots, \\ p'_h = \frac{n_h}{N_j} \cdot \frac{N_j}{N_j + 1} = p_h \cdot \frac{N_j}{N_j + 1} \end{array} \right. \quad (3.4)$$

prenons maintenant les termes suivants :

$$P_j = (p_1 \ p_2 \ \dots \ p_h), P'_j = (p'_1 \ p'_2 \ \dots \ p'_h), T_j = \frac{N_j}{N_j + 1}, Z_j = \left(0 \ \dots \ \frac{1}{N_j + 1} \ 0 \ \dots \right)$$

où P'_j est le vecteur des nouvelles probabilités, P_j est le vecteur des probabilités avant la classification du nouveau point. Nous construisons Z_j en mettant le terme $\frac{1}{N_j + 1}$ dans la colonne correspondant à l'indice de barre dans laquelle le nouveau point se trouve, et zéros dans les autres colonnes. L'équation de l'apprentissage incrémental devient :

$$P'_j = T_j \cdot P_j + Z_j \quad (3.5)$$

L'équation (3.5) montre que la mise à jour de l'histogramme de probabilité ne nécessite pas le calcul de la nouvelle répartition des points, elle a seulement besoin de l'indice de la barre où se trouve ce nouveau point. En conséquence le temps de cette mise à jour est constant et indépendant de la taille de l'ensemble d'apprentissage. La taille de mémoire nécessaire pour cette mise à jour est également constante et indépendante de la taille de l'ensemble d'apprentissage.

L'information apportée par la classification d'un nouveau point diminue avec le nombre de points d'apprentissage jusqu'à un niveau où chaque nouveau point devient redondant, ce qui est confirmé par la Figure 3.10. L'équation suivante montre les limites de l'information apportée par l'apprentissage incrémental :

$$\left. \begin{array}{l} \lim_{N_j \rightarrow \infty} \frac{N_j}{N_j + 1} = 1 \\ \lim_{N_j \rightarrow \infty} \frac{1}{N_j + 1} = 0 \end{array} \right\} \Rightarrow P'_j \rightarrow P_j \quad (3.6)$$

Dans le paragraphe suivant, nous allons appliquer notre approche sur plusieurs exemples afin de confirmer son efficacité par rapport au temps de calcul, à la taille de mémoire et au nombre de points rejetés en appartenance.

3.7.3. Résultats expérimentaux

3.7.3.1. Temps de mise à jour et taille de mémoire

Nous comparons le temps de calcul et la taille de mémoire nécessaires pour la mise à jour des densités de possibilité, après la classification d'un nouveau point, entre FPM avec

l'apprentissage incrémental et FPM avec l'apprentissage continu hors-ligne. Nous commençons par l'exemple de la Figure 3.2. Nous initialisons l'ensemble d'apprentissage avec 320 points et nous calculons le temps de mise à jour des densités de possibilité après la classification d'un nouveau point pour différentes tailles de l'ensemble d'apprentissage. Le Tableau 3.2 et les Figure 3.16.a et Figure 3.16.b montrent les résultats obtenus .

Nombre de points d'apprentissage	FPM avec l'apprentissage continu hors-ligne		FPM avec l'apprentissage incrémental	
	Temps de mise à jour en secondes	Taille de mémoire $N + h . a . c$ en unités	Temps de mise à jour en secondes	Taille de mémoire $h . a . c$ en unités
320	0.44	400	0.06	80
400	0.55	480	0.06	80
480	0.71	560	0.06	80
560	0.83	640	0.06	80
640	0.93	720	0.06	80
720	1.04	800	0.06	80
800	1.1	880	0.06	80
880	1.2	960	0.06	80
960	1.37	1040	0.06	80
1040	1.48	1120	0.06	80
1120	1.59	1200	0.06	80
1200	1.7	1280	0.06	80
1280	1.81	1360	0.06	80
1360	1.92	1440	0.06	80
1440	1.98	1520	0.06	80

Tableau 3.2. Comparaison du temps de mise à jour et de la taille de mémoire entre FPM avec l'apprentissage incrémental et FPM avec l'apprentissage continu hors-ligne pour l'exemple de la Figure 3.2.

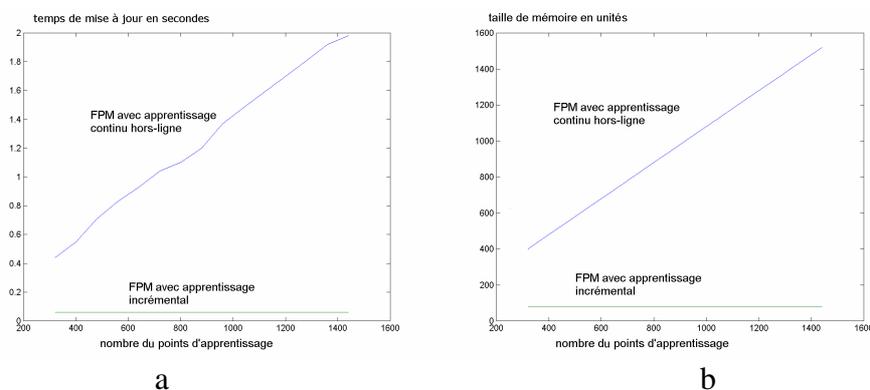


Figure 3.16. Relation graphique entre le nombre de points d'apprentissage et : a) le temps de mise jour et b) la taille de mémoire pour FPM avec l'apprentissage incrémental et FPM avec l'apprentissage continu hors-ligne pour l'exemple de la Figure 3.2.

Les Tableaux 3.3, 3.4, 3.5 et 3.6, et les Figures 3.17, 3.18, 3.19 et 3.20 montrent ces résultats pour respectivement : les données Matières Plastiques, les données Machine à Laver, les données Codes Métalliques et les données Injection Plastique. Nous avons fixé h à 20 pour tous ces exemples, afin de voir les facteurs qui influencent le temps de mise à jour pour les deux méthodes. Les données Matières Plastiques, Machine à Laver et Codes Métalliques ont le même nombre d'attributs mais diffèrent par le nombre de classes. En revanche les données Injection Plastique diffèrent des autres par le nombre d'attributs et le nombre de classes.

Nombre de points d'apprentissage	FPM avec l'apprentissage continu hors-ligne		FPM avec l'apprentissage incrémental	
	Temps de mise à jour en secondes	Taille de mémoire $N + h . a . c$ en unités	Temps de mise à jour en secondes	Taille de mémoire $h . a . c$ en unités
40	0.07	160	0.06	120
80	0.13	200	0.06	120
120	0.18	240	0.06	120
160	0.23	280	0.06	120
200	0.27	320	0.06	120
240	0.33	360	0.06	120
280	0.39	400	0.06	120
320	0.44	440	0.06	120
360	0.51	480	0.06	120
400	0.55	520	0.06	120

Tableau 3.3. Comparaison du temps de mise à jour et de la taille de mémoire entre FPM avec l'apprentissage incrémental et FPM avec l'apprentissage continu hors-ligne, pour les données Matières Plastiques.

Nombre de points d'apprentissage	FPM avec l'apprentissage continu hors-ligne		FPM avec l'apprentissage incrémental	
	Temps de mise à jour en secondes	Taille de mémoire $N + h . a . c$ en unités	Temps de mise à jour en secondes	Taille de mémoire $h . a . c$ en unités
51	0.08	211	0.06	160
102	0.14	262	0.06	160
153	0.2	313	0.06	160
204	0.27	364	0.06	160
255	0.33	415	0.06	160
306	0.38	466	0.06	160
357	0.45	517	0.06	160
408	0.52	568	0.06	160
459	0.58	619	0.06	160
510	0.63	670	0.06	160

Tableau 3.4. Comparaison du temps de mise à jour et de la taille de mémoire entre FPM avec l'apprentissage incrémental et FPM avec l'apprentissage continu hors-ligne, pour les données Machine à Laver.

Nombre de points d'apprentissage	FPM avec l'apprentissage continu hors-ligne		FPM avec l'apprentissage incrémental	
	Temps de mise à jour en secondes	Taille de mémoire $N + h . a . c$ en unités	Temps de mise à jour en secondes	Taille de mémoire $h . a . c$ en unités
45	0.06	405	0.06	360
90	0.09	450	0.06	360
135	0.14	595	0.06	360
180	0.18	540	0.06	360
225	0.21	585	0.06	360
270	0.26	630	0.06	360
315	0.29	675	0.06	360
360	0.34	720	0.06	360
405	0.38	765	0.06	360
450	0.41	810	0.06	360

Tableau 3.5. Comparaison du temps de mise à jour et de la taille de mémoire entre FPM avec l'apprentissage incrémental et FPM avec l'apprentissage continu hors-ligne, pour les données Codes Métalliques.

Nombre de points d'apprentissage	FPM avec l'apprentissage continu hors-ligne		FPM avec l'apprentissage incrémental	
	Temps de mise à jour en secondes	Taille de mémoire $N + h . a . c$ en unités	Temps de mise à jour en secondes	Taille de mémoire $h . a . c$ en unités
219	0.87	519	0.25	300
438	1.68	738	0.25	300
657	2.47	957	0.25	300
876	3.23	1176	0.25	300
1095	4.01	1395	0.25	300
1314	4.84	1614	0.25	300
1533	5.57	1833	0.25	300
1752	6.35	2052	0.25	300
1971	7.23	2271	0.25	300
2190	8.61	2490	0.25	300

Tableau 3.6. Comparaison du temps de mise à jour et de la taille de mémoire entre FPM avec l'apprentissage incrémental et FPM avec l'apprentissage continu hors-ligne, pour les données Injection Plastique.

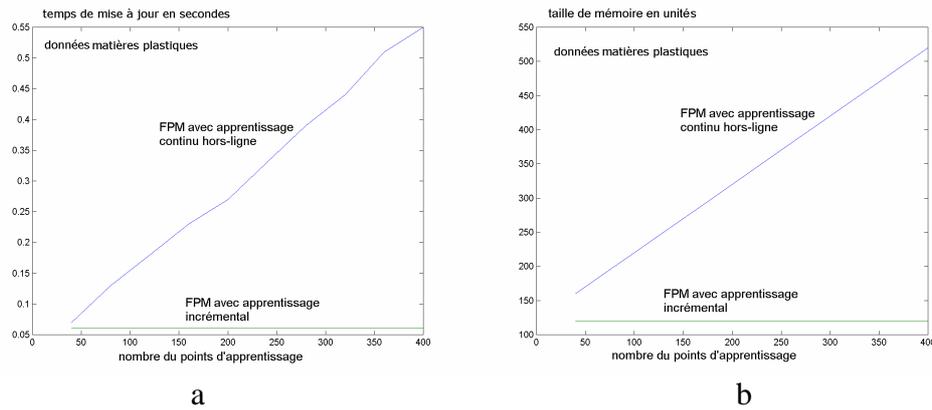


Figure 3.17. Relation graphique entre le nombre de points d'apprentissage et : a) le temps de mise à jour et b) la taille de mémoire pour FPM avec l'apprentissage incrémental et FPM avec l'apprentissage continu hors-ligne, pour les données Matières Plastiques.

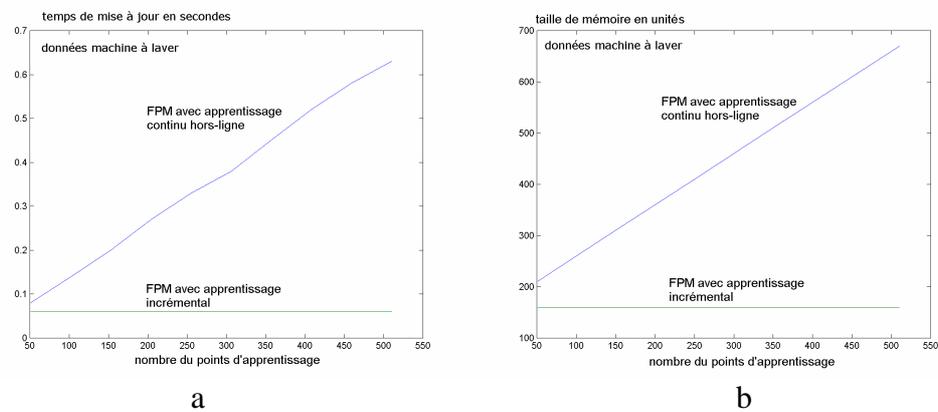


Figure 3.18. Relation graphique entre le nombre de points d'apprentissage et : a) le temps de mise à jour et b) la taille de mémoire pour FPM avec l'apprentissage incrémental et FPM avec l'apprentissage continu hors-ligne, pour les données Machine à Laver.

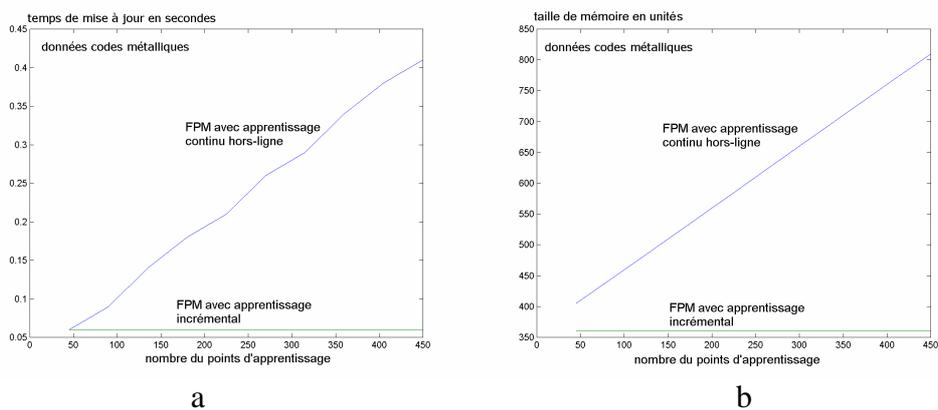


Figure 3.19. Relation graphique entre le nombre de points d'apprentissage et : a) le temps de mise à jour et b) la taille de mémoire pour FPM avec l'apprentissage incrémental et FPM avec l'apprentissage continu hors-ligne, pour les données Codes Métalliques.

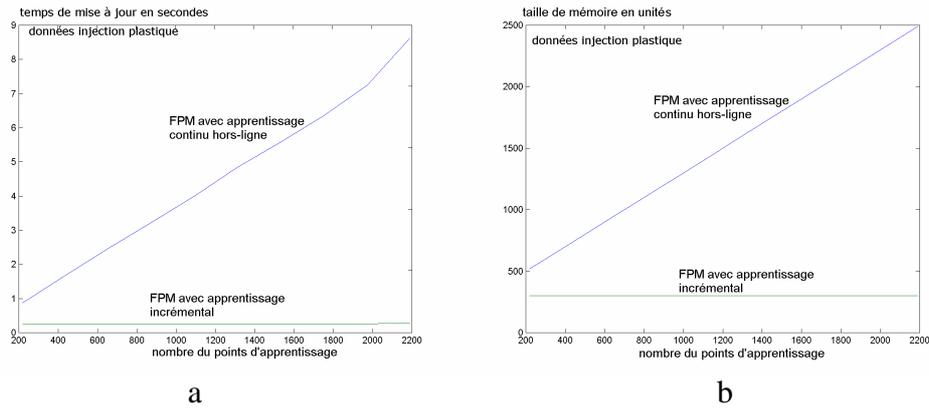


Figure 3.20. Relation graphique entre le nombre de points d'apprentissage et : a) le temps de mise à jour et b) la taille de mémoire pour FPM avec l'apprentissage incrémental et FPM avec l'apprentissage continu hors-ligne, pour les données Injection Plastique.

Nous comparons le temps de mise à jour en fonction de h pour les deux méthodes et pour les exemples précédents comme le montre les Figure 3.21.a, Figure 3.21.b, Figure 3.21.c et Figure 3.21.d.

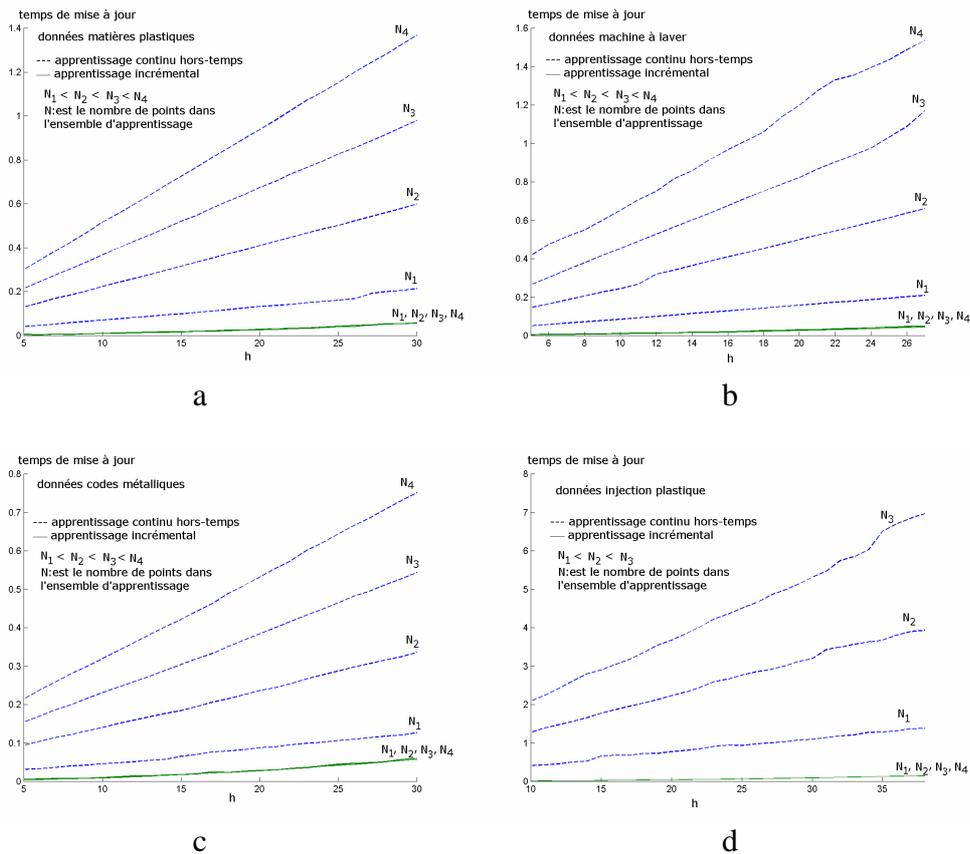


Figure 3.21. Relation entre le temps de mise à jour et le nombre de barres h pour différentes tailles de l'ensemble d'apprentissage : a) données Matières Plastiques, b) données Machine à Laver, c) données Codes Métalliques et d) données Injection Plastique.

En analysant ces résultats, nous constatons que :

- le temps de mise à jour T_{hl} , en utilisant l'apprentissage continu hors-ligne, augmente linéairement avec une pente proportionnelle au nombre d'observations N_j de la classe j à laquelle la nouvelle observation appartient, au nombre de barres h et au nombre d'attributs a : $T_{hl} \propto N_j \cdot h \cdot a$,
- le volume de mémoire M_{hl} nécessaire pour cette mise à jour, en utilisant l'apprentissage continu hors-ligne, augmente linéairement avec une pente proportionnelle d'une part au nombre des observations N , et d'autre part au nombre de barres h , au nombre d'attributs a et au nombre de classes c : $M_{hl} \propto N + h \cdot a \cdot c$,
- le temps de mise à jour T_{inc} , en utilisant l'apprentissage incrémental, est constant et indépendant du nombre d'observations dans l'ensemble d'apprentissage. Ce temps est proportionnel au carré du nombre de barres et au nombre d'attributs : $T_{inc} \propto h^2 \cdot a$. Cela est dû à la spécificité du calcul des densités de possibilité en utilisant la transformation de Dubois et Prade. Nous voyons que T_{inc} n'est pas seulement constant mais est également plus petit que T_{hl} , puisque h est toujours largement inférieur à N_j ,
- la taille de mémoire M_{inc} , en utilisant l'apprentissage incrémental, est constante et indépendante du nombre d'observations dans l'ensemble d'apprentissage. Cette taille est proportionnelle au nombre de barres h , au nombre d'attributs a et au nombre de classes c : $M_{inc} \propto h \cdot a \cdot c$.

3.7.3.2. Nombre de points rejetés en appartenance

La mise à jour des densités de possibilité permet de suivre l'évolution de la forme des classes. Cela pourrait diminuer le nombre de points rejetés en appartenance. Nous étudions l'effet de l'apprentissage incrémental sur le nombre de points rejetés en appartenance en utilisant les données Injection Plastique. Nous initialisons l'apprentissage avec 125 points, la Figure 3.22 montre le nuage initial dans l'espace de représentation. Nous disposons de 640 points qui forment l'ensemble de test.

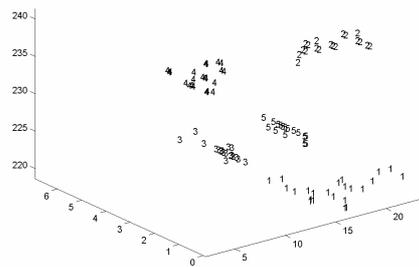


Figure 3.22. Nuage initial de points utilisés pour la phase d'apprentissage.

Lors de la première phase, nous essayons de classer les 640 points de l'ensemble de test. Les points correctement classifiés sont ajoutés à l'ensemble d'apprentissage tandis que les points rejetés forment l'ensemble de points qui seront classés lors de l'étape suivante. Plusieurs apprentissages sont réalisés successivement jusqu'à ce que tous les points soient affectés aux classes.

Ces opérations sont réalisées 2 fois : sans apprentissage continu et avec l'apprentissage incrémental. La Figure 3.23.a et la Figure 3.23.b montrent les résultats graphiques après le premier et le dernier apprentissage sans l'utilisation de l'apprentissage continu et la Figure 3.24.a et la Figure 3.24.b montrent les mêmes résultats avec l'apprentissage incrémental.

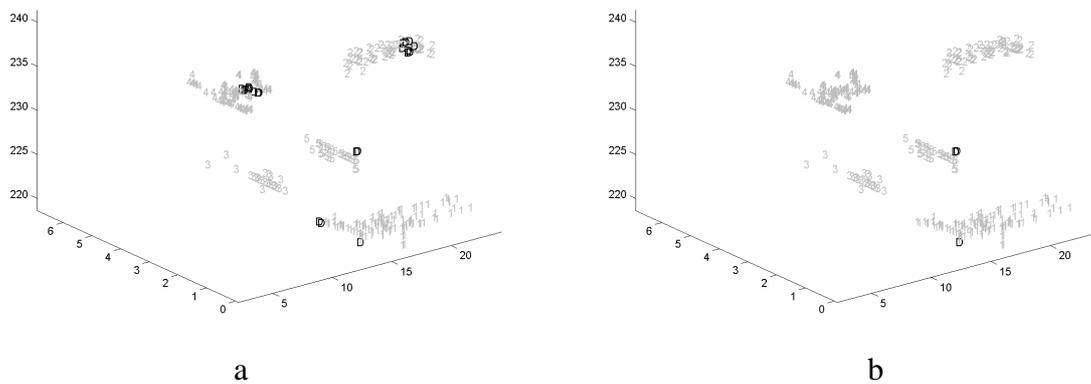


Figure 3.23. Résultat de classification après a) le premier apprentissage et b) le dernier apprentissage sans l'utilisation de l'apprentissage continu.

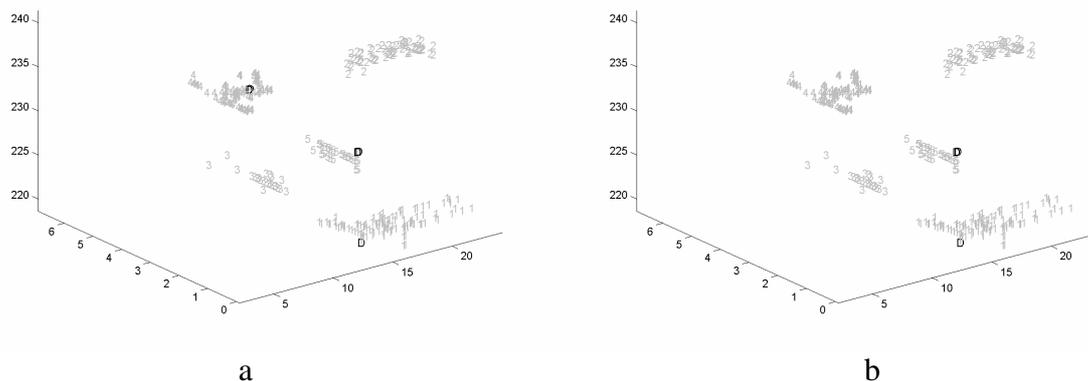


Figure 3.24. Résultat de classification après a) le premier apprentissage et b) le dernier apprentissage avec l'utilisation de l'apprentissage incrémental.

Le Tableau 3.4 montre les résultats pour 5 étapes d'apprentissage dans les 2 cas.

Sans apprentissage continu			Avec apprentissage incrémental	
étape	Points mal classés	Rejets en appartenance	Points mal classés	Rejets en appartenance
1	0	25	0	5
2	0	3	0	3
3	0	3	0	3
4	0	3	0	3
5	0	3	0	3

Tableau 3.7. Evolution du nombre de points rejetés en appartenance au fur et à mesure des apprentissage.

Nous constatons que le nombre de points rejetés en appartenance diminue avec le nombre de points d'apprentissage. En utilisant l'apprentissage incrémental, le nombre de points rejetés en appartenance est moins élevé que celui sans apprentissage continu, cela est dû à la mise à jour des densités de possibilité après la classification de chaque nouveau point. Nous ne pouvons pas descendre en dessous de 3 points rejetés en appartenance, en effet ces 3 points sont trop éloignés des classes pour y être affectés.

Dans l'expérience précédente, les points étaient rejetés si leur valeur d'appartenance était nulle pour toutes les classes. Nous voulons maintenant fixer un seuil de rejet en appartenance afin de n'accepter que les points assez proches ou se trouvant dans les zones centrales des classes. La Figure 3.25.a et la Figure 3.25.b montrent les résultats graphiques de classification après le premier et le dernier apprentissage sans apprentissage continu et la Figure 3.26.a et la Figure 3.26.b montrent ces résultats avec l'apprentissage incrémental.

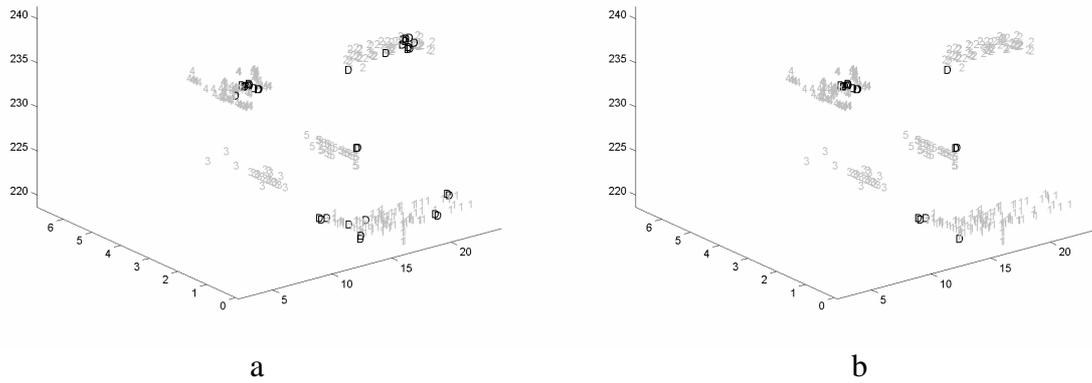


Figure 3.25. Résultat de classification avec un seuil de rejet en appartenance après a) le premier apprentissage et b) le dernier apprentissage sans l'utilisation de l'apprentissage continu.

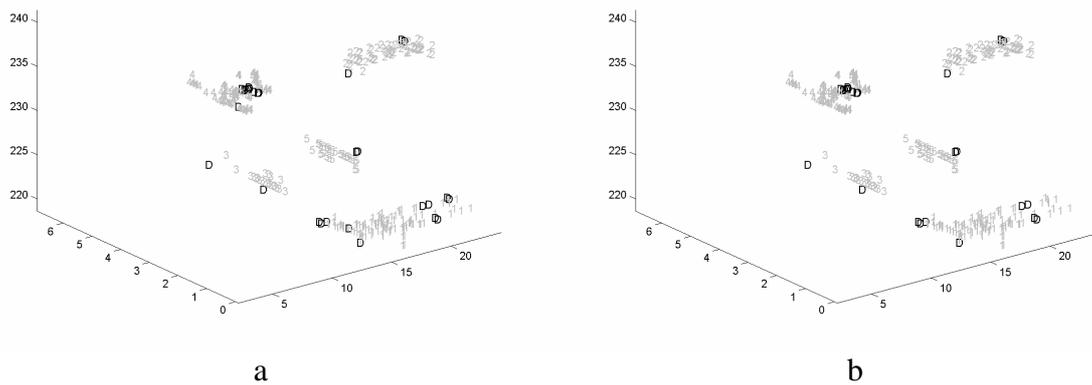


Figure 3.26. Résultat de classification avec un seuil de rejet en appartenance après a) le premier apprentissage et b) le dernier apprentissage avec l'utilisation de l'apprentissage incrémental.

Le Tableau 3.8 montre le nombre de points rejetés pour les deux cas et pour 5 étapes d'apprentissage. Le seuil de rejet en appartenance est fixé à 0.1.

étape	Sans apprentissage continu		Avec apprentissage incrémental	
	Points mal classés	Rejets en appartenance	Points mal classés	Rejets en appartenance
1	0	39	0	32
2	0	22	0	27
3	0	18	0	27
4	0	18	0	27
5	0	18	0	27

Tableau 3.8. Evolution du nombre de points rejetés au fur et à mesure des apprentissages avec un seuil de rejet en appartenance.

Nous constatons qu'avec l'apprentissage incrémental, les points se trouvant dans les zones périphériques, ou zones de faible densité, sont toujours rejetés en appartenance, et les points se trouvant dans les zones centrales, ou zones de forte densité, sont affectés aux classes. En revanche, dans le cas sans apprentissage continu, des points se trouvant dans des zones de faible densité sont acceptés.

3.8. Adaptation de FPM pour des classes de forme complexe

La méthode FPM se comporte de manière différente sur deux types de forme de classes :

- les classes de forme complexe ou les classes de forme simple dont l'orientation principale n'est parallèle à aucun axe de l'espace de représentation,
- les classes de forme simple dont l'orientation principale est parallèle à un axe de l'espace de représentation.

Les classes sphériques, ou elliptiques dont l'orientation principale est parallèle à un axe, sont des classes de forme simple, en revanche les classes obliques ou paraboliques sont des classes de forme complexe. Toute méthode de classification doit donner une fonction discriminante respectant la forme des classes. La méthode du FPM ne comporte pas de fonction analytique de discrimination. Toutefois nous pouvons visualiser la discrimination engendrée par un opérateur, produit, minimum ..., en traçant les courbes de niveau d'appartenance qui lui correspondent.

Dans cette section nous comparons la forme des courbes de niveau d'appartenance avec la forme des classes, pour différents jeux de données. Nous utilisons les deux opérateurs d'agrégation produit et minimum. Nous verrons que la discrimination résultante en utilisant la méthode FPM est correcte lorsque les classes ont des formes simples et sont assez séparables. Nous proposons une approche pour développer la méthode FPM afin qu'elle soit capable de discriminer correctement des classes qui ne sont pas assez séparables et qui sont de forme complexe [MOU02d]. Nous comparerons notre approche avec celles qui existent déjà.

3.8.1. Application de FPM aux classes de forme simple

La Figure 3.27 montre les courbes de niveau d'appartenance obtenues avec $h = 10$, pour les deux opérateurs d'agrégation, produit et minimum, pour l'exemple de deux classes de forme circulaire.

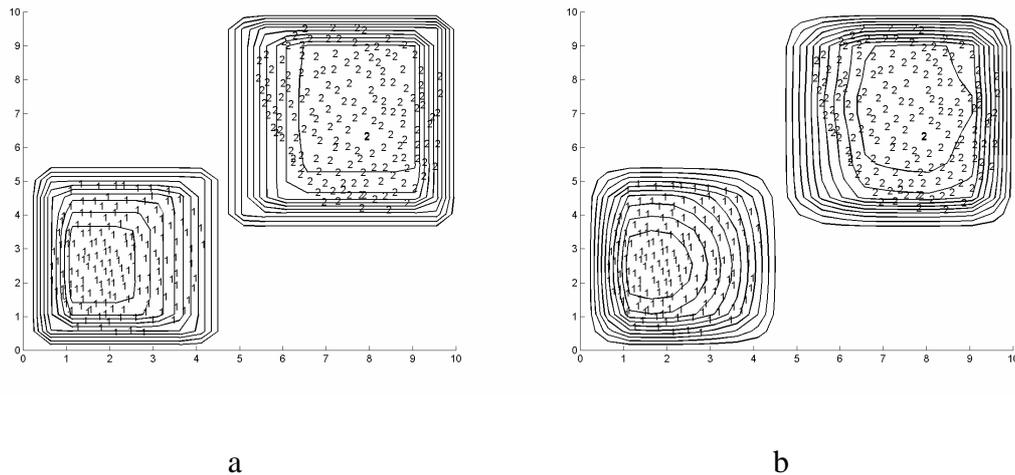


Figure 3.27. Courbes de niveau d'appartenance obtenues avec FPM a) avec l'opérateur minimum et b) avec l'opérateur produit.

Nous constatons que les courbes de niveau d'appartenance résultants de l'utilisation de l'opérateur d'agrégation produit respectent la forme des classes. En revanche, les courbes de niveau d'appartenance pour l'opérateur minimum sont de forme rectangulaire ce qui signifie le non respect de la forme circulaires des classes.

3.8.2. Application de FPM aux classes de forme complexe

Pour illustrer les limites de la méthode FPM, nous nous sommes intéressés aux nuages de points comportant des classes obliques et paraboliques dont trois exemples sont présentés sur les Figure 3.28, Figure 3.29 et Figure 3.30. Nous constatons que les courbes de niveau d'appartenance, obtenues pour l'opérateur minimum, sont de forme rectangulaire et ne respectent absolument pas la forme allongée et inclinée des classes. Ceci peut provoquer des erreurs de classification puisqu'un point excentré, ou situé dans la concavité, par rapport à la classe peut avoir une possibilité d'appartenance équivalente à un point situé dans celle-ci. Il lui est par conséquent affecté bien qu'il ne lui appartienne pas. L'opérateur produit donne des résultats légèrement meilleurs. Les formes des courbes de niveau d'appartenance sont légèrement plus proches de la forme des classes mais elles ne respectent toujours pas la forme de celles-ci.

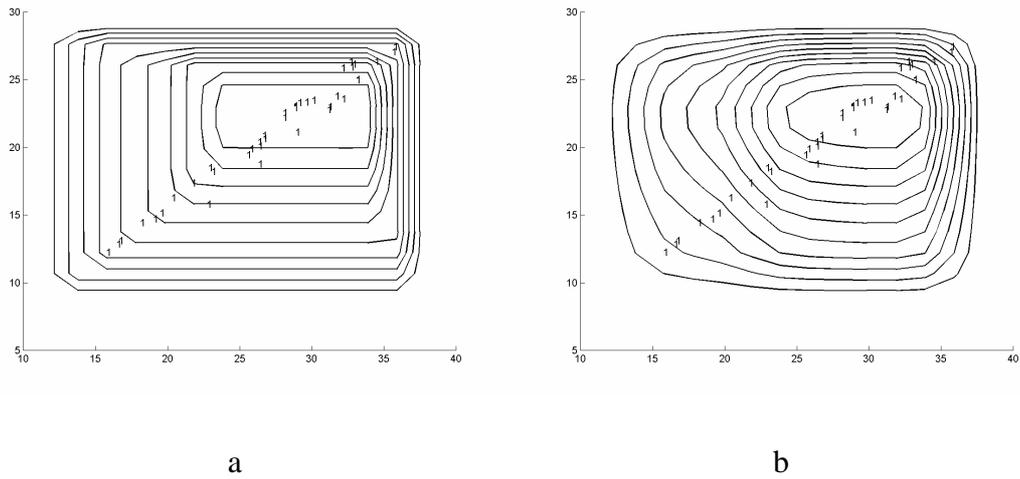


Figure 3.28. Courbes de niveau d'appartenance obtenues avec FPM pour une classe oblique a) avec l'opérateur minimum et b) avec l'opérateur produit.

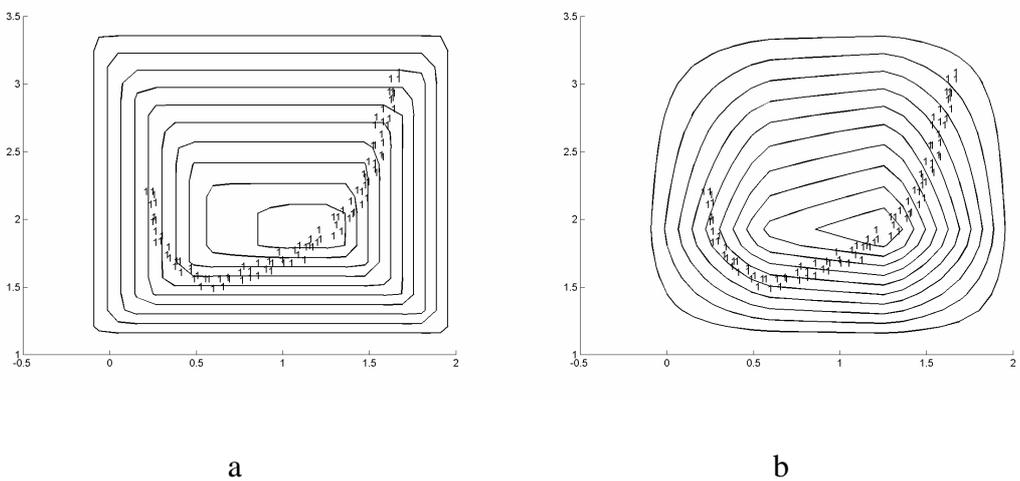


Figure 3.29. Courbes de niveau d'appartenance obtenues avec FPM pour une classe parabolique a) avec l'opérateur minimum et b) avec l'opérateur produit.

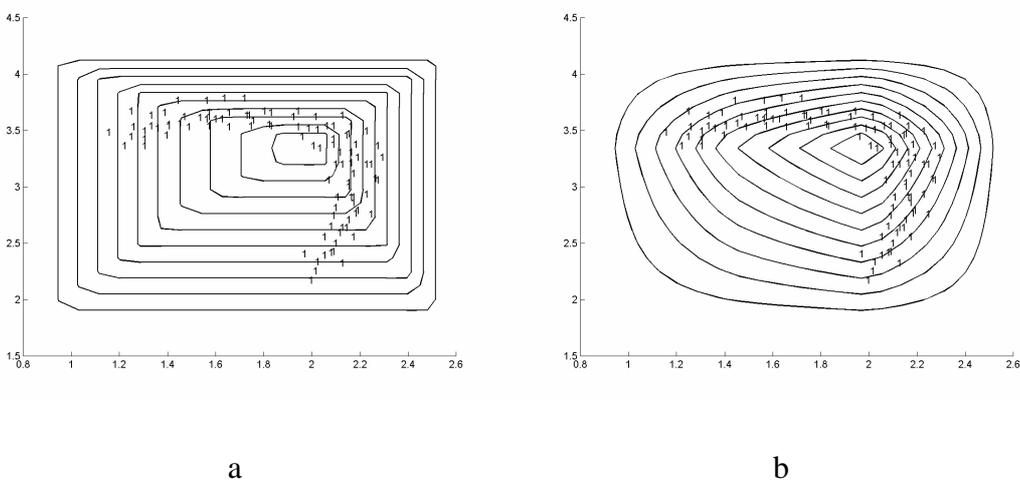


Figure 3.30. Courbes de niveau d'appartenance obtenues avec FPM pour une classe de forme fer à cheval a) avec l'opérateur minimum et b) avec l'opérateur produit.

Si une deuxième classe apparaît dans la concavité d'une de ces trois classes, la méthode FPM devient inopérante comme le montre la Figure 3.31. En effet, les courbes de niveau d'appartenance de la classe 1 englobent complètement la classe 2.

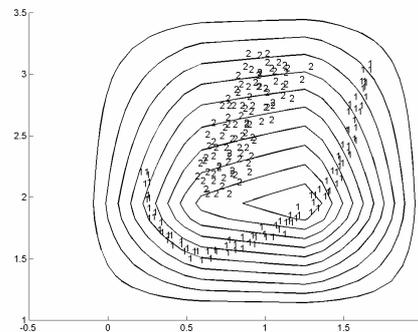


Figure 3.31. Nuage comportant deux classes, pour lequel FPM est inopérante.

Dès que les classes ont des formes complexes et qu'elles sont non séparables par un hyperplan, la méthode FPM devient inopérante quel que soit l'opérateur utilisé. La Figure 3.32 montre les courbes de niveau d'appartenance pour les 3 classes du nuage de données Matières Plastiques. Les points éloignés des classes sont affectés à l'une d'entre elles avec une possibilité d'appartenance élevée. La forme et l'orientation des classes ne sont pas respectées.

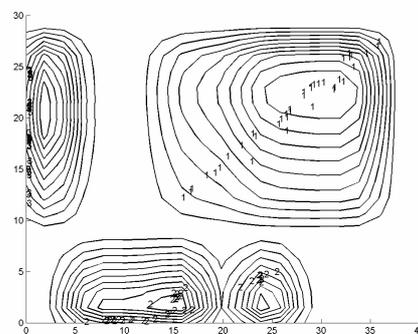


Figure 3.32. Résultat de classification des données matières plastiques.

Dans le cas des classes de formes complexe, le rejet d'un point, ayant une possibilité d'appartenance inférieure à un seuil déterminé, perd tout sens. La Figure 3.33 illustre une erreur de rejet. Le seuil est fixé à 0.1. Le point A est affecté à la classe tandis que le point B est rejeté [DEV99].

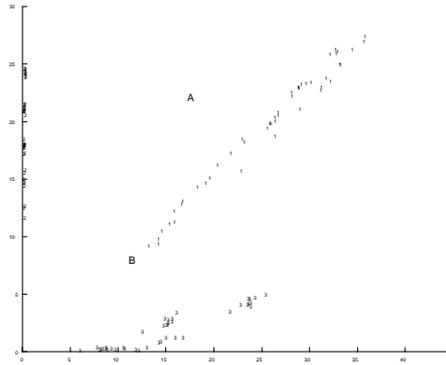


Figure 3.33. Utilisation d'un seuil de rejet en appartenance dégradant le taux de classification quand les classes ont une forme complexe.

Le problème du non respect de la forme des classes provient de l'estimation attribut par attribut, des densités, ce qui s'apparente à une projection des classes sur chacun des axes. Il ne peut pas être résolu par les opérateurs d'agrégation quels qu'ils soient. Deux solutions ont été proposées [DEV99] pour remédier à ce problème : Fuzzy Pattern Matching Multi-prototype (FPMM) et Fuzzy Pattern Matching avec une fonction Exponentielle FPME. La première consiste à augmenter le nombre de profils de possibilités en créant puis en fusionnant des sous classes. La seconde consiste à enrichir l'apprentissage en introduisant une information, indépendante des axes, sur la forme des classes. Ces deux solutions sont basées sur l'approche multi-prototype.

3.8.3. Fuzzy Pattern Matching Multiprofil (FPMM)

3.8.3.1. Décomposition des classes en sous-classes et calcul des densités de possibilité

Cette méthode consiste à enrichir l'apprentissage en utilisant l'approche multi-prototype. L'algorithme d'apprentissage présenté ci-dessous comporte deux étapes principales : la décomposition de chaque classe j en s_j sous-classes, en utilisant la méthode FCM, et le calcul des densités de possibilité pour chaque sous-classe et chaque attribut [DEV99] :

DEBUT

Fixer h

POUR j variant de 1 à c

Décomposer la classe j en s_j sous-classes par l'algorithme FCM

POUR i variant de 1 à s_j

POUR k variant de 1 à a

Calculer les h valeurs de l'histogramme de probabilité de l'ensemble d'apprentissage

Classer les valeurs dans l'ordre décroissant

Calculer la distribution de possibilité par une transformation probabilité-possibilité

FIN POUR

FIN POUR

FIN POUR

FIN

3.8.3.2. Utilisation des opérateurs d'agrégation

Lorsqu'un point apparaît dans l'espace de classification, une valeur de possibilité d'appartenance à chaque sous-classe i est calculée pour chaque attribut k et pour chaque classe j . L'agrégation de ces valeurs, par un opérateur, donne une possibilité d'appartenance μ_j^i à chaque sous-classe i de la classe j :

$$\mu_j^i(x) = H\left(\pi_j^{i1}(x), \pi_j^{i2}(x), \dots, \pi_j^{ia}(x)\right),$$

$\pi_j^{ik}(x)$ est la possibilité d'appartenance du point x à la sous-classe i de la classe j , pour le $k^{\text{ième}}$ attribut.

3.8.3.3. Fusion des possibilités d'appartenance

Le caractère possibiliste de l'algorithme impose l'utilisation de l'opérateur somme bornée, défini dans [MAS96], pour fusionner les possibilités d'appartenance. Lorsqu'un nombre s_j de sous-classes existe à l'intérieur d'une classe j , la fonction d'appartenance à cette classe est définie par :

$$u_j(\underline{x}) = \mu_j^1(\underline{x}) \cup \dots \cup \mu_j^{s_j}(\underline{x}) = \min \left[1, \sum_{i=1}^{s_j} \mu_j^i(\underline{x}) \right]$$

Le point est affecté à la classe pour laquelle il a la possibilité d'appartenance la plus forte. L'algorithme de classification est présenté ci dessous [DEV99] :

DEBUT

POUR un point \underline{x}

POUR j variant de 1 à c

POUR i variant de 1 à s_j

POUR k variant de 1 à a

Calculer la possibilité d'appartenance $\pi_j^{i,k}$ à la sous-classe i de la classe j et pour l'attribut k par lecture des profils

FIN POUR

Calculer la possibilité d'appartenance μ_j^i à la sous-classe i par agrégation des valeurs $\pi_j^{i,1}, \dots, \pi_j^{i,a}$

FIN POUR

Calculer $u_j(\underline{x})$ la possibilité d'appartenance à la classe j par une somme bornée des valeurs μ_j^i

FIN POUR

FIN POUR

FIN

3.8.3.4. Performances de l'algorithme

Deux paramètres influencent le respect de la forme des classes : le nombre de barres h et le nombre de sous-classes s_j créés dans la classe j . La valeur de h a été fixée expérimentalement à 20 pour l'exemple test dont les courbes de niveau d'appartenance sont présentées Figure 3.34 [DEV99].

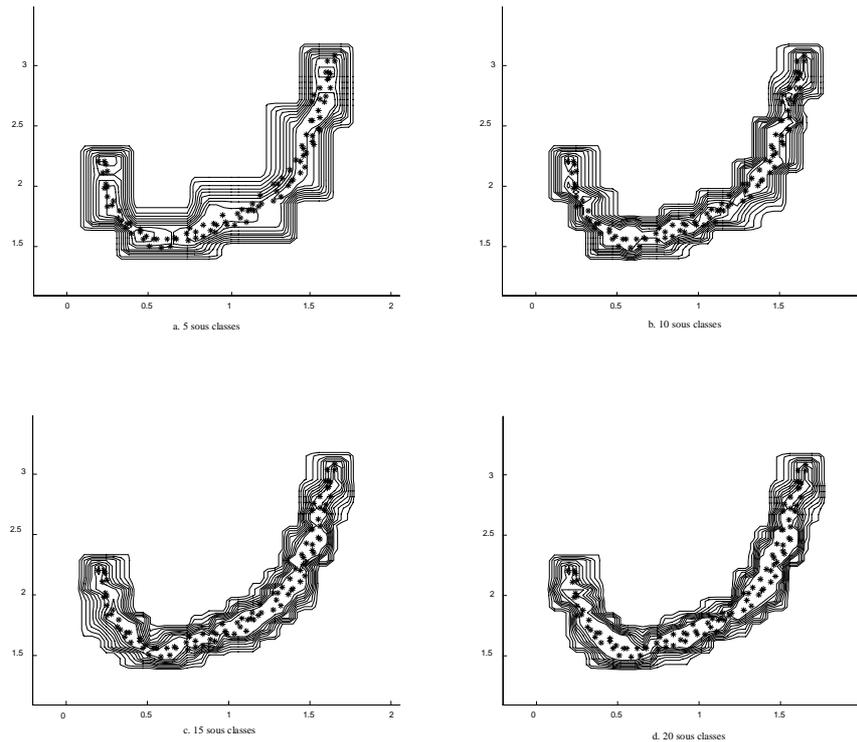


Figure 3.34. Courbes de niveau d'appartenance obtenues par l'algorithme FPMM pour l'opérateur minimum et $h = 20$. La classe est décomposée en a) 5 sous classes, b) 10 sous classes, c) 15 sous classes et d) 20 sous classes.

Nous constatons que le nombre de sous classes créées dans chaque classe a une forte influence sur la qualité des courbes de niveau d'appartenance, en terme de respect de la forme des classes. Plus ce nombre est important, plus les courbes de niveau d'appartenance respectent la forme des classes. Mais l'augmentation du nombre de sous-classes provoque un accroissement du temps de calcul, tant dans la phase d'apprentissage que dans la phase de classification, ce qui peut interdire l'application en temps réel de l'algorithme FPMM. De plus, cette augmentation cause une perte de l'information sur la distribution des points dans la classe. En effet, un nombre élevé de sous-classes conduit à attribuer à tous les points de l'ensemble d'apprentissage une valeur d'appartenance égale à 1, ce qui signifie que l'approche floue n'est plus appliquée. En conséquence FPMM ne peut pas distinguer les zones de haute densité de points dans la classe, comme le montre la Figure 3.35.a. Nous remarquons que l'algorithme FPMM ne distingue pas l'existence de deux zones de haute densité de points qui peuvent signifier des sous-modes de fonctionnement ou deux classes liées par du bruit représenté par des points aberrants. En revanche, un nombre petit de sous-classes provoque une hétérogénéité non-significative des niveau d'appartenance, comme le montre la Figure 3.35.b. Les courbes d'appartenance montrent l'existence de trois zones de

haute densité qui ne reflètent pas la vraie distribution des points dans la partie centrale de classe. Cela est dû à l'approche multiprofil utilisée par l'algorithme FPMM.

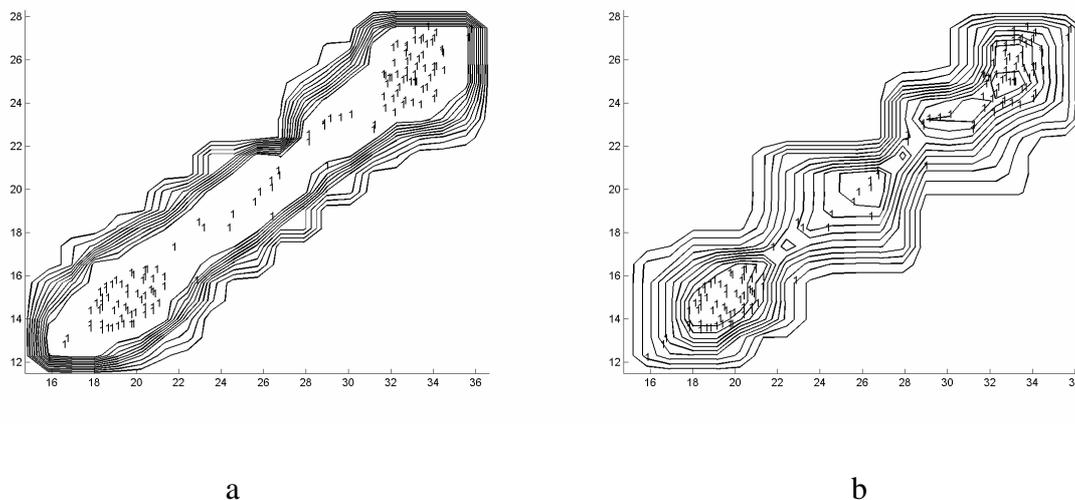


Figure 3.35. Courbes de niveau d'appartenance obtenues par l'algorithme FPMM pour l'opérateur minimum et $h = 10$. La classe est décomposée en a) 30 sous classes, b) 5 sous classes.

3.8.3.5. Apprentissage incrémental pour la méthode FPMM

L'utilisation de l'apprentissage continu dans la méthode FPMM est très coûteuse en temps de calcul. L'intégration de l'apprentissage incrémental est donc très utile.

Nous comparons FPMM avec apprentissage incrémental et FPMM avec l'apprentissage continu hors-ligne, en terme de temps de calcul et de nombre de points rejetés en appartenance sur les données Injection Plastique. Nous initialisons l'apprentissage avec 125 points et 3 sous-classes. Nous réalisons la première phase de classification avec 640 points et avec un seuil de rejet en appartenance égal à 0.2. Après chaque phase de classification, nous relançons un nouvel apprentissage en ajoutant les points correctement classés à l'ensemble d'apprentissage et en incrémentant le nombre de sous-classes. Les points rejetés forment l'ensemble de test pour la phase de classification suivante. Le Tableau 3.9 montre les résultats pour 8 étapes d'apprentissage.

	FPMM avec apprentissage continu hors-ligne	FPMM avec apprentissage incrémental
Nombre de sous classes	Temps moyen de mise à jour pour un point	Temps moyen de mise à jour pour un point
3	1.2	0.09
4	2.6	0.09
5	2.8	0.09
6	3.0	0.09
7	3.2	0.09
8	3.5	0.09
9	3.5	0.09
10	3.7	0.09

Tableau 3.9. Temps de mise à jour, après la classification d'un nouveau point, au fur et à mesure des apprentissage pour les méthodes FPMM avec apprentissage continu hors-ligne et FPMM avec apprentissage incrémental.

	FPMM sans apprentissage continu	FPMM avec apprentissage incrémental
Nombre de sous classes	Nombre de points rejetés en appartenance	Nombre de points rejetés en appartenance
3	42	28
4	30	18
5	25	16
6	19	16
7	19	15
8	19	15
9	19	13
10	17	11

Tableau 3.10. Evolution du nombre de points rejetés en appartenance au fur et à mesure des apprentissage pour les méthodes FPMM sans apprentissage continu et FPMM avec l'apprentissage incrémental.

Nous constatons que :

- avec l'apprentissage incrémental, le temps de mise à jour des densités de possibilité, après la classification de chaque nouveau point, est constant et indépendant de la taille de l'ensemble d'apprentissage,
- ce temps est moins élevé que celui obtenu avec l'apprentissage continu hors-ligne,
- ce temps est constant par rapport au nombre de sous-classes, contrairement au temps obtenu avec l'apprentissage continu hors-ligne qui augmente avec le nombre de sous-classes. Cela donne un nouvel avantage pour l'apprentissage incrémental quand on utilise la méthode FPMM,

- le nombre de points rejetés en appartenance est moins élevé en utilisant FPMM avec apprentissage incrémental que celui obtenu en utilisant FPMM sans l'apprentissage continu, grâce à la mise à jour après la classification de chaque nouveau point.

3.8.4. Fuzzy Pattern Matching avec une fonction Exponentielle (FPME)

La méthode Fuzzy Pattern Matching avec une fonction Exponentielle (FPME), développée dans [DEV99] consiste à diviser chaque classe j en s_j sous-classes. Chacune de ces sous-classes a une fonction d'appartenance représentée par un fonction exponentielle. La valeur d'appartenance d'un nouveau point \underline{x} à la sous-classe i de la classe j est donnée par l'expression suivante :

$$\mu_j^i(\underline{x}) = \exp[-\lambda_j^i \cdot d(\underline{x}, g_j^i)] \quad (3.7)$$

où g_j^i est le centroïde de la sous-classe i et $d(\underline{x}, g_j^i)$ est la distance de Mahalanobis entre le nouveau point et le centroïde. L'introduction de cette distance permet de respecter la forme elliptique des sous-classes. Le paramètre λ_j^i permet d'étaler la courbe de la fonction exponentielle par rapport à son centroïde.

Pour chaque classe j , nous déterminons une fonction exponentielle globale par l'union des s_j fonctions exponentielles définies sur les s_j sous-classes :

$$\mu_j(\underline{x}) = \mu_j^1(\underline{x}) \cup \dots \cup \mu_j^{s_j}(\underline{x}) = \min \left[1, \sum_{i=1}^{s_j} \mu_j^i(\underline{x}) \right] \quad (3.8)$$

Le degré d'appartenance d'un point pour chaque classe est obtenu par agrégation de la valeur de la fonction exponentielle globale et des valeurs de possibilités sur chaque attribut. Cette agrégation est réalisée par l'opérateur minimum :

$$u_j(\underline{x}) = \min [\pi_j^1(\underline{x}), \pi_j^2(\underline{x}), \pi_j^3(\underline{x}), \dots, \pi_j^a(\underline{x}), \mu_j(\underline{x})] \quad (3.9)$$

où $\pi_j^k(\underline{x})$ est la possibilité d'appartenance du point \underline{x} à la classe j par rapport à l'attribut k .

Nous constatons sur la Figure 3.36 que les courbes de niveau d'appartenance respectent la forme de la classe pour le nuage test. En conséquence, l'introduction de la fonction exponentielle rend l'utilisation de la méthode FPM possible pour les classes de forme complexe.

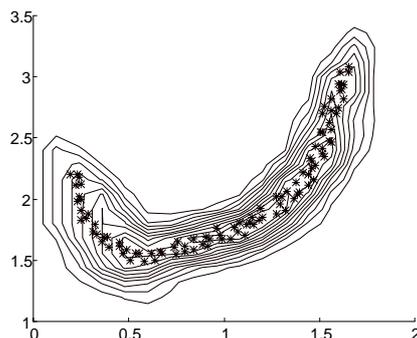


Figure 3.36. Courbes de niveau d'appartenance obtenues par agrégation des possibilités et de la fonction exponentielle globale pour la classe parabolique.

Cette méthode souffre, cependant, de plusieurs inconvénients :

- elle nécessite la détermination de plusieurs paramètres qui influencent fortement la qualité du respect de la forme des classes. Ces paramètres sont : le nombre de barres h , le nombre de sous-classes s_j pour chaque classe j et le paramètre λ_j^i de chaque sous-classe i de chaque classe j . Il n'existe actuellement aucune méthode pour le choix de ces paramètres,
- le temps de calcul est très élevé aussi bien pour la phase d'apprentissage que pour la phase de classification,
- puisque FPME utilise l'approche multiprofil, elle n'est pas capable de distinguer les zones de haute densité, si elles existent, comme on l'a vu pour la méthode FPMM.

Pour remédier à ces problèmes, nous proposons comme solution la méthode Fuzzy Pattern Matching Corrélée (FPMC) que nous développons dans la section suivante.

3.8.5. Fuzzy Pattern Matching Corrélée (FPMC)

La méthode FPM suppose que les différents axes de l'espace de représentation sont orthogonaux ou autrement dit indépendants. Cette supposition conduit à des courbes de niveau d'appartenance rectangulaires ou circulaires pour toutes les formes des classes. En conséquence la phase d'apprentissage ne cherche pas à déterminer la forme des classes.

Nous proposons donc d'inclure, dans la phase d'apprentissage, une information sur la forme des classes. Cette information est représentée par une liaison ou une corrélation entre les différentes densités de possibilité de tous les axes de l'espace de représentation. L'espace de représentation n'est plus orthogonal. Nous appelons notre méthode Fuzzy Pattern

Matching Corrélée (FPMC) [MOU02d].

3.8.5.1. Phase d'apprentissage

Considérons un ensemble d'apprentissage Ω , étiqueté, contenant N points répartis dans c classes, dans un espace de représentation à a attributs. Chaque classe j contient N_j points projetés sur un histogramme, contenant h barres, pour chaque attribut i . La matrice de corrélation B est composée de plusieurs matrices de corrélation, une pour chaque classe j :

$$B = [B_j]_{j=1 \dots c} \quad (3.10)$$

La construction de la matrice de corrélation B_j , de la classe j , commence par la recherche du vecteur de corrélation $(B_{\alpha i})_j$ entre la barre $\alpha = 1 \dots h$, du premier attribut et toutes les barres de l'attribut $i = 2 \dots a$, pour la classe j . Par exemple, $(B_{12})_1$ est le vecteur de corrélation entre la première barre du premier attribut et toutes les barres du deuxième attribut pour la classe 1. On détermine ensuite la matrice de corrélation $(B_i)_j$ entre toutes les barres du premier attribut et toutes les barres de l'attribut, $i = 2 \dots a$, pour la classe j :

$$(B_i)_j = [(B_{\alpha i})_j]_{\alpha=1 \dots h} \quad (3.11)$$

la matrice de corrélation de la classe j est alors :

$$B_j = [(B_i)_j]_{i=2 \dots a} \quad (3.12)$$

$(B_{\alpha i})_j$ est calculé de la manière suivante : si, au moins, un point d'apprentissage de la classe j se trouve dans la barre α du premier attribut et, en même temps, dans la barre β de l'attribut i , nous mettons la valeur β dans la colonne β du vecteur de corrélation $(B_{\alpha i})_j$, sinon nous mettons zéro. $(B_{\alpha i})_j$ est donc déterminé par :

$$\forall x_k^1 \in b_{1\alpha} \Rightarrow \text{si } x_k^i \in b_{i\beta} : \text{colonne } \beta \text{ de } (B_{\alpha i})_j = \beta : k = 1 \dots N_j \quad (3.13)$$

$b_{i\beta}$ est la barre β de l'attribut i .

Pour bien comprendre cette méthode, prenons l'exemple suivant d'une classe oblique. La Figure 3.37.a montre le nuage comportant une classe de forme oblique et la Figure 3.37.b montre les histogrammes de probabilité pour chaque axe, avec $h = 10$.

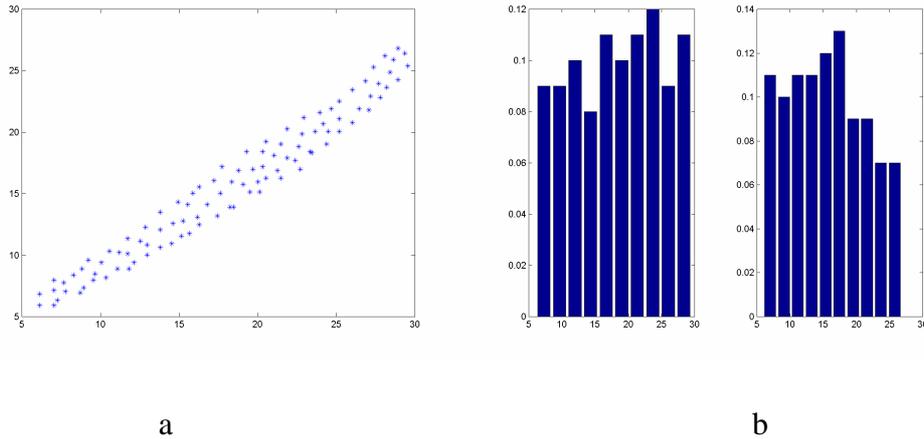


Figure 3.37. a) Nuage comportant une classe oblique, et b) histogrammes de probabilité obtenus avec $h = 10$.

La matrice de corrélation B pour cet exemple est la suivante :

$$\beta = \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{matrix}$$

$$B = \begin{pmatrix}
 1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha = 1, i = 2, j = 1 \\
 1 & 2 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha = 2, i = 2, j = 1 \\
 0 & 2 & 3 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha = 3, i = 2, j = 1 \\
 0 & 0 & 3 & 4 & 5 & 0 & 0 & 0 & 0 & 0 & \alpha = 4, i = 2, j = 1 \\
 0 & 0 & 3 & 4 & 5 & 6 & 0 & 0 & 0 & 0 & \alpha = 5, i = 2, j = 1 \\
 0 & 0 & 0 & 4 & 5 & 6 & 0 & 0 & 0 & 0 & \alpha = 6, i = 2, j = 1 \\
 0 & 0 & 0 & 0 & 5 & 6 & 7 & 0 & 0 & 0 & \alpha = 7, i = 2, j = 1 \\
 0 & 0 & 0 & 0 & 0 & 6 & 7 & 8 & 0 & 0 & \alpha = 8, i = 2, j = 1 \\
 0 & 0 & 0 & 0 & 0 & 0 & 7 & 8 & 9 & 0 & \alpha = 9, i = 2, j = 1 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 9 & 10 & \alpha = 10, i = 2, j = 1
 \end{pmatrix} \tag{3.14}$$

Chaque ligne de cette matrice montre la corrélation entre une barre α du premier attribut et toutes les barres du second attribut. La première ligne signifie qu'au moins un point d'apprentissage se trouvant dans la première barre d'historgramme du premier attribut, se trouve également dans une des deux premières barres d'historgramme du deuxième attribut.

3.8.5.2. Phase de classification

La classification d'un nouveau point x commence par la détermination des barres, $b_{l\alpha}, b_{i\beta} : \alpha = 1..h, \beta = 1..h, i = 2..a$, dans lesquelles ce nouveau point se trouve. Puis la valeur de possibilité d'appartenance π_j , à la classe j , est calculée par la projection linéaire sur les densités de possibilité de cette classe si et seulement si :

$$\forall i = 2 \dots a : \text{colonne } \beta \text{ de } (B_{\alpha i})_j = \beta \quad (3.15)$$

sinon le point sera rejeté par rapport à la classe j et $\pi_j = 0$. Enfin, le point sera classifié dans la classe pour laquelle il a la possibilité d'appartenance la plus élevée. Les algorithmes du fonctionnement de la méthode FPMC, dans sa phase d'apprentissage, et dans sa phase de classification sont montrés respectivement Figure 3.38.a et Figure 3.38.b.

<pre> Initialiser la matrice B à 0 Pour $j = 1 \dots c$ Pour $i = 2 \dots a$ Pour $\alpha = 1 \dots h$ Pour $k = 1 \dots N_j$ Si $x_k^1 \in b_{1\alpha}$ Trouver la barre β telle que $x_k^i \in b_{i\beta}$ colonne β de $(B_{\alpha i})_j = \beta$ Fin Si Fin Pour k Fin Pour α Fin Pour i Fin Pour j </pre> <p style="text-align: center;">a</p>	<pre> Pour un nouveau point \underline{x} Pour $j = 1 \dots c$ Trouver $\alpha : x^1 \in b_{1\alpha}$ Pour $i = 2 \dots a$ Trouver $\beta : x^i \in b_{i\beta}$ Si colonne β de $(B_{\alpha i})_j = \beta$ Calculer $\pi_j^i(\underline{x})$ Sinon $\pi_j(\underline{x}) = 0$ Sortir de la boucle i Fin Si Fin Pour j Fin Pour </pre> <p style="text-align: center;">b</p>
--	---

Figure 3.38. Algorithmes du fonctionnement de la méthode FPMC : pour a) la phase d'apprentissage et b) pour la phase de classification.

La Figure 3.39 montre les courbes de niveau d'appartenance, pour l'exemple de la Figure 3.37.a.

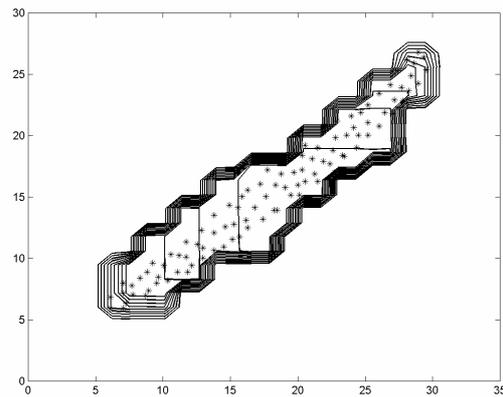


Figure 3.39. Courbes de niveau d'appartenance obtenues par la méthode FPMC avec $h = 10$ pour une classe oblique.

Nous constatons que ces courbes respectent la forme oblique de la classe. En revanche, les courbes dans la partie centrale de la classe forment des îlots d'appartenance en fonction de la densité de répartition des points.

Nous pouvons augmenter la précision du respect de la forme des classes en augmentant le nombre de barres h . Mais cela provoque l'augmentation du nombre d'îlots d'appartenance puisque l'augmentation de h met plus en valeur la diversité de la distribution des points de l'ensemble d'apprentissage, comme le montre la Figure 3.40 obtenue pour $h = 16$.

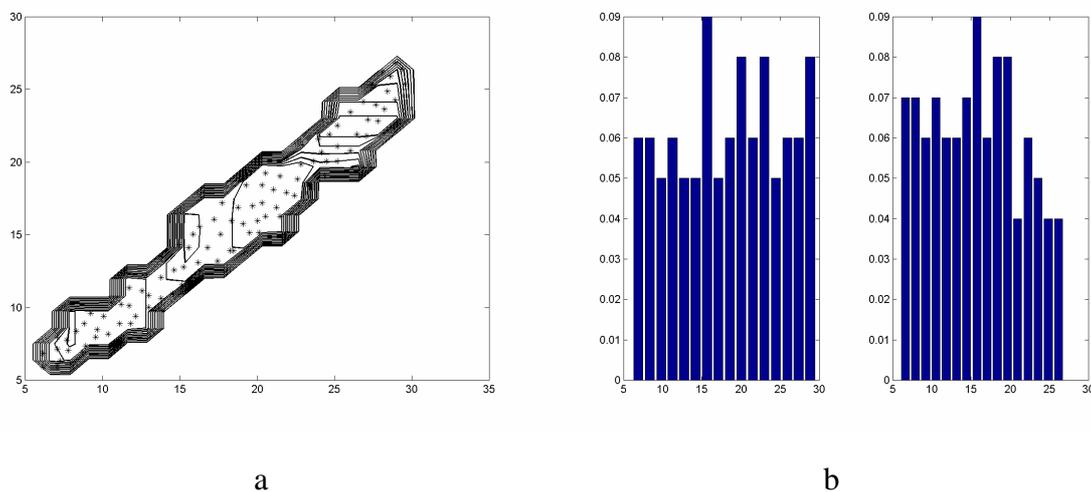


Figure 3.40. a) Courbes de niveau d'appartenance obtenues par la méthode FPMC avec $h = 16$ pour une classe oblique et b) histogrammes de probabilité.

3.8.6. Performances de FPMC

L'application de la méthode FPMC aux données Matières Plastiques donne des courbes de niveau d'appartenance respectant la forme des classes. La Figure 3.41 montre ces courbes obtenues avec $h = 10$.

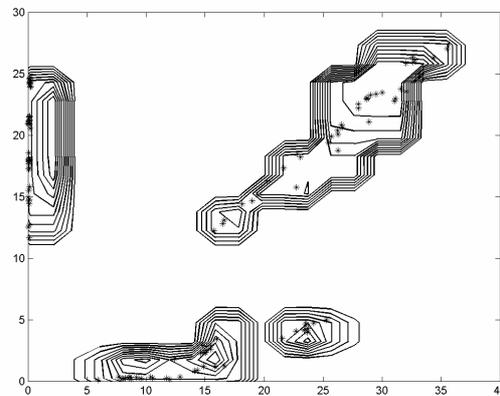


Figure 3.41. Courbes de niveau d'appartenance obtenues par la méthode FPMC avec $h = 10$ pour les données Matières Plastiques.

Nous remarquons l'existence d'îlots d'appartenance dus à la diversité de densité des points dans l'ensemble d'apprentissage. Les courbes de niveau d'appartenance élevé se trouvent autour des zones les plus denses.

La Figure 3.42 montre les courbes de niveau d'appartenance obtenues par FPMC, pour un exemple contenant deux sous-classes représentées par deux zones de haute densité. Nous constatons que les courbes de niveau d'appartenance se concentrent sur ces deux zones. En revanche, les deux méthodes FPMM et FPME ne distinguent pas ces deux sous-classes comme on l'a vu Figure 3.35.

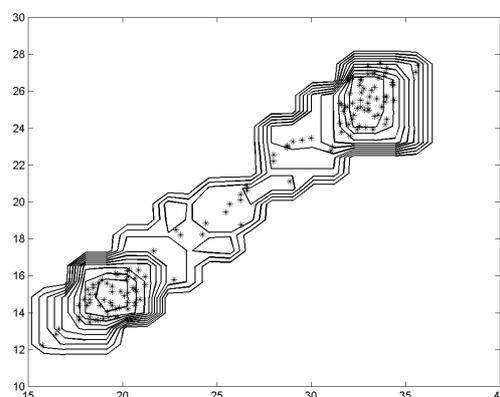


Figure 3.42. Courbes de niveau d'appartenance obtenues par la méthode FPMC avec $h = 10$ pour une classe oblique contenant deux sous-classes.

3.9. Conclusion

FPM souffre de deux inconvénients majeurs :

- dans un environnement évolutif, FPM doit adapter ses densités de possibilités après la classification de chaque nouvelle observation. Le temps nécessaire pour cette adaptation augmente avec le nombre de points de l'ensemble d'apprentissage, ce qui empêche la mise à jour des densités de possibilité en temps réel,
- les performances de FPM se dégradent dans le cas des classes de forme complexe ou dans le cas de classes qui ne sont pas bien séparées. Le problème du non respect de la forme des classes provient de l'estimation, attribut par attribut, des densités de possibilité. Il ne peut pas être résolu par les opérateurs d'agrégation quels qu'ils soient.

Pour remédier au premier problème, nous avons proposé une approche incrémentale pour réaliser un apprentissage continu après la classification de chaque nouveau point. L'information apportée par ce dernier est ajoutée à la connaissance existante sans avoir à calculer de nouveau les histogrammes de probabilité. En conséquence le temps de calcul est constant et faible. L'application de l'apprentissage incrémental est limitée au cas où les nouvelles observations sont bien classifiées. Si la nouvelle observation est mal classée, l'utilisation de l'apprentissage incrémental dégraderait les performances de classification. Dans le chapitre suivant, nous proposons un critère pour décider si un nouveau point peut être utilisé par l'apprentissage incrémental.

En ce qui concerne le second problème, deux solutions avaient déjà été proposées : Fuzzy Pattern Matching Multi-prototype (FPMM) et Fuzzy Pattern Matching avec une fonction Exponentielle (FPME). Ces deux méthodes souffrent de plusieurs inconvénients : leurs performances dépendent fortement de plusieurs paramètres, les temps de calcul sont très élevés et elles sont incapables de distinguer des zones de haute densité dans les classes.

Nous avons proposé une nouvelle approche pour rendre l'utilisation de la méthode FPM possible, en temps réel, pour des classes de forme complexe ayant des zones de haute densité. Cette approche consiste à inclure, dans la phase d'apprentissage, une information sur la forme des classes. Cette information est représentée par une liaison ou une corrélation entre les différentes densités de possibilité de tous les axes de l'espace de représentation. Nous appelons cette approche Fuzzy Pattern Matching Corrélée.

Les performances de FPM et de ses dérivées dépendent du nombre de barres h des histogrammes. La détermination de la valeur optimale de h reste une chose primordiale pour

obtenir les meilleures performances. Dans le chapitre suivant, nous introduisons une approche pour déterminer la valeur optimale de h .

4. Optimisation des paramètres de la méthode FPM

4.1. Introduction

La fonction d'appartenance de FPM, est construite en utilisant des histogrammes de probabilité. Chaque histogramme est défini par deux paramètres : le pas représenté par le nombre de barres et la largeur de l'intervalle représentée par les bornes inférieure et supérieure. En changeant ces deux paramètres, un nombre illimité d'histogrammes peuvent être définis pour le même ensemble de données. Le choix de ces paramètres influence les performances de la méthode FPM [MOU02f].

Il existe peu d'approches pour la détermination de la valeur des paramètres d'histogramme. Ces approches ont montré que le nombre de barres dépend du nombre de points de l'ensemble d'apprentissage. Elles utilisent des formules mathématiques en se basant sur un critère de minimisation d'une fonction d'erreur.

Dans le cas des méthodes non-paramétriques pour lesquelles les densités de probabilité sont inconnues, la minimisation d'une fonction d'erreur est impossible. C'est la cas de FPM où l'on ne dispose d'aucune information sur la forme de la densité de probabilité. De plus, contrairement aux approches précédentes, nous cherchons à optimiser les paramètres de plusieurs histogrammes à la fois, en tenant compte de la relation entre ces histogrammes qui peut affecter les performances générales de la méthode. L'utilisation des techniques de validation croisée donnent des résultats très dépendants de l'ensemble d'apprentissage et elles doivent être appliquées hors-ligne. Pour ces raisons, les paramètres sont déterminés expérimentalement ou heuristiquement.

Dans ce chapitre, nous proposons une nouvelle approche afin d'optimiser la valeur de ces paramètres en respectant à la fois les particularités de l'ensemble d'apprentissage et de la méthode FPM.

4.2. Paramètres d'un histogramme

4.2.1. Nombre de barres

Dans notre cas, l'histogramme est un outil d'estimation de l'occurrence des points d'apprentissage, dans une classe, dont découle la densité de probabilité. Deux paramètres sont nécessaires pour définir un histogramme : le pas représenté par le nombre de barres h et la

largeur de l'intervalle représentée par les bornes inférieure Y_1^k et supérieure Y_2^k , pour chaque attribut k . La largeur d'une barre b^k , suivant l'attribut k , est calculée par :

$$b^k = \frac{Y_2^k - Y_1^k}{h} \quad (4.1)$$

Une valeur trop élevée de h entraîne une représentation des données avec trop de détail alors qu'une valeur trop faible cause une perte d'information. Dans les deux cas, l'histogramme résultant est une fausse estimation de la densité de probabilité. Il faut donc choisir h pour révéler l'essentiel de la structure de la densité de probabilité.

Malgré l'importance de ce paramètre, peu d'études ont été consacrées à la détermination de la valeur optimale de h , surtout dans le cas où la forme de la densité de probabilité est inconnue. En supposant que cette forme est gaussienne, Scott [SCO79] propose de choisir la largeur de barre b par :

$$b = 3,49 \cdot s \cdot N^{-1/3} \quad (4.2)$$

où s est l'écart type de l'ensemble d'apprentissage et N est le nombre de points.

Freedman et Diaconis proposent une formule similaire mais plus robuste [IZE91, SCO91] :

$$b = 2 \cdot (IQR) \cdot N^{-1/3} \quad (4.3)$$

où IQR est l'interquartile. Cette formule est plus robuste parce que l'interquartile est moins sensible que la variance aux points aberrants.

Sturges donne une formule théorique pour le calcul de h , dans le cas d'une densité de probabilité normale. Cette formule donne la valeur minimale de h , utilisée comme une valeur par défaut dans les modules statistiques [SCO91] :

$$h = 1 + \log_2 N \quad (4.4)$$

He [HE97] propose de choisir h qui minimise la fonction d'erreur définie par :

$$h = \min \left(\frac{\sum_{m=1}^h \frac{p_m \cdot (1 - p_m)}{(N + 1)}}{\{(1 + (1 - r))\}}, \quad 5 \leq h \leq 20 \right) \quad (4.5)$$

$$\left\{ \frac{(1 - \frac{1}{h})}{(N + 1)} \right\}$$

où r est l'entropie de Shannon divisée par le nombre de barres h et p_m est la probabilité de la barre m de l'histogramme. L'entropie de Shannon est utilisée pour mesurer l'incertitude contenue dans l'ensemble d'apprentissage. Cette formule traduit l'effet qu'un ensemble d'apprentissage pauvre en information a besoin d'un histogramme avec moins de barres qu'un ensemble d'apprentissage riche.

En supposant que la densité de probabilité est normale et pour un niveau de coupe $\alpha = 0.05$, Otnes [OTN72] calcule h par :

$$h = 1,87 \cdot (N - 1)^{2/5} \quad (4.6)$$

Klein [KLE98] utilise une formule empirique pour la méthode FPM, en faisant l'analogie avec la formule de détermination du nombre de plus proches voisins de la méthode k-ppv :

$$h = \frac{\sqrt{N}}{2} \quad (4.7)$$

Rudemo [RUD82] a proposé de minimiser l'erreur carrée intégrale définie par :

$$ISE = \int \{ \hat{f}_h(x) - f(x) \}^2 . dx \quad (4.8)$$

où $\hat{f}_h(x)$ est la densité de probabilité estimée de $f(x)$, en faisant intervenir le paramètre h . On utilise les couples d'apprentissage $\{y_k = f(x_k), x_k : k = 1 .. n \text{ et } k \neq i\}$ pour estimer $\hat{f}_{h,-i}(x)$ la densité de probabilité sans le $i^{\text{ème}}$ point. Cette estimation est validée en calculant l'erreur d'estimation $y_i - \hat{f}_{h,-i}(x_i)$ définie par :

$$CV(h) = \frac{1}{N^2} \sum_{i=1}^N \{ y_i - \hat{f}_{h,-i}(x_i) \}^2 \quad (4.9)$$

On choisit h qui minimise (4.9) :

$$h = \arg_{h_i} \min [CV(h)] \quad (4.10)$$

Toutes ces méthodes nécessitent soit la connaissance de la forme de la densité de probabilité, soit la minimisation d'une fonction d'erreur. Dans notre cas, nous ne disposons d'aucune information sur la forme de densité de probabilité. La minimisation d'une fonction d'erreur demande beaucoup de calcul et la complexité de ce calcul augmente considérablement dans le cas d'une dimension supérieure à 1, ce qui est souvent le cas dans

les problèmes industriels. De plus, le minimum obtenu est très dépendant de l'ensemble d'apprentissage. Cela pose un grand problème si l'ensemble d'apprentissage est incomplet ce qui est le cas souvent dans la plupart des applications industrielles. Il n'existe aucune méthode pour déterminer le nombre de barres maximisant les performances de la méthode FPM, nous proposons, dans la suite de ce chapitre, une méthode fiable pour remédier à ce problème.

4.2.2. Bornes inférieure et supérieure

La distance entre les bornes inférieure Y_1^k et supérieure Y_2^k exprime l'intervalle de variation des valeurs pour un processus par rapport au paramètre k . Toutes les observations à l'extérieur de cet intervalle ne sont pas considérées par l'histogramme. Un intervalle de variation trop petit cause donc une perte d'information et un intervalle trop large entraîne une sensibilité de l'histogramme aux points aberrants ou au bruit.

Dans la littérature, les bornes d'un histogramme sont déterminées soit comme les valeurs minimum et maximum de l'ensemble d'apprentissage, si on ne connaît pas la fonction de densité de probabilité [HE97], soit en cherchant ces bornes pour que chaque barre ait au moins deux points d'apprentissage, si on connaît la fonction de densité de probabilité [OTN72]. Actuellement, il n'existe aucune méthode pour déterminer les bornes d'un histogramme pour maximiser les performances de FPM.

4.3. Estimation des performances de FPM

La probabilité d'erreur et le temps de classification d'un nouveau point déterminent les performances de toute méthode de classification. L'erreur de Bayes est l'erreur minimale de classification, elle est indépendante de la méthode de classification utilisée. Son calcul, comme pour toute estimation d'une probabilité d'erreur, nécessite l'estimation de la densité de probabilité. Si la forme de la densité de probabilité est inconnue, comme dans la plupart des cas réels, cette estimation est très difficile, voire impossible, surtout dans un espace de dimension supérieure à 2 [SAN00].

En pratique, la probabilité d'erreur est estimée en utilisant un ensemble composé de N points. Cet ensemble est divisé en deux, la première partie est l'ensemble d'apprentissage qui contient N_{app} points et la deuxième partie est l'ensemble de test qui contient N_t points. La probabilité d'erreur P_e est estimée par le calcul du nombre de points mal classés, N_{mc} , par rapport au nombre total de points dans l'ensemble de test :

$$P_e = \frac{N_{mc}}{N_t} \quad (4.11)$$

Il existe plusieurs méthodes d'estimation de la probabilité d'erreur :

- la méthode de resubstitution où le même ensemble de points est utilisé pour mettre au point le classifieur et le tester. L'estimation de la probabilité d'erreur par cette méthode est optimiste puisqu'on obtient une valeur inférieure à la valeur vraie [DUB90],
- la méthode de validation croisée : l'ensemble de points est divisé en deux sous ensembles, un ensemble d'apprentissage et un ensemble de test. L'estimation de la probabilité d'erreur reste biaisée mais pessimiste, c'est-à-dire supérieure à la valeur vraie [DUB90]. Il existe deux méthodes pour réaliser cette division : la méthode de laissé de coté, leaving-one-out, où l'ensemble d'apprentissage est de taille $N - 1$, le point restant formant l'ensemble de test. L'opération est répétée N fois, c'est-à-dire jusqu'à ce que chaque point ait été utilisé comme point test. L'inconvénient de cette méthode est son temps élevé de calcul et son avantage est l'exploitation de façon beaucoup plus performante de l'ensemble de points. La deuxième méthode utilise N_{app} points pour l'ensemble d'apprentissage et le reste forme l'ensemble de test.

La variance $V\hat{P}$ de l'estimation de la probabilité d'erreur, pour la méthode de validation croisée, est calculée par [RAU91] :

$$V\hat{P} = \frac{E\hat{P} \cdot (1 - E\hat{P})}{N_t} \quad (4.12)$$

où $E\hat{P}$ est la valeur moyenne de la probabilité d'erreur estimée \hat{P}_e . La probabilité d'erreur estimée, \hat{P}_∞ , dans le cas d'un ensemble de points de taille infinie, peut être calculée à l'aide des probabilités d'erreur estimées par la méthode de resubstitution \hat{P}_R et par la méthode de leaving-one-out \hat{P}_L [RAU91] :

$$\hat{P}_\infty = \frac{\hat{P}_L + \hat{P}_R}{2} \quad (4.13)$$

les performances d'une méthode de classification dépendent de deux facteurs :

- le nombre de points d'apprentissage par rapport au nombre d'attributs,

- la séparabilité des classes.

Puisque le premier facteur est fixé a priori, nous allons utiliser la séparabilité des classes comme indicateur des performances de la méthode FPM. Nous introduisons une mesure afin de quantifier cette séparabilité, ensuite nous cherchons h qui maximise cette mesure.

4.4. Mesures de séparabilité des classes

Les performances d'une méthode de classification dépendent de la structure des classes que présentent les données. Si ces classes ne sont pas bien séparables, la probabilité d'erreur aura une valeur trop élevée quelque soit la méthode de classification utilisée.

Il existe plusieurs mesures de séparabilité dans la littérature. [ZAD65, CHE76, BEZ79, BEZ 81, FRI 96]. Bezdek [BEZ81] a proposé la mesure de séparabilité entre les deux classes i et j définie par :

$$G = 1 - \max_{i+1 \leq j \leq c} \left\{ \max_{1 \leq i \leq c-1} \left\{ \frac{r_i + r_j}{c_{ij}} \right\} \right\} \quad (4.14)$$

où r_i et r_j sont respectivement les diamètres des plus petites sphères centrées sur les centroïdes \hat{v}_i et \hat{v}_j des classes i et j :

$$r_i = \max_{1 \leq k \leq N_i} \{d(x_k, \hat{v}_i)\}$$

$$r_j = \max_{1 \leq m \leq N_j} \{d(x_m, \hat{v}_j)\}$$

$$\hat{v}_i = \frac{1}{N_i} \sum_{k=1}^{N_i} x_{ik}$$

$$\hat{v}_j = \frac{1}{N_j} \sum_{m=1}^{N_j} x_{jm}$$

N_i et N_j sont respectivement les nombres de points d'apprentissage appartenant aux classes i et j , et $d(x_k, \hat{v}_i)$ est la distance entre le point x_k , appartenant à la classe i , et le centroïde de cette classe. c_{ij} est la distance entre les deux centroïdes des classes i et j . La mesure de séparabilité G est égale à 0 si les deux classes sont tangentes, les deux classes sont séparables si $0 < G < 1$ et elles sont chevauchantes si $G < 0$. La maximisation de G peut être utilisée pour trouver le nombre de classes dans un nuage de données [BEZ81].

Une autre mesure de séparabilité est la matrice intra-classe qui est généralement utilisée pour sélectionner les paramètres les plus discriminants pour l'espace de représentation. Cette matrice est définie par :

$$\hat{\Sigma} = \frac{1}{N} \sum_{j=1}^c N_j \cdot (\hat{v}_j - \hat{v}) \cdot (\hat{v}_j - \hat{v})^T \quad (4.15)$$

\hat{v} est le le centre de gravité general de toutes les classes défini par :

$$\hat{v} = \frac{1}{N} \sum_{j=1}^c N_j \cdot \hat{v}_j \quad (4.16)$$

Frigui et Krishnapuram [FRI96] ont proposé une mesure de similarité entre deux classes, utilisée comme critère de détermination du nombre de classes pour la méthode Unsupervised Robust C Prototype (URCP) [FRI91]. Si C_i et C_j sont respectivement les ensembles de points appartenant aux classes i et j , si C_{ij} est l'union de C_i et C_j et si u_{ik} et u_{jk} sont respectivement les degrés d'appartenance de x_k aux classes i et j , alors les cardinalités des ensembles C_i et C_j sont respectivement : $|C_i| = \sum_{x_k \in C_i} u_{ik}$ et $|C_j| = \sum_{x_k \in C_j} u_{jk}$. La mesure de similarité SI_{ij} entre les

deux classes i et j est définie par :

$$0 \leq SI_{ij} = 1 - \frac{\sum_{x_k \in C_{ij}} |u_{ik} - u_{jk}|}{\sum_{x_k \in C_i} u_{ik} + \sum_{x_k \in C_j} u_{jk}} \leq 1 \quad (4.17)$$

Quand les deux classes i et j sont identiques, $SI_{ij} = 1$, en revanche $SI_{ij} = 0$ si ces deux classes sont complètement séparées. Une valeur de SI_{ij} comprise entre 0 et 1 signifie que les deux classes se chevauchent.

Toutes ces mesures sont basées soit sur la distance entre les points d'apprentissage et les centroïdes des classes soit sur les degrés d'appartenance obtenu par URCP. Elles ne font pas intervenir h et en conséquence elles ne peuvent pas servir à estimer ce paramètre. De plus, elles demandent des temps de calcul très élevés et un volume de mémoire très important puisqu'elles utilisent tous les points de l'ensemble d'apprentissage. Enfin, ces mesures ne sont pas adaptées à la particularité de FPM qui considère deux classes complètement séparables si leurs densités de probabilité, ou de possibilité, sont également complètement séparables. Prenons l'exemple de la Figure 4.1, les mesures précédentes considèrent les deux classes

complètement séparables. En revanche pour FPM, ces deux classes sont partiellement chevauchantes à cause des zones de chevauchement entre les densités de possibilité. La Figure 4.2 montre un exemple de deux classes qui sont séparables par FPM puisque les densités de possibilité des deux classes sont complètement séparables par l'attribut 1.

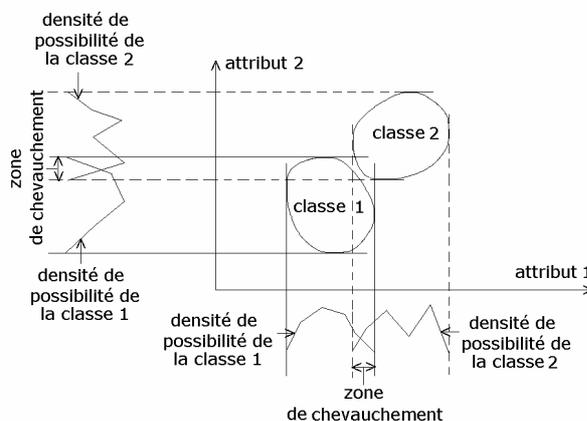


Figure 4.1. Exemple de deux classes non-séparables par FPM.

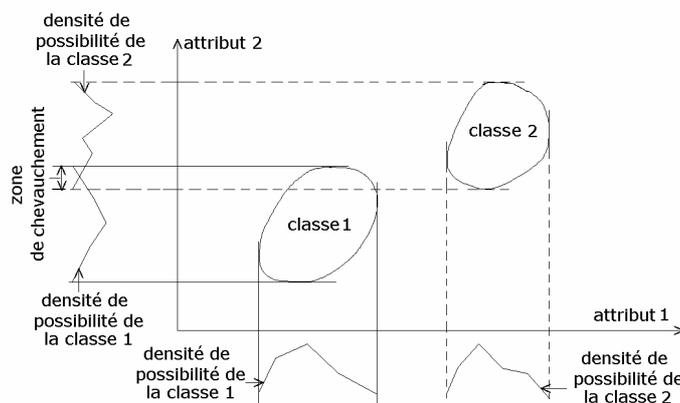


Figure 4.2. Exemple de deux classes séparables par FPM.

Pour remédier à ces inconvénients, nous proposons une nouvelle mesure de séparabilité des classes adaptée à la méthode FPM.

4.5. Mesure de séparabilité adaptée à FPM

Nous cherchons une mesure qui quantifie la zone de chevauchement ou d'intersection entre les densités de possibilités ou de probabilités. Nous choisissons les densités de probabilité afin d'obtenir une mesure de séparabilité comprise entre 0 et 1.

Nous proposons une mesure de séparabilité S , entre c classes, cette mesure est calculée en plusieurs étapes [MOU02f]. La première consiste à déterminer la mesure de séparabilité s_{ji}^k

entre la classe i et la classe j , pour l'attribut k . Les valeurs de probabilité des histogrammes présentés Figure 4.3 sont $p_j^k = (p_{j1}^k, p_{j2}^k, \dots, p_{jm}^k, \dots, p_{jh}^k)$ et $p_i^k = (p_{i1}^k, p_{i2}^k, \dots, p_{im}^k, \dots, p_{ih}^k)$. Chacune des classes possède respectivement un support, ou un intervalle, $[y_{1j}^k, y_{2j}^k]$ et $[y_{1i}^k, y_{2i}^k]$, où la densité de probabilité est supérieure à zéro. Chaque barre m est définie par l'intervalle $[y_{1jm}^k, y_{2jm}^k]$ comme le montre la Figure 4.3.

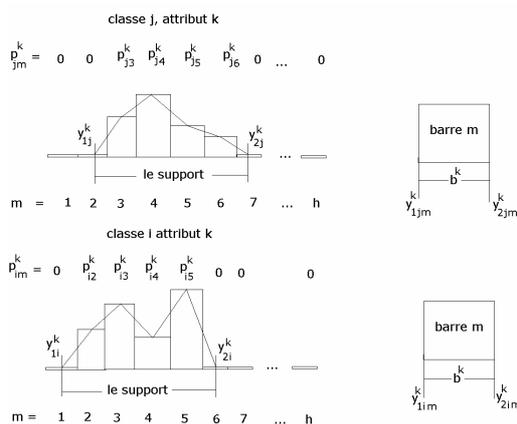


Figure 4.3. Support de la densité de probabilité.

s_{ji}^k est calculée par :

$$s_{ji}^k = 1 - \sum_{m=1}^h I_{im}^k :$$

pour $m = 1 \dots h$:

$$\text{si } y_{1im}^k \geq y_{1j}^k \text{ et } y_{2im}^k \leq y_{2j}^k : I_{im}^k = p_{im}^k$$

$$\text{sinon : si } y_{1im}^k < y_{1j}^k \text{ et } y_{2im}^k > y_{1j}^k : I_{im}^k = \frac{1}{2} \cdot p_{im}^k$$

$$\text{sinon : si } y_{1im}^k < y_{2j}^k \text{ et } y_{2im}^k > y_{2j}^k : I_{im}^k = \frac{1}{2} \cdot p_{im}^k$$

$$\text{sinon : } I_{im}^k = 0$$

(4.18)

Si la dispersion des points est différente pour les deux classes, on a $s_{ji}^k \neq s_{ij}^k$. Il existe alors des zones non chevauchantes entre les deux classes comme le montre la Figure 4.4 . Pour en tenir compte et avoir une seule mesure de séparabilité entre les deux classes j et i , nous avons choisi de prendre la valeur moyenne :

$$S_{ij}^k = S_{ji}^k = \frac{S_{ji}^k + S_{ij}^k}{2} \quad (4.19)$$

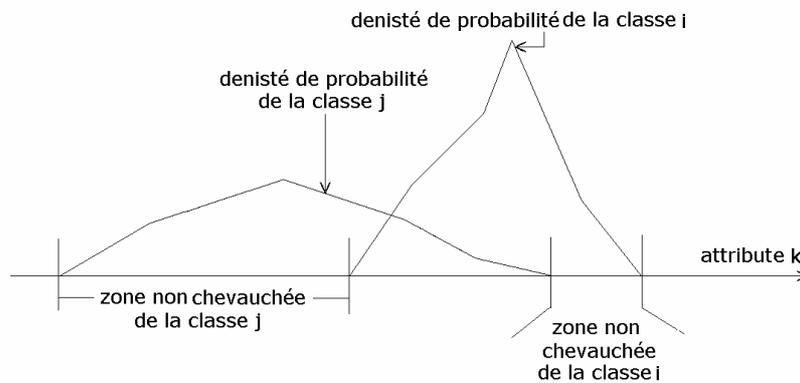


Figure 4.4. Zones non chevauchées entre les classes.

Nous déterminons ensuite la mesure de séparabilité S_{ji} entre la classe j et la classe i pour tous les attributs :

$$\left. \begin{array}{l} \text{si } j \neq i : S_{ji} = \max(S_{ji}^k) \\ \text{sinon : } S_{ii} = 0 \end{array} \right\} k = 1 \dots a \quad (4.20)$$

Nous avons choisi l'opérateur d'agrégation maximum car deux classes sont complément séparables si elles le sont par rapport à au moins un seul attribut. La mesure de séparabilité S_{ji} est égale à 1 si la classe j est totalement séparable de la classe i et égale à 0 si la classe j est totalement superposée à la classe i . S_{ji} est également égale à 0 si $j = i$ puisqu'une classe est complètement superposée à elle-même.

La troisième étape consiste à calculer la mesure de séparabilité de la classe j :

$$S_j = \min_{i=1 \dots c, j \neq i} (S_{ji}) \quad (4.21)$$

Nous avons utilisé l'opérateur minimum pour tenir compte du fait qu'une classe est séparable si elle l'est par rapport à toutes les autres classes.

La dernière étape est le calcul de la mesure globale de séparabilité :

$$0 \leq S = \frac{\sum_{j=1}^c S_j}{c} \leq 1 \quad (4.22)$$

Pour expliquer le calcul de la mesure globale de séparabilité, prenons l'exemple des données Matières Plastiques. La matrice suivante montre les mesures de séparabilité entre les classes résultantes par l'application de FPM avec $h = 2$ et pour chacun des deux attributs :

$$\begin{bmatrix} s_{11}^1 = 0 & s_{12}^1 = 0 & s_{13}^1 = 0 \\ s_{21}^1 = 0 & s_{22}^1 = 0 & s_{23}^1 = 0 \\ s_{31}^1 = 0.46 & s_{32}^1 = 0.12 & s_{33}^1 = 0 \\ s_{11}^2 = 0 & s_{12}^2 = 0 & s_{13}^2 = 0 \\ s_{21}^2 = 0.46 & s_{22}^2 = 0 & s_{23}^2 = 0.46 \\ s_{31}^2 = 0 & s_{32}^2 = 0 & s_{33}^2 = 0 \end{bmatrix} \quad (4.23)$$

En regardant la Figure 4.5, nous pouvons voir le chevauchement entre les densités de possibilité.

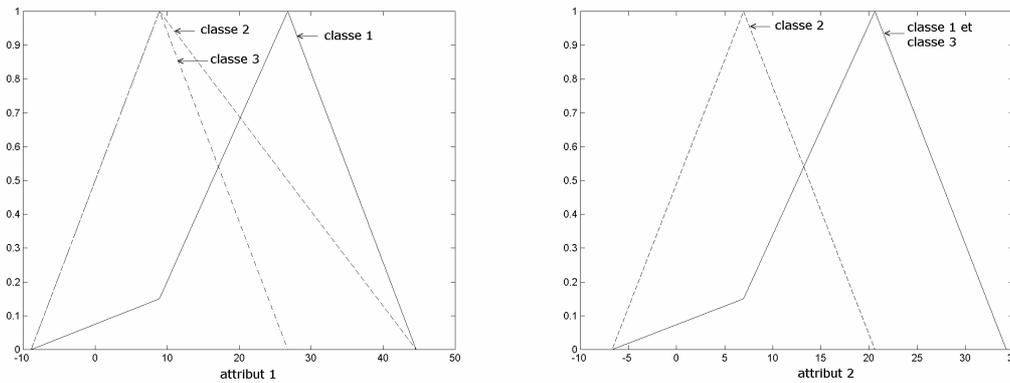


Figure 4.5. Densités de possibilité pour les données Matières Plastiques avec $h = 2$.

Nous remarquons que pour l'attribut 1, la classe 1 couvre complètement la classe 3 : $s_{13}^1 = 0$ alors que la classe 3 couvre partiellement la classe 1 $s_{31}^1 = 0.46$. Nous choisissons la valeur moyenne comme indicateur de séparabilité entre ces deux classes $S_{13}^1 = S_{31}^1 = \frac{0 + 0.46}{2} = 0.23$.

En répétant cela pour toutes les autres mesures de tous les attribut, la matrice asymétrique définie par (4.23) devient donc symétrique :

$$\begin{bmatrix} S_{11}^1 = 0 & S_{12}^1 = 0 & S_{13}^1 = 0.23 \\ S_{21}^1 = 0 & S_{22}^1 = 0 & S_{23}^1 = 0.06 \\ S_{31}^1 = 0.23 & S_{32}^1 = 0.06 & S_{33}^1 = 0 \\ S_{11}^2 = 0 & S_{12}^2 = 0.23 & S_{13}^2 = 0 \\ S_{21}^2 = 0.23 & S_{22}^2 = 0 & S_{23}^2 = 0.23 \\ S_{31}^2 = 0 & S_{32}^2 = 0.23 & S_{33}^2 = 0 \end{bmatrix} \quad (4.24)$$

Nous agrégeons les mesures de séparabilité des deux attributs par l'opérateur maximum :

$$\begin{bmatrix} S_{11} = 0 & S_{12} = 0.23 & S_{13} = 0.23 \\ S_{21} = 0.23 & S_{22} = 0 & S_{23} = 0.23 \\ S_{31} = 0.23 & S_{32} = 0.23 & S_{33} = 0 \end{bmatrix} \quad (4.25)$$

La mesure de séparabilité de chaque classe est obtenue par l'agrégation de toutes les mesures de séparabilité entre cette classe et les autres, en utilisant l'opérateur minimum :

$$[S_1 = 0.23 \quad S_2 = 0.23 \quad S_3 = 0.23] \quad (4.26)$$

La mesure globale de séparabilité est obtenue en prenant la valeur moyenne de toutes les mesures de séparabilité de chaque classe :

$$S = 0.23 \quad (4.27)$$

Cette faible valeur de séparabilité indique que les classes sont peu séparables avec $h = 2$. Les densités de possibilité de la Figure 4.5 le confirme.

Répetons maintenant les mêmes calculs mais avec $h = 5$:

$$\begin{bmatrix} s_{11}^1 = 0 & s_{12}^1 = 0.22 & s_{13}^1 = 1 \\ s_{21}^1 = 0.25 & s_{22}^1 = 0 & s_{23}^1 = 0 \\ s_{31}^1 = 1 & s_{32}^1 = 0.77 & s_{33}^1 = 0 \\ s_{11}^2 = 0 & s_{12}^2 = 1 & s_{13}^2 = 0 \\ s_{21}^2 = 1 & s_{22}^2 = 0 & s_{23}^2 = 1 \\ s_{31}^2 = 0 & s_{32}^2 = 1 & s_{33}^2 = 0 \end{bmatrix} \quad (4.28)$$

$$\begin{bmatrix} S_{11}^1 = 0 & S_{12}^1 = 0.24 & S_{13}^1 = 1 \\ S_{21}^1 = 0.24 & S_{22}^1 = 0 & S_{23}^1 = 0.39 \\ S_{31}^1 = 1 & S_{32}^1 = 0.39 & S_{33}^1 = 0 \\ S_{11}^2 = 0 & S_{12}^2 = 1 & S_{13}^2 = 0 \\ S_{21}^2 = 1 & S_{22}^2 = 0 & S_{23}^2 = 1 \\ S_{31}^2 = 0 & S_{32}^2 = 1 & S_{33}^2 = 0 \end{bmatrix} \quad (4.29)$$

$$\begin{bmatrix} S_{11} = 0 & S_{12} = 1 & S_{13} = 1 \\ S_{21} = 1 & S_{22} = 0 & S_{23} = 1 \\ S_{31} = 1 & S_{32} = 1 & S_{33} = 0 \end{bmatrix} \quad (4.30)$$

$$[S_1 = 1 \quad S_2 = 1 \quad S_3 = 1] \quad (4.31)$$

$$S = 1 \quad (4.32)$$

Cette valeur indique que les classes sont complètement séparables avec $h = 5$. En effet, la Figure 4.6 montre que l'attribut 1 permet de séparer la classe 1 de la classe 3 alors que l'attribut 2 permet de séparer la classe 2 des deux autres. En conséquence, les trois classes sont complètement séparables pour FPM.

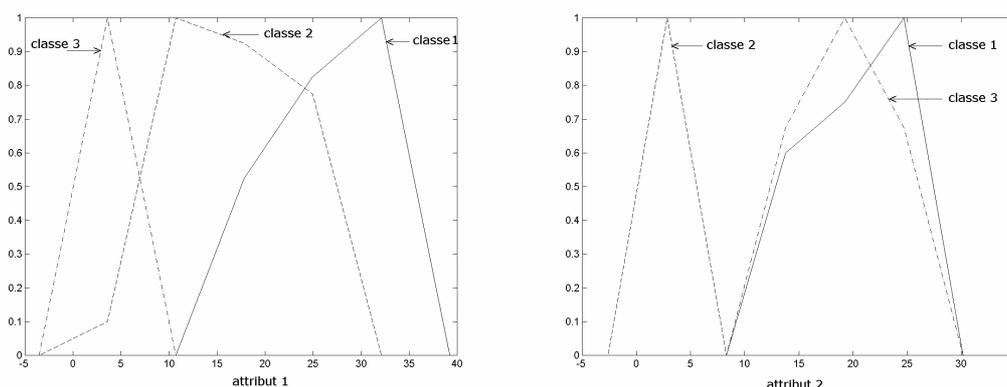


Figure 4.6. Densités de possibilité obtenues avec $h = 5$ pour les données Matières Platiques.

4.6. Zones de rejet dans les classes

Si le nombre de barres h est trop élevé, des barres nulles peuvent apparaître. En effet, le pas de l'histogramme devenant trop petit, certaines barres correspondent aux espaces vides entre les points d'apprentissage. Ces barres nulles causent l'apparition de zones de rejet de points dans les parties centrales des classes ce qui entraîne la dégradation des performances de FPM.

Il faut donc chercher h qui maximise la mesure de séparabilité entre les classes et en même temps ne cause pas l'apparition de zones de rejet. Le nombre de ces zones zr_j^k , par rapport à la classe j suivant l'attribut k , est calculé par :

$$zr_j^k = \sum_{m=u+1}^{v-1} \text{nul}(p_{jm}^k) \text{ avec } \begin{cases} \text{nul}(p_{jm}^k) = 1 \text{ si } p_{jm}^k = 0 \\ \text{nul}(p_{jm}^k) = 0 \text{ si } p_{jm}^k \neq 0 \end{cases} \quad (4.33)$$

p_{ju}^k et p_{jv}^k sont respectivement les probabilités des première et dernière barres du support de l'histogramme. Le nombre de zones de rejet est compris entre 0 et $h - 2$. La Figure 4.7 montre

$$zr_1^1 = 0, zr_1^2 = 0, zr_2^1 = 2, zr_2^2 = 0, zr_3^1 = 0 \text{ et } zr_3^2 = 1$$

Les nombres de zones de rejet pour chaque classe sont les maximums des nombres des zones de rejet par rapport à tous les attributs :

$$zr_1 = 0, zr_2 = 2 \text{ et } zr_3 = 1$$

Enfin, le nombre global de zones de rejet est le maximum des nombres de zones de rejet par rapport à toutes les classes :

$$zr = 2$$

4.7. Exemples

Nous montrons, dans les exemples suivants, la relation entre le nombre de barres h et la mesure globale de séparabilité S . Puis nous montrons la relation entre h et le nombre global de zones de rejet zr . Enfin, nous choisissons h qui maximise d'une part S et d'autre part qui ne cause pas l'apparition de zones de rejet.

4.7.1. Données Matières Plastiques

La Figure 4.8.a montre la relation entre h et la mesure globale de séparabilité S et la Figure 4.8.b montre la relation entre h et le nombre global de zones de rejet zr . Nous constatons que $h = 5$ est la valeur optimale recherchée puisqu'il donne une mesure de séparabilité $S = 1$ et un nombre de zones de rejet $zr = 0$.

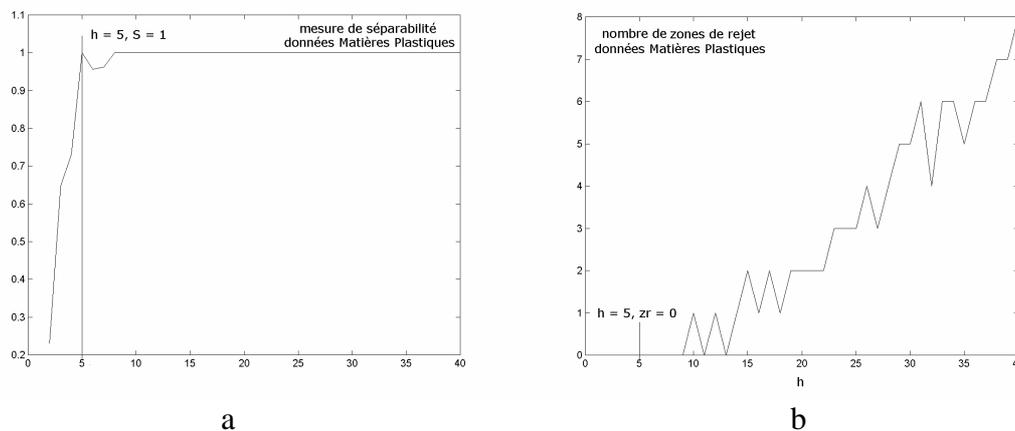


Figure 4.8. Influence du nombre de barres h sur a) la mesure de séparabilité S , et sur b) le nombre de zones de rejet, pour les données Matières Plastiques

4.7.2. Données Machine à Laver

La Figure 4.9.a montre la relation entre h et la mesure globale de séparabilité S et la Figure 4.9.b montre la relation entre h et le nombre global de zones de rejet zr . Nous constatons que $h = 14$ est un bon compromis qui donne une mesure de séparabilité $S = 0.63$ et un nombre de zones de rejet $zr = 0$.

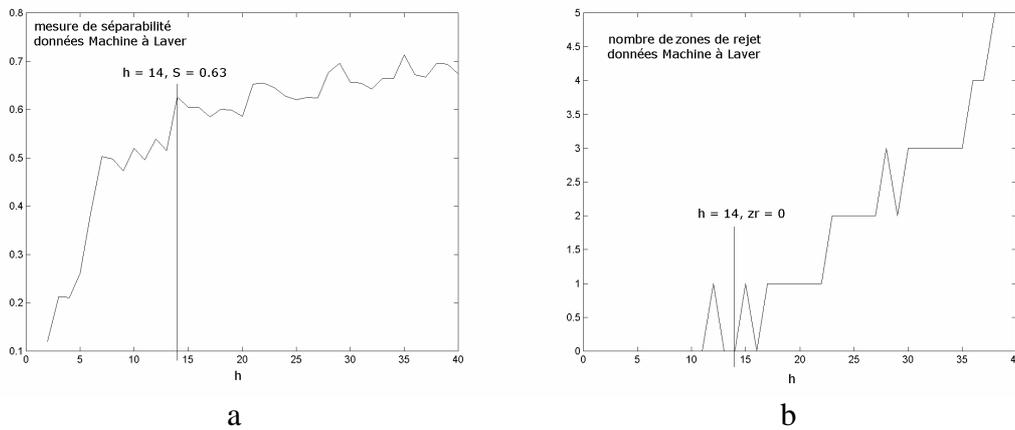


Figure 4.9. Influence du nombre de barres h sur a) la mesure de séparabilité S et sur b) le nombre de zones de rejet, pour les données Machine à Laver

La matrice symétrique des mesures de séparabilité de chaque classe et chaque attribut pour $h = 14$ est :

$$\begin{bmatrix} S_{11}^1 = 0 & S_{12}^1 = 0.03 & S_{13}^1 = 0.96 & S_{14}^1 = 1 \\ S_{21}^1 = 0.03 & S_{22}^1 = 0 & S_{23}^1 = 1 & S_{24}^1 = 1 \\ S_{31}^1 = 0.96 & S_{32}^1 = 1 & S_{33}^1 = 0 & S_{34}^1 = 0.94 \\ S_{41}^1 = 1 & S_{42}^1 = 1 & S_{43}^1 = 0.94 & S_{44}^1 = 0 \\ S_{11}^2 = 0 & S_{12}^2 = 0.28 & S_{13}^2 = 0.01 & S_{14}^2 = 1 \\ S_{21}^2 = 0.28 & S_{22}^2 = 0 & S_{23}^2 = 0.29 & S_{24}^2 = 1 \\ S_{31}^2 = 0.01 & S_{32}^2 = 0.29 & S_{33}^2 = 0 & S_{34}^2 = 0.98 \\ S_{41}^2 = 1 & S_{42}^2 = 1 & S_{43}^2 = 0.98 & S_{44}^2 = 0 \end{bmatrix}$$

Nous constatons que les classes 1 et 2 se chevauchent alors que les classes 3 et 4 sont bien séparables. La Figure 4.10 montre les densités de possibilités pour les quatre classes et les deux attributs. Nous remarquons que la classe 4 est complètement séparable des classes 1 et 2 par les deux attributs : $S_{41}^1 = S_{41}^2 = 1$, et elle se chevauche peu avec la classe 3 : $S_{43}^1 = 0.98$.

Les mesures de séparabilité de chaque classe sont :

$$[S_1 = 0.28 \quad S_2 = 0.28 \quad S_3 = 0.96 \quad S_4 = 0.98]$$

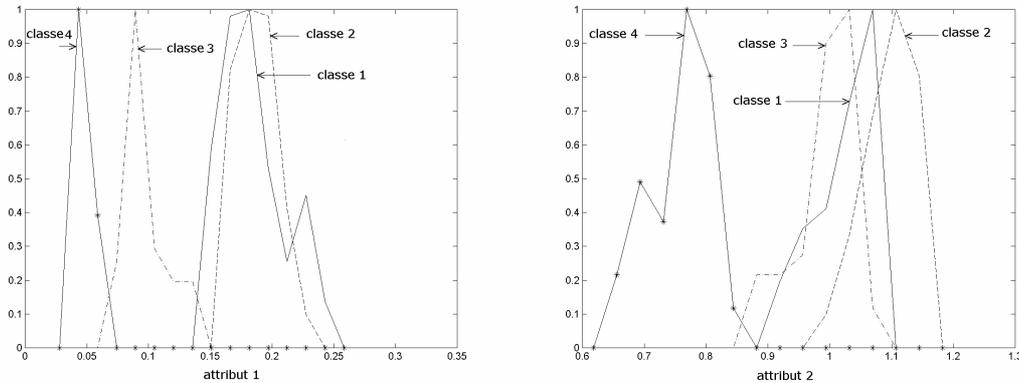


Figure 4.10. Densités de possibilités obtenues avec $h = 14$ et pour les données Machine à Laver

4.7.3. Données Codes Métalliques

La Figure 4.11.a montre la relation entre h et la mesure globale de séparabilité S et la Figure 4.11.b montre la relation entre h et le nombre global de zones de rejet zr . Nous constatons que $h = 12$ est un bon compromis qui donne une mesure de séparabilité $S = 0.57$ et un nombre de zones de rejet $zr = 0$. Une valeur de h plus élevée peut améliorer la mesure globale de séparabilité mais cause l'apparition des zones de rejet.

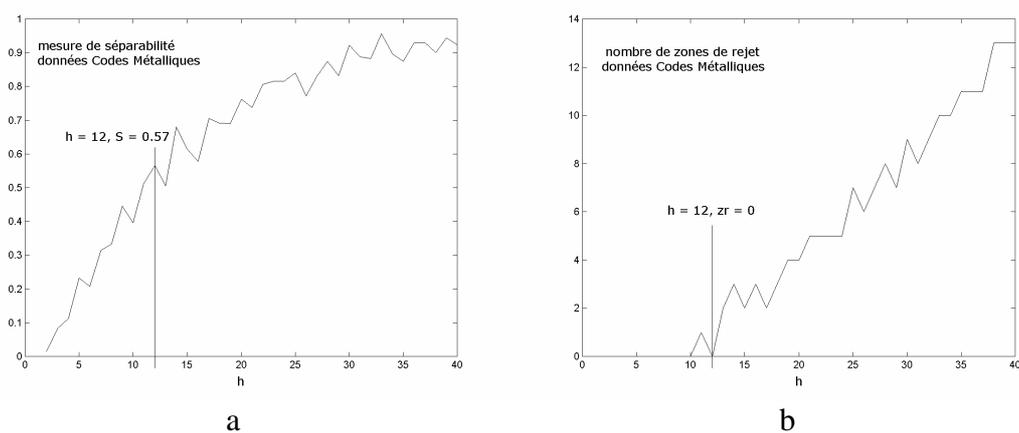


Figure 4.11. Influence du nombre de barres h sur a) la mesure de séparabilité S et sur b) le nombre de zones de rejet, pour les données Codes Métalliques

Les mesures de séparabilité pour chaque classes sont :

$$[S_1 = 1 \quad S_2 = 0.53 \quad S_3 = 0.45 \quad S_4 = 0.45 \quad S_5 = 0.47 \quad S_6 = 0.62 \quad S_7 = 0.45 \quad S_8 = 0.45 \quad S_9 = 0.65]$$

nous constatons que seule la classe 1 est complètement séparable des autres classes qui se chevauchent partiellement.

4.7.4. Données Injection Plastique

La Figure 4.12.a montre la relation entre h et la mesure globale de séparabilité S et la Figure 4.12.b montre la relation entre h et le nombre global de zones de rejet zr . Nous constatons que $h = 5$ est la valeur optimale recherchée puisqu'elle donne une mesure de séparabilité $S = 1$ et un nombre de zones de rejet $zr = 0$.

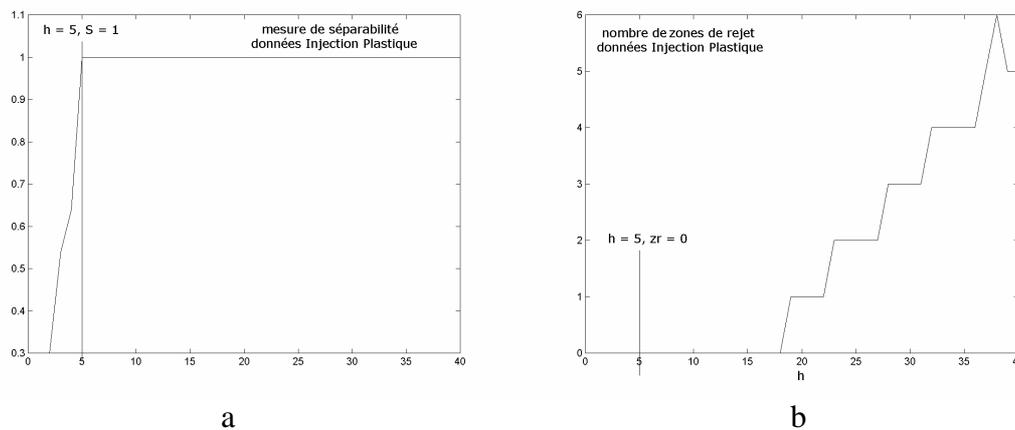


Figure 4.12. Influence du nombre de barres h sur a) la mesure de séparabilité S et sur b) le nombre de zones de rejet, pour les données Injection Plastique

La matrice symétrique des mesures de séparabilité de chaque classe et chaque attribut, pour $h = 5$, est :

$$\begin{bmatrix} S_{11}^1 = 0 & S_{12}^1 = 0.03 & S_{13}^1 = 0.09 & S_{14}^1 = 1 & S_{15}^1 = 0.21 \\ S_{21}^1 = 0.03 & S_{22}^1 = 0 & S_{23}^1 = 0.06 & S_{24}^1 = 1 & S_{25}^1 = 0.19 \\ S_{31}^1 = 0.09 & S_{32}^1 = 0.06 & S_{33}^1 = 0 & S_{34}^1 = 1 & S_{35}^1 = 0.5 \\ S_{41}^1 = 1 & S_{42}^1 = 1 & S_{43}^1 = 1 & S_{44}^1 = 0 & S_{45}^1 = 1 \\ S_{51}^1 = 0.21 & S_{52}^1 = 0.19 & S_{53}^1 = 0.5 & S_{54}^1 = 1 & S_{55}^1 = 0 \\ S_{11}^2 = 0 & S_{12}^2 = 0 & S_{13}^2 = 1 & S_{14}^2 = 0 & S_{15}^2 = 0 \\ S_{21}^2 = 0 & S_{22}^2 = 0 & S_{23}^2 = 1 & S_{24}^2 = 0 & S_{25}^2 = 0 \\ S_{31}^2 = 1 & S_{32}^2 = 1 & S_{33}^2 = 0 & S_{34}^2 = 1 & S_{35}^2 = 1 \\ S_{41}^2 = 0 & S_{42}^2 = 0 & S_{43}^2 = 1 & S_{44}^2 = 1 & S_{45}^2 = 0 \\ S_{51}^2 = 0 & S_{52}^2 = 0 & S_{53}^2 = 1 & S_{54}^2 = 0 & S_{55}^2 = 0 \\ S_{11}^3 = 0 & S_{12}^3 = 1 & S_{13}^3 = 0 & S_{14}^3 = 1 & S_{15}^3 = 1 \\ S_{21}^3 = 1 & S_{22}^3 = 0 & S_{23}^3 = 1 & S_{24}^3 = 0 & S_{25}^3 = 0 \\ S_{31}^3 = 0 & S_{32}^3 = 1 & S_{33}^3 = 0 & S_{34}^3 = 1 & S_{35}^3 = 1 \\ S_{41}^3 = 1 & S_{42}^3 = 0 & S_{43}^3 = 1 & S_{44}^3 = 0 & S_{45}^3 = 1 \\ S_{51}^3 = 1 & S_{52}^3 = 1 & S_{53}^3 = 1 & S_{54}^3 = 1 & S_{55}^3 = 0 \end{bmatrix}$$

nous remarquons, a travers cette matrice, que la classe 1 est totalement séparable de toutes les autres classes, en effet elle est séparable :

- de la classe 4 par rapport à l'attribut 1 : $S_{14}^1 = 1$,
- de la classe 3 par rapport à l'attribut 2 : $S_{13}^2 = 1$,
- des classe 2 et 5 par rapport à l'attribut 3 : $S_{12}^3 = 1$ et $S_{15}^3 = 1$.

Nous pouvons faire le même raisonnement pour les autres classes.

4.7.5. Données Iris

La Figure 4.13.a montre la relation entre h et la mesure globale de séparabilité S et la Figure 4.13.b montre la relation entre h et le nombre global de zones de rejet zr . Nous constatons que $h = 8$ est un bon compromis qui donne une mesure de séparabilité $S = 0.52$ et un nombre de zones de rejet $zr = 0$. Les mesures de séparabilité de chaque classe sont : $[S_1 = 1 \quad S_2 = 0.28 \quad S_3 = 0.28]$. Nous constatons que la classe 1 est totalement séparable alors que les classes 2 et 3 se chevauchent.

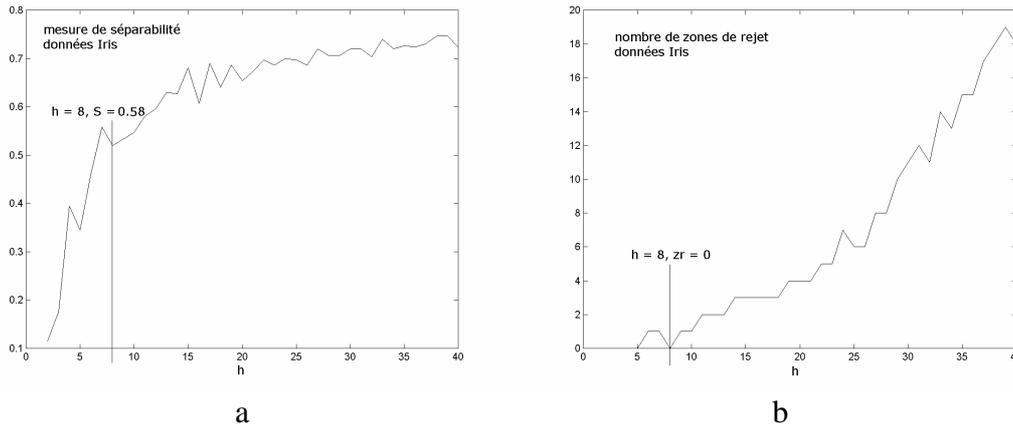


Figure 4.13. Influence du nombre de barres h sur a) la mesure de séparabilité S et sur b) le nombre de zones de rejet, pour les données Iris

4.8. Influence des bornes des histogrammes

Terrell et Scott [TER85] recommandent de considérer les valeurs minimale x_{\min}^k et maximale x_{\max}^k , de l'ensemble d'apprentissage, comme bornes inférieure et supérieure de l'histogramme suivant l'attribut k . Scott [SCO91] considère qu'un histogramme possède une borne inférieure et un nombre infini de barres dont un nombre fini n'est pas nul. Un histogramme est donc totalement déterminé par la largeur des barres b^k et la borne inférieure. Pour éliminer le problème de la détermination de la borne inférieure, Scott construit un nombre Q d'histogrammes : $\hat{f}_1^k, \hat{f}_2^k, \dots, \hat{f}_Q^k$ ayant la même largeur de barre mais différentes bornes inférieures :

$$Y_1^k = 0, Y_1^k = \frac{b^k}{Q}, Y_1^k = \frac{2.b^k}{Q}, \dots, Y_1^k = \frac{(Q-1).b^k}{Q} \quad (4.36)$$

L'histogramme recherché est l'histogramme moyen :

$$\hat{f}^k = \frac{1}{Q} \cdot \sum_{m=1}^Q \hat{f}_m^k \quad (4.37)$$

Pour la méthode FPM, les bornes inférieures et supérieures des histogrammes doivent être fixées. Pour cela et afin d'étudier l'influence des bornes d'histogramme sur les performances de FPM, nous proposons de construire $(Q + 1)$ histogrammes avec un nombre de barres fixe mais avec différentes bornes inférieures Y_1^k et supérieures Y_2^k pour chaque attribut k :

$$\left. \begin{aligned} Y_1^k &= \frac{(Q - q) \cdot x_{\min}^k}{Q} \\ Y_2^k &= \frac{(Q + q) \cdot x_{\max}^k}{Q} \end{aligned} \right\} q = 0 \dots Q \tag{4.38}$$

puis nous calculons la mesure globale de séparabilité et le nombre global de zones de rejet pour chacun de ces histogrammes. Le premier histogramme utilise les valeurs minimale x_{\min}^k et maximale x_{\max}^k de l'ensemble d'apprentissage comme bornes inférieure et supérieure.

Ensuite les bornes inférieures des $q^{i\text{ème}}$ histogrammes sont décalées à gauche de $\frac{q \cdot x_{\min}^k}{Q}$ et les

bornes supérieures sont décalées à droite de $\frac{q \cdot x_{\max}^k}{Q}$. Nous utilisons

$$\frac{(Y_2^k - Y_1^k)_q - (x_{\max}^k - x_{\min}^k)}{x_{\max}^k + x_{\min}^k} = \frac{q}{Q}, q = 0 \dots Q, \text{ comme indicateur du changement des bornes}$$

inférieures et supérieures des histogrammes, afin de visualiser les résultats dans un espace à 2 axes. La Figure 4.14 montre les résultats obtenus pour les données précédentes avec $Q = 5$ et en utilisant les valeurs de h choisies auparavant pour obtenir les meilleures performances de FPM.

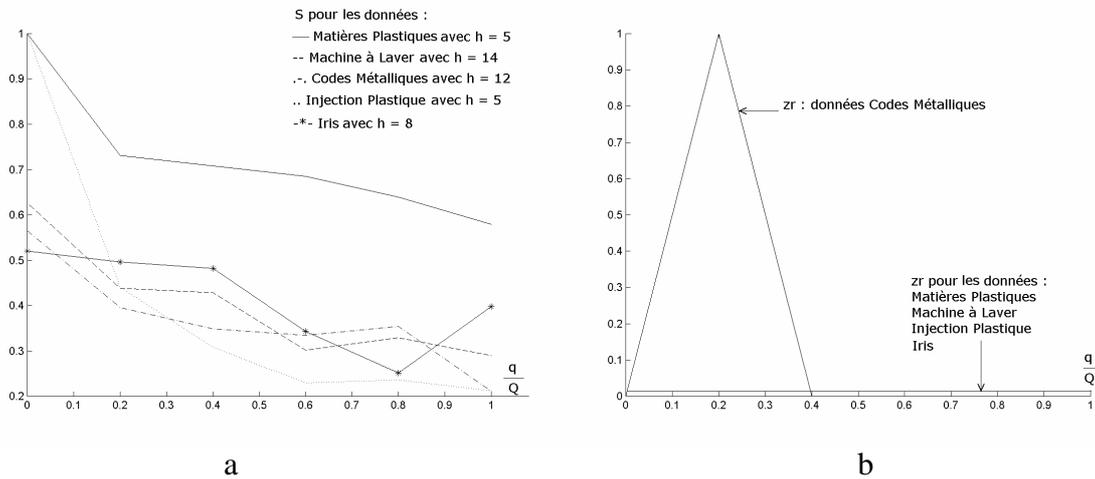


Figure 4.14. Influence des bornes inférieures et supérieures des histogrammes, pour les données précédentes avec $Q = 5$, sur a) la mesure globale de séparabilité S , b) le nombre global de zones de rejet z_r , pour les valeurs de h optimales.

Nous constatons que :

- les bornes inférieures et supérieures des histogrammes influencent la mesure globale de séparabilité et le nombre global de zones de rejet,

- les valeurs minimales et maximales de l'ensemble d'apprentissage, utilisées comme bornes inférieures et supérieures des histogrammes, $q = 0$, donnent les meilleures performances.

4.9. Détermination des paramètres des histogrammes

Deux contraintes sont utilisées pour déterminer la valeur optimale de h :

- le temps de calcul t_{cl} pour la classification d'un nouveau point : ce temps doit être assez faible et constant quelque soit le nombre de points d'apprentissage, pour que la classification soit réalisée en temps réel. Ce temps augmente avec le nombre h de barres des histogrammes,
- le taux de classification T_c qui est l'indicateur de l'efficacité du module de classification.

Les méthodes d'estimation de T_c nécessitent soit la connaissance de la densité de probabilité soit un nombre très grand des points d'apprentissage par rapport au nombre d'attributs. Cela est rarement vérifié dans les applications industrielles. Nous utilisons notre mesure de séparabilité comme une estimation du taux de classification T_c en considérant que tous les points se trouvant dans une zone de chevauchement entre deux classes sont susceptibles d'être mal classés. La mesure global de séparabilité S est donc une estimation pessimiste du taux de classification.

Les valeurs minimales et maximales des points d'apprentissage, suivant chaque attribut, sont considérées respectivement comme bornes inférieures et supérieures des histogrammes puisque d'une part nous ne connaissons pas a priori ces bornes et d'autre part les valeurs minimales et maximales des points d'apprentissage donnent la meilleure séparabilité entre les classes.

La valeur de h est déterminée itérativement. Nous calculons la mesure de séparabilité S pour différents h puis nous choisissons h qui maximise S . La valeur minimale de h est égale à 2 afin de conserver le caractère flou de la méthode FPM. La valeur maximale de h est choisie d'une part en fonction du temps de calcul maximal accepté pour que la classification soit réalisée en temps réel et d'autre part en fonction du nombre de points d'apprentissage.

Si la valeur choisie de h ne donne pas une mesure globale de séparabilité satisfaisante, l'espace de représentation est mis en cause, les paramètres de cet espace ne sont pas assez discriminants pour les classes. Il faut donc chercher d'autres paramètres afin de séparer mieux

les classes et arriver au taux recherché de classification pour le système de RdF.

4.10. Méthode incrémentale pour la mise à jour de la mesure de séparabilité

La mise à jour des densités de possibilité est utile lorsqu'un nouveau point est bien classé, sinon cette mise à jour dégrade les performances de FPM. Nous pouvons utiliser la mesure de séparabilité comme indice de bonne ou de mauvaise classification d'un nouveau point. Si la mesure de séparabilité a diminué après la classification du nouveau point, cela signifie que celui-ci se trouve dans une zone de chevauchement. Nous considérons donc ce point comme mal classé et nous ne l'utilisons pas pour la mise à jour des densités de possibilité. Les points qui augmentent la mesure de séparabilité sont considérés comme des points apportant de l'information utile.

Nous proposons une approche incrémentale pour la mise à jour de la mesure de séparabilité après la classification de chaque nouveau point. Le temps de mise à jour est faible, constant et indépendant de la taille de l'ensemble d'apprentissage.

Pour expliquer cette approche incrémentale, considérons l'histogramme de probabilité $p_j^k = (p_{j1}^k, p_{j2}^k, \dots, p_{jm}^k, \dots, p_{jh}^k)$ de la classe j suivant l'attribut k dont le support est $[y_{1j}^k, y_{2j}^k]$. Les bornes de chaque barre m sont y_{1jm}^k et y_{2jm}^k , la largeur de chaque barre $b^k = y_{2jm}^k - y_{1jm}^k$ est constante par rapport à chaque attribut k . Les densités de probabilité sont mises à jour après la classification du nouveau point \underline{x} dans la classe j [MOU02a] :

$$p_{jm}'^k = \frac{N_j}{N_j + 1} \cdot p_{jm}^k + \frac{1}{N_j + 1} \quad \text{si } x^k \in [y_{1jm}^k, y_{2jm}^k],$$

$$p_{jm}'^k = \frac{N_j}{N_j + 1} \cdot p_{jm}^k \quad \text{sinon}$$
(4.39)

De l'équation (4.18), la mesure de séparabilité s_{ij}^k entre la classe i et la classe j , suivant l'attribut k , est calculée par :

$$s_{ij}^k = 1 - \left(\frac{1}{2} \cdot p_{ju}^k + p_{j(u+1)}^k + \dots + p_{jz}^k + \dots + \frac{1}{2} \cdot p_{jv}^k \right)$$
(4.40)

où la barre u vérifie : $y_{1ju}^k = y_{1i}^k - \frac{b^k}{2}$ et la barre v vérifie $y_{2jv}^k = y_{2i}^k + \frac{b^k}{2}$. En revanche la

barre z vérifie : $\forall z : u < z < v : y_{1jz}^k > y_{1i}^k$ et $y_{2jz}^k < y_{2i}^k$ comme le montre la Figure 4.15.

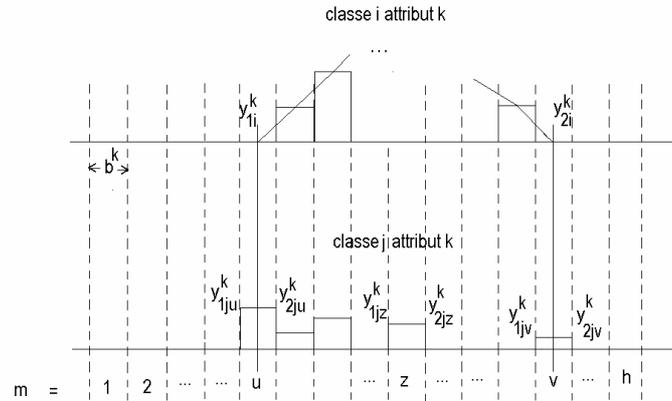


Figure 4.15. Calcul de la séparabilité entre la classe i et la classe j .

Si après la classification du nouveau point celui-ci se trouve dans la barre m de l'attribut k de la classe j , la nouvelle mesure de séparabilité $s_{ij}^{\prime k}$ est calculée selon un de 3 cas suivants :

- si la barre m est la barre u ou la barre v , alors $s_{ij}^{\prime k}$ est :

$$s_{ij}^{\prime k} = 1 - \left\{ \frac{N_j}{N_j + 1} \cdot \left(\frac{P_{jm}^k}{2} + p_{j(m+1)}^k + \dots + \frac{P_{jv}^k}{2} \right) + \frac{1}{2} \cdot \frac{1}{N_j + 1} \right\} \Rightarrow$$

$$s_{ij}^{\prime k} = 1 - \left\{ \frac{N_j}{N_j + 1} \cdot (1 - s_{ij}^k) + \frac{1}{2} \cdot \frac{1}{N_j + 1} \right\} \Rightarrow$$

$$s_{ij}^{\prime k} = \frac{N_j}{N_j + 1} \cdot s_{ij}^k + \frac{1}{2} \cdot \frac{1}{N_j + 1} \quad (4.41)$$

- Si la barre m est la barre z , c'est-à-dire que cette barre se trouve complètement dans le support de la classe i , alors $s_{ij}^{\prime k}$ est :

$$s_{ij}^{\prime k} = 1 - \left\{ \frac{N_j}{N_j + 1} \cdot \frac{P_{ju}^k}{2} + \dots + \left(\frac{N_j}{N_j + 1} \cdot p_{jm}^k + \frac{1}{N_j + 1} \right) + \dots + \frac{N_j}{N_j + 1} \cdot \frac{P_{jv}^k}{2} \right\} \Rightarrow$$

$$s_{ij}^{\prime k} = 1 - \left\{ \frac{N_j}{N_j + 1} \cdot \left(\frac{P_{ju}^k}{2} + \dots + p_{jm}^k + \dots + \frac{P_{jv}^k}{2} \right) + \frac{1}{N_j + 1} \right\} \Rightarrow$$

$$s'_{ij}{}^k = 1 - \left\{ \frac{N_j}{N_j + 1} \cdot (1 - s_{ij}^k) + \frac{1}{N_j + 1} \right\} \Rightarrow$$

$$s'_{ij}{}^k = \frac{N_j}{N_j + 1} \cdot s_{ij}^k \quad (4.42)$$

- Si toute la barre m du nouveau point se trouve à l'extérieur du support de la classe i ,

$s'_{ij}{}^k$ peut être calculée de la manière suivante :

$$s'_{ij}{}^k = 1 - \left(\frac{N_j}{N_j + 1} \cdot \frac{p_{ju}^k}{2} + \dots + \frac{N_j}{N_j + 1} \cdot p_{jm}^k + \dots + \frac{N_j}{N_j + 1} \cdot \frac{p_{jv}^k}{2} \right) \Rightarrow$$

$$s'_{ij}{}^k = 1 - \left(\frac{N_j}{N_j + 1} \cdot (1 - s_{ij}^k) \right) \Rightarrow$$

$$s'_{ij}{}^k = \frac{N_j}{N_j + 1} \cdot s_{ij}^k + \frac{1}{N_j + 1} \quad (4.43)$$

Si le nouveau point entraîne une évolution simple de la classe j , les mesures de séparabilité

$s'_{ji}{}^k$ entre la classe j et la classe i doivent également être mises à jour. Nous avons deux cas :

- cas d'évolution simple du côté gauche : la nouvelle mesure de séparabilité $s'_{ji}{}^k$ est :

$$\forall x^k \in [y_{1jm}^k, y_{2jm}^k[: \text{si } y_{1j}^k > y_{1i}^k \text{ et } y_{1jm}^k = y_{1j}^k - \frac{b^k}{2} \Rightarrow$$

$$s'_{ji}{}^k = 1 - \left\{ \frac{1}{2} \cdot p_{i(m-1)}^k + p_{im}^k + p_{i(m+1)}^k + \dots + \frac{1}{2} \cdot p_{iv}^k \right\} \Rightarrow$$

$$s'_{ji}{}^k = \left\{ 1 - \left(\frac{1}{2} \cdot p_{im}^k + p_{i(m+1)}^k + \dots + \frac{1}{2} \cdot p_{iv}^k \right) \right\} + \frac{1}{2} \cdot (p_{i(m-1)}^k + p_{im}^k) \Rightarrow$$

$$s'_{ji}{}^k = 1 - (1 - s_{ji}^k) + \frac{1}{2} \cdot (p_{i(m-1)}^k + p_{im}^k) \Rightarrow$$

$$s'_{ji}{}^k = s_{ji}^k - \frac{1}{2} \cdot (p_{i(m-1)}^k + p_{im}^k) \quad (4.44)$$

- cas d'évolution simple du côté droite : la nouvelle mesure de séparabilité $s'_{ji}{}^k$ est :

$$\begin{aligned}
\forall x^k \in [y_{1jm}^k, y_{2jm}^k[: \text{si } y_{2j}^k < y_{2i}^k \text{ et } y_{2jm}^k = y_{2j}^k + \frac{b^k}{2} \Rightarrow \\
s_{ji}^k &= 1 - \left\{ \frac{1}{2} p_{iu}^k + p_{i(u+1)}^k + \dots + p_{im}^k + \frac{1}{2} p_{i(m+1)}^k \right\} \Rightarrow \\
s_{ji}^k &= \left\{ 1 - \left(\frac{1}{2} p_{iu}^k + p_{i(u+1)}^k + \dots + \frac{1}{2} p_{im}^k \right) \right\} + \frac{1}{2} \cdot (p_{im}^k + p_{i(m+1)}^k) \Rightarrow \\
s_{ji}^k &= 1 - (1 - s_{Ci}^k) + \frac{1}{2} \cdot (p_{im}^k + p_{i(m+1)}^k) \Rightarrow \\
s_{ji}^k &= s_{ji}^k - \frac{1}{2} \cdot (p_{im}^k + p_{i(m+1)}^k) \tag{4.45}
\end{aligned}$$

4.11. Amélioration de la mesure de séparabilité

4.11.1. FPM avec des barres de largeur Adaptative (FPMA)

Auparavant, nous avons calculé la mesure de séparabilité en utilisant des histogrammes avec des barres de largeur constante pour toutes les classes suivant chaque attribut k . Nous proposons d'utiliser FPM avec des barres de largeur Adaptative (FPMA) en fonction de la dispersion des points d'apprentissage dans chaque classe j . Le but est d'améliorer la séparabilité entre les classes en diminuant la valeur de h qui reste constante. Chaque histogramme a ses bornes inférieure Y_{1j}^k et supérieure Y_{2j}^k qui sont choisies comme, respectivement, les valeurs minimale et maximale des points d'apprentissage de la classe j suivant l'attribut k . La largeur de barre b_j^k , pour la classe j et par rapport à l'attribut k , est plus grande dans le cas d'une classe étendue que dans le cas d'une classe étroite comme le montre la Figure 4.16. La largeur des barres s'adapte donc en fonction du détail ou de l'information qu'apportent les points d'apprentissage.

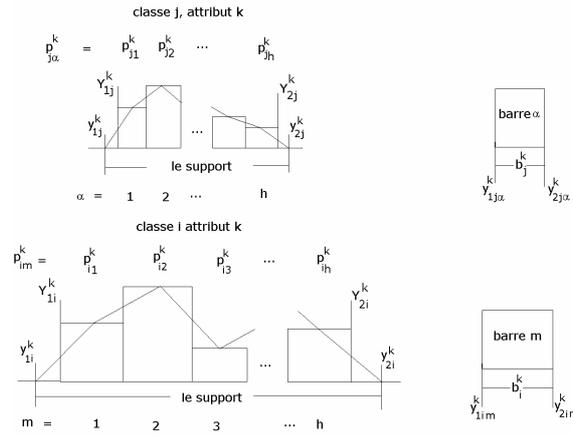


Figure 4.16. Histogrammes avec des barres adaptatives

La méthode FPMA augmente la mesure de séparabilité entre les classes en limitant l'intervalle de variation de chaque classe. La mesure de séparabilité s_{ji}^k entre la classe j et la classe i par rapport à l'attribut k est calculée par :

$$\begin{aligned}
 s_{ji}^k &= 1 - \sum_{m=1}^h I_{im}^k : \\
 &\text{si } y_{1im}^k \geq y_{1j}^k \text{ et } y_{2im}^k \leq y_{2j}^k : I_{im}^k = p_{im}^k \\
 &\text{sinon : si } y_{1im}^k < y_{1j}^k \text{ et } y_{2im}^k > y_{1j}^k \text{ et } y_{2im}^k < y_{2j}^k : I_{im}^k = \left(\frac{y_{2im}^k - y_{1j}^k}{b^k} \right) p_{im}^k \\
 &\text{sinon : si } y_{1im}^k < y_{2j}^k \text{ et } y_{2im}^k > y_{1j}^k \text{ et } y_{2im}^k > y_{2j}^k : I_{im}^k = \left(\frac{y_{1im}^k - y_{2j}^k}{b^k} \right) p_{im}^k \\
 &\text{sinon : si } y_{1im}^k < y_{1j}^k \text{ et } y_{2im}^k > y_{2j}^k : I_{im}^k = \left(\frac{y_{2j}^k - y_{1j}^k}{b_i^k} \right) p_{im}^k \\
 &\text{sinon : } I_{im}^k = 0
 \end{aligned} \tag{4.46}$$

Le calcul de la mesure globale de séparabilité S se fait de la même manière que pour la méthode FPM avec des barres de largeur fixe (FPM).

La section suivante compare les résultats obtenus par la méthode FPMA et ceux obtenus par la méthode FPM.

4.11.2. Résultats expérimentaux

4.11.2.1. Données Matières Plastiques

La Figure 4.17.a compare la mesure globale de séparabilité S obtenue par la méthode FPMA et celle obtenue par la méthode FPM. La Figure 4.17.b compare les nombres globaux

de zones de rejet zr obtenus par les deux méthodes. Nous constatons qu'avec FPMA, $h = 2$ est suffisant pour donner une mesure globale de séparabilité $S = 1$ et un nombre global de zones de rejet $zr = 0$ tandis qu'avec FPM, le meilleur compromis était $h = 5$.

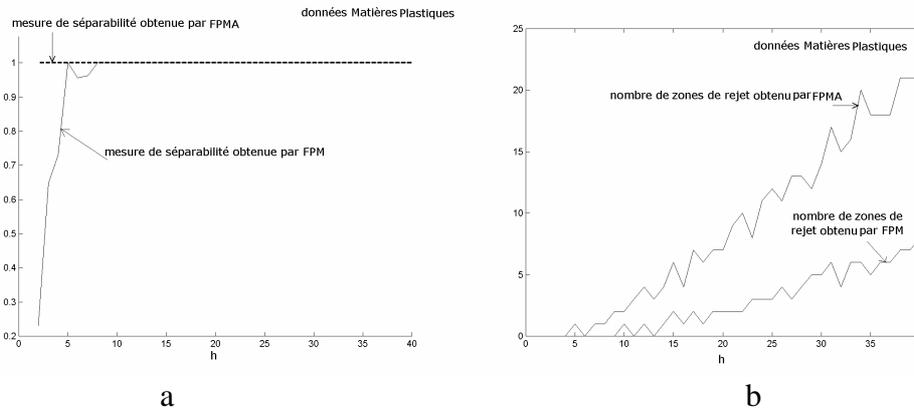


Figure 4.17. Comparaison entre FPMA et FPM par rapport a) à la mesure globale de séparabilité b) au nombre global de zones de rejet, pour les données Matières Plastiques

4.11.2.2. Données Machine à Laver

La Figure 4.18.a compare la mesure globale de séparabilité S obtenue par la méthode FPMA et celle obtenue par la méthode FPM. La Figure 4.18.b compare les nombres globaux de zones de rejet zr obtenus par les deux méthodes. Nous constatons qu'avec FPMA, $h = 6$ est suffisant pour donner une mesure globale de séparabilité $S = 0.83$ et un nombre global de zones de rejet $zr = 0$ alors qu'avec FPM, le meilleur compromis était $h = 14$ qui donnait une mesure globale de séparabilité $S = 0.63$.

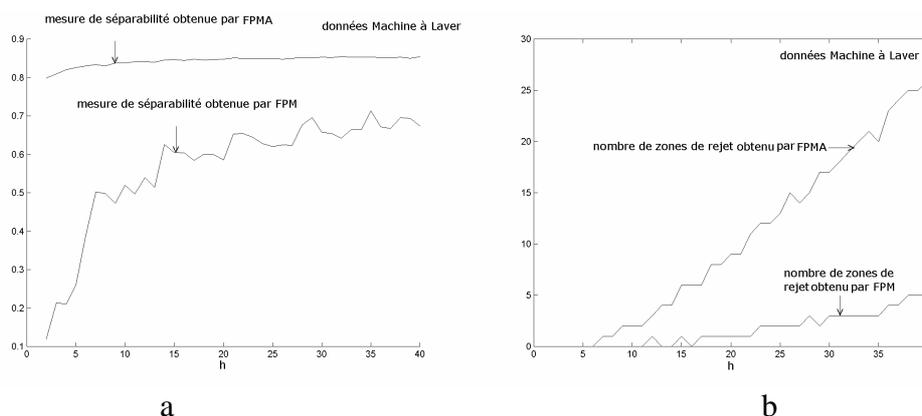


Figure 4.18. Comparaison entre FPMA et FPM par rapport a) à la mesure de séparabilité b) au nombre des zones de rejet, pour les données Machine à Laver

4.11.2.3. Données Codes Métalliques

La Figure 4.19.a compare la mesure globale de séparabilité S obtenue par la méthode FPMA et celle obtenue par la méthode FPM. La Figure 4.19.b compare les nombres globaux de zones de rejet zr obtenus par les deux méthodes. Nous constatons qu'avec FPMA, $h = 3$ est suffisant pour donner une mesure globale de séparabilité $S = 0.98$ et un nombre global de zones de rejet $zr = 0$. En revanche, avec FPM, le meilleur compromis était $h = 12$ qui donnait une mesure globale de séparabilité $S = 0.57$.

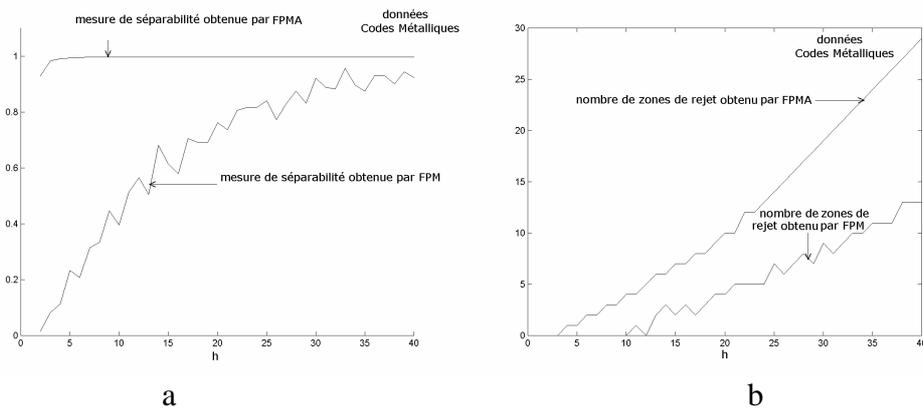


Figure 4.19. Comparaison entre FPMA et FPM par rapport a) à la mesure globale de séparabilité b) au nombre global de zones de rejet, pour les données Codes Métalliques

4.11.2.4. Données injection plastique

La Figure 4.20.a compare la mesure globale de séparabilité S obtenue par la méthode FPMA et celle obtenue par la méthode FPM. La Figure 4.20.b compare les nombres globaux de zones de rejet zr obtenus par les deux méthodes. Nous constatons qu'avec FPMA, $h = 2$ est suffisant pour donner une mesure globale de séparabilité $S = 1$ et un nombre global de zones de rejet $zr = 0$ alors qu'avec FPM, le meilleur compromis était $h = 5$ qui donnait une mesure globale de séparabilité $S = 1$.

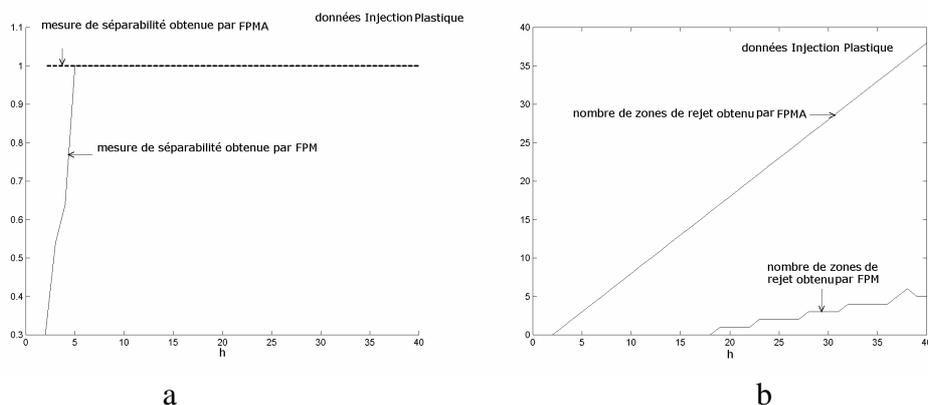


Figure 4.20. Comparaison entre FPMA et FPM par rapport a) à la mesure globale de séparabilité b) au nombre de zones de rejet, pour les données Injection Plastique

4.11.2.5. Données Iris

La Figure 4.21.a compare la mesure globale de séparabilité S obtenue par la méthode FPMA et celle obtenue par la méthode FPM. La Figure 4.21.b compare les nombres globaux de zones de rejet z_r obtenus par les deux méthodes. Nous constatons qu'avec FPMA, $h = 6$ est suffisant pour donner une mesure globale de séparabilité $S = 0.86$ et un nombre global de zones de rejet $z_r = 0$. En revanche, avec FPM, le meilleur compromis était $h = 8$ qui donnait une mesure globale de séparabilité $S = 0.58$.

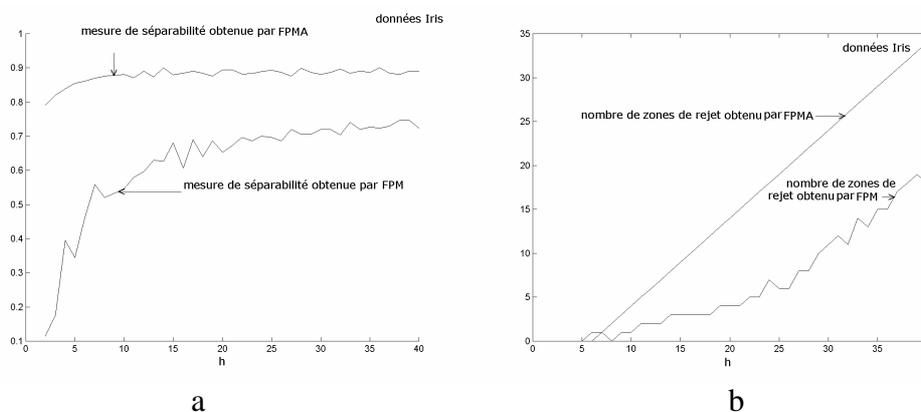


Figure 4.21. Comparaison entre FPMA et FPM par rapport a) à la mesure globale de séparabilité b) au nombre global de zones de rejet, pour les données Iris

Nous constatons que la méthode FPMA améliore la séparabilité en diminuant le nombre de barre h . En conséquence elle améliore le taux de classification et diminue le temps de calcul.

Lors de l'utilisation de FPM, les bornes des histogrammes sont les valeurs minimales et maximales de l'ensemble d'apprentissage, en conséquence les histogrammes sont plus étendus que les classes et comportent des barres nulles à l'extérieur de celles-ci. Ces barres nulles permettent de tenir compte de l'évolution simple. En revanche avec FPMA ces barres nulles n'existent pas, un nouveau point qui se trouve dans le support d'une classe j mais à l'extérieur de la première ou la dernière barres, pour un attribut k , ne peut pas être pris en compte dans la mise à jour des histogrammes. Il faut alors refaire un apprentissage de l'histogramme en prenant le nouveau point comme nouvelle borne inférieure ou supérieure.

4.12. Méthode incrémentale pour la mise à jour de la mesure de séparabilité pour FPMA

Si le nouveau point est classé dans la classe j , seules les mesures de séparabilité

s_{ij} , $i = 1 \dots c$ doivent être mises à jour, les mesures de séparabilité s_{ji} ne changent pas parce que la méthode FPMA n'est pas capable de mettre à jour les histogrammes de probabilité en cas d'évolution simple. Le raisonnement utilisé pour le calcul incrémental des nouvelles mesures de séparabilité dans le cas de FPMA est le même que pour le cas de FPM.

Considérons que la barre du nouveau point classifié dans la classe j par rapport à l'attribut k est m , la nouvelle mesure de séparabilité s'_{ij} est calculée par :

$$\text{Si } y_{2jm}^k \leq y_{1i}^k \text{ ou } y_{1jm}^k \geq y_{2i}^k \Rightarrow s'_{ij} = \frac{N_j}{N_j + 1} \cdot s_{ij}^k + \frac{1}{N_j + 1} \quad (4.47)$$

$$\text{Si } y_{1jm}^k \geq y_{1i}^k \text{ et } y_{2jm}^k \leq y_{2i}^k \Rightarrow s'_{ij} = \frac{N_j}{N_j + 1} \cdot s_{ij}^k \quad (4.48)$$

$$\text{Si } y_{1jm}^k < y_{1i}^k \text{ et } y_{2jm}^k > y_{1i}^k \text{ et } y_{2jm}^k < y_{2i}^k \Rightarrow s'_{ij} = \frac{N_j}{N_j + 1} \cdot s_{ij}^k + \left(1 - \frac{(y_{2jm}^k - y_{1i}^k)}{b_j^k}\right) \cdot \frac{1}{N_j + 1} \quad (4.49)$$

$$\text{Si } y_{1jm}^k < y_{2i}^k \text{ et } y_{2jm}^k > y_{2i}^k \text{ et } y_{1jm}^k > y_{1i}^k \Rightarrow s'_{ij} = \frac{N_j}{N_j + 1} \cdot s_{ij}^k + \left(1 - \frac{(y_{2i}^k - y_{1jm}^k)}{b_j^k}\right) \cdot \frac{1}{N_j + 1} \quad (4.50)$$

$$\text{Si } y_{1jm}^k \leq y_{1i}^k \text{ et } y_{2jm}^k \geq y_{2i}^k \Rightarrow s'_{ij} = \frac{N_j}{N_j + 1} \cdot s_{ij}^k + \left(1 - \frac{(y_{2i}^k - y_{1i}^k)}{b_j^k}\right) \cdot \frac{1}{N_j + 1} \quad (4.51)$$

L'équation (4.47) montre le cas où toute la barre m , du nouveau point \underline{x} classifié dans la classe j , se trouve complètement dans le support de la classe i . En revanche les équations (4.49), (4.50) et (4.51) montre le cas où une partie de la barre m se trouve dans le support de la classe i . L'équation (4.48) montre le cas où la barre m se trouve complètement à l'extérieur de la classe i .

L'augmentation maximale Δ_1 de la mesure de séparabilité s_{ij}^k , après la classification du nouveau point \underline{x} dans la classe j , a lieu dans le cas où toute la barre m du point \underline{x} se trouve dans le support de la classe i , c'est-à-dire pour l'équation (4.47) :

$$\Delta_1 = s'_{ij} - s_{ij}^k = \frac{1 - s_{ij}^k}{N_j + 1} \quad (4.52)$$

La diminution maximale Δ_2 de la mesure de séparabilité s_{ij}^k a lieu dans le cas où toute la barre m se trouve à l'extérieur de la classe i , c'est-à-dire pour l'équation (4.48) :

$$\Delta_2 = s_{ij}^k - s_{ij}^k = \frac{-s_{ij}^k}{N_j + 1} \quad (4.53)$$

En conséquence, la variation positive ou négative de la mesure de séparabilité, après la classification d'un nouveau point, est bornée par :

$$\frac{-s_{ij}^k}{N_j + 1} \leq s_{ij}^k - s_{ij}^k \leq \frac{1 - s_{ij}^k}{N_j + 1} \quad (4.54)$$

cette inégalité est également valable pour la méthode FPM avec des barres de largeur fixe.

4.13. Conclusion

La fonction d'appartenance de FPM, est construite en utilisant des histogrammes de probabilité. Chaque histogramme est défini par deux paramètres : le pas représenté par le nombre de barres et la largeur de l'intervalle représentée par les bornes inférieure et supérieure. Actuellement, il n'existe aucune méthode pour la détermination de ces paramètres, ils sont déterminés expérimentalement ou heuristiquement.

Les performances d'une méthode de classification augmentent avec la séparabilité entre les classes. Par conséquent, nous cherchons les paramètres d'histogramme qui maximisent une mesure de séparabilité entre les classes.

Il existe plusieurs mesures de séparabilité dans la littérature. Ces mesures sont basées sur le principe de distance entre les centroïdes des classes et les points de l'ensemble d'apprentissage. En conséquence elles ne font pas intervenir les paramètres des histogrammes. De plus, elles ne sont pas adaptées au principe de la méthode FPM qui consiste à considérer que deux classes sont complètement séparables si leur densités de probabilité, ou de possibilité, sont complètement séparables. Pour cette raison, nous avons proposé une mesure de séparabilité S adaptée à la méthode FPM. Cette mesure est basée sur le chevauchement entre les différents histogrammes de probabilité. Le temps de calcul est donc constant et faible. Nous cherchons les paramètres d'histogrammes qui maximisent cette mesure et en conséquence qui maximisent les performances de FPM.

Nous avons proposé également d'utiliser la méthode FPM avec des barres de largeur

adaptative en fonction de la dispersion des points dans chaque classe, suivant chaque attribut. Le but est d'améliorer la mesure de séparabilité entre les classes en diminuant le temps de calcul. Enfin nous avons développé deux approches incrémentales pour la mise à jour de la mesure de séparabilité après la classification de chaque nouveau point. La nouvelle valeur de la mesure de séparabilité pourrait être utilisée comme indicateur de bonne ou de mauvaise classification d'un nouveau point. Si le nouveau point classifié augmente la mesure de séparabilité, il est bien classifié et il doit être utilisé pour la mise à jour des histogrammes de probabilité.

5. Elimination du bruit et détection des zones de haute densité

5.1. Introduction

Dans le contexte du diagnostic par RdF, l'effet du bruit se manifeste par l'apparition de points aberrants dans l'espace de représentation. Ces points dégradent les performances de la méthode FPM par l'élargissement du champ de classification. Ils sont acceptés dans des classes au lieu d'être rejetés.

D'autre part, un mode de fonctionnement peut avoir plusieurs sous-modes. Ces derniers sont interprétés comme des sous-classes, ou des zones de haute densité de points, dans une classe. Classifier une observation dans une de ces sous-classes permet de déterminer plus précisément le fonctionnement d'un système à un instant donné.

Dans ce chapitre, nous montrons qu'un seuil paraît nécessaire pour éliminer les points aberrants ou détecter les zones de haute densité. Dans un premier temps, ce seuil est fixé par l'utilisateur. La méthode FPM utilisant la Transformation Variable (TV) donne de bons résultats en perdant un minimum d'information. Dans un second temps, nous introduisons une approche pour déterminer ce seuil automatiquement. Nous comparons les performances de FPM utilisant la TV avec celles obtenues par FPM utilisant les transformations courantes. Enfin nous appliquons FPM avec la TV à la surveillance d'une ligne de peinture au trempé de foyers de chaudières [MOU02e].

5.2. Paramètre k et détection des zones de haute densité

5.2.1. Nécessité d'un seuil

Les transformations vérifiant la condition de cohérence donnent des valeurs de possibilité plus élevées que les valeurs correspondantes de probabilité. Or, en diagnostic, nous avons besoin d'une fonction d'appartenance qui donne des valeurs faibles pour des points se situant dans des espaces de faible densité. Cela a pour but :

- de détecter les points aberrants,
- de détecter les zones de haute densité qui correspondent à des sous-classes ou à des sous-modes de fonctionnement.

Nous voulons donc une transformation qui donne une valeur de possibilité élevée pour les

points ayant une grande valeur de probabilité et une valeur de possibilité plus faible que celle de la probabilité, voire même nulle, lorsque celle-ci est basse. Cela peut être avantageusement réalisé par la TV en jouant sur la valeur du paramètre k .

Pour éliminer les points aberrants ou distinguer les zones de haute-densité, il faut utiliser un seuil de probabilité ou de possibilité. En choisissant un seuil de probabilité, la distribution de probabilité devient sous-normalisée, ce qui donne également une distribution de possibilité sous-normalisée, lorsqu'on utilise les transformations de Dubois et Prade classique et optimale [MOU02b]. Cela diminue la zone floue des classes. Les autres transformations donnent une distribution de possibilité normalisée. La TV est plus spécifique que les transformations normalisée et de Klir et en conséquence elle est plus informative [MOU02b]. De plus une transformation plus spécifique fournit une meilleure distinction entre les cas très proches correspondants à une décision difficile à prendre. Le choix de la TV pour éliminer le bruit et détecter les zones de haute densité est donc meilleur lorsque l'on utilise un seuil de probabilité.

Le choix d'un seuil de possibilité est d'autant plus facile que la distribution de possibilité, obtenue par une transformation probabilité-possibilité, est spécifique. C'est généralement le cas des densités de possibilité obtenues par la TV, elle est donc la mieux adaptée [MOU02b].

Nous avons préféré utiliser un seuil de probabilité, parce que d'une part les histogrammes de probabilité font mieux ressortir les points aberrants et les zone de haute densité. D'autre part, choisir un seuil de possibilité nécessite plus du temps de calcul pour faire la transformation de toutes les valeurs de probabilité.

5.2.2. Rôle du paramètre k dans la Transformation Variable

Une valeur de k élevée renforce la contribution des barres de l'histogramme ayant une haute densité et minimise la contribution de toutes les autres barres [MOU02b]. Une valeur faible de k donne une fonction d'appartenance très tolérante qui met en valeur toutes les barres de l'histogramme.

Pour une certaine valeur de k , la TV se situe entre la transformation classique de Dubois et Prade et la transformation optimale. Quant k diminue, la TV approche la transformation classique de Dubois et Prade. En revanche quand k augmente, la TV approche la transformation optimale. Pour des valeurs élevées de k , la TV devient très spécifique ou sélective et travaille comme un filtre. Mais dans ce cas la TV viole la condition de cohérence pour les petites valeurs de probabilité.

5.2.3. Choix du seuil de probabilité pour déterminer k

Considérons un histogramme de probabilité dont les probabilités de ses barres sont $p = (p_1, p_2, \dots, p_h)$ et dont le nombre de barres de valeur supérieure à zéro est Se_m . Nous déterminons la valeur du k en fonction de nombre de barres Se que la TV doit amplifier. Si les probabilités sont classées en ordre strictement descendant : $p = (p_1 > p_2 > \dots > p_h > 0)$, la valeur de k , correspondant au seuil de probabilité p_{se} , est calculé par :

$$k = \frac{\log p_{Se}}{(1 - p_{Se}) \cdot \log \frac{p_{Se}}{P_{\max}}} \quad 1 < Se \leq Se_m \quad (5.1)$$

Se est déterminé en choisissant les barres de haute probabilité pour à la fois éliminer les bruits autour des classes et se concentrer sur les zones de haute densité. Se se situe dans l'intervalle suivant :

$$1 \leq Se \leq Se_m \quad (5.2)$$

Quand $Se = 1$, la TV permet d'amplifier seulement la barre la plus élevée en lui attribuant la valeur de possibilité maximum qui est égale à 1, les autres barres auront des valeurs de possibilité moins élevées que celles de probabilité. En conséquence la TV est la plus spécifique possible. Lorsque Se augmente, la TV devient moins spécifique en acceptant plus de barres. La TV permet de concentrer la distribution de possibilité obtenue sur la zone la plus dense, en éliminant les barres de probabilité qui vérifient l'inégalité suivante :

$$\pi_m = 0 \text{ si } \pi_m < p_m \quad (5.3)$$

Choisir une valeur trop petite de Se , correspondant à une valeur élevée du seuil de probabilité, entraîne une perte d'information qui se traduit par le rejet de points appartenant à la classe. Le choix judicieux de Se incombe à l'utilisateur. En regardant les histogrammes de probabilité, l'utilisateur peut fixer le seuil en choisissant le nombre de barres qui doivent être conservées et amplifiées par la TV. Nous pouvons remarquer que cette dernière fonctionne comme un filtre dont la bande passante est déterminée par le seuil choisi.

5.2.4. Relation entre spécificité et seuil

Lorsqu'on utilise la TV avec un seuil de probabilité, la spécificité pour les barres de probabilité restantes augmente quand la valeur du seuil choisi augmente. L'augmentation de la spécificité avec la valeur du seuil accroît l'information contenue dans les distributions de

possibilité.

Proposition 1.

La spécificité des distributions de possibilité obtenues par la TV augmente pour les probabilités restantes quand la valeur du seuil augmente.

Preuve

Considérons la distribution de probabilité $p = (p_1, p_2, \dots, p_h)$ où $p_1 > p_2 > \dots > p_h > 0$ et $\pi = (\pi_1, \pi_2, \dots, \pi_h)$ la distribution de possibilité correspondante obtenue avec la TV, si k_h est calculé selon (5.1) avec $p_{se} = p_h$, la distribution de possibilité π correspondante sera :

$$\pi_1 = 1, \pi_2 = \left(\frac{p_2}{p_1}\right)^{k_h \cdot (1-p_2)}, \pi_3 = \left(\frac{p_3}{p_1}\right)^{k_h \cdot (1-p_3)}, \dots, \pi_h = 0$$

si on augmente le seuil : $p_{se} = p_m$ où $1 < m < h$, on a $k_m > k_h$ et la distribution de possibilité π correspondante est :

$$\pi'_1 = 1, \pi'_2 = \left(\frac{p_2}{p_1}\right)^{k_m \cdot (1-p_2)}, \pi'_3 = \left(\frac{p_3}{p_1}\right)^{k_m \cdot (1-p_3)}, \dots,$$

$$\pi'_{m-1} = \left(\frac{p_{m-1}}{p_1}\right)^{k_m \cdot (1-p_{m-1})}, \pi'_m = 0, \dots, \pi'_h = 0$$

comme $k_m > k_h$ on a, pour les probabilités restantes :

$$\pi'_1 = \pi_1 = 1, \pi'_2 < \pi_2, \pi'_3 < \pi_3, \dots, \pi'_{m-1} < \pi_{m-1}$$

cette inégalité signifie que la distribution de possibilité π' est inférieure à π , ce qui prouve, selon la condition de spécificité, que π' est plus spécifique que π pour les probabilités restantes.

Proposition 2

La spécificité des distributions de possibilité obtenues avec la transformation normalisée, pour les probabilités restantes, reste constante quand la valeur du seuil augmente.

Preuve

Supposons que $p = (p_1, p_2, \dots, p_h)$ où $p_1 > p_2 > \dots > p_h > 0$, est une distribution de probabilité et $\pi = (\pi_1, \pi_2, \dots, \pi_h)$ est la distribution de possibilité correspondante en utilisant la transformation normalisée, si le seuil de probabilité choisi est $p_{se} = p_h$, la distribution de

possibilité π est :

$$\pi_1 = \frac{p_1}{p_1} = 1, \pi_2 = \frac{p_2}{p_1}, \pi_3 = \frac{p_3}{p_1}, \dots, \pi_h = 0 \quad (5.4)$$

si on choisi le seuil : $p_s = p_m$ où $1 < m < h$, la distribution de possibilité π' devient :

$$\pi'_1 = \frac{p_1}{p_1} = 1, \pi'_2 = \frac{p_2}{p_1}, \pi'_3 = \frac{p_3}{p_1}, \dots, \pi'_{m-1} = \frac{p_{m-1}}{p_1}, \pi'_m = 0, \dots, \pi'_h = 0 \quad (5.5)$$

en comparant (5.4) et (5.5), pour les probabilités restantes, nous trouvons que :

$$\pi_1 = \pi'_1 = 1, \pi_2 = \pi'_2, \pi_3 = \pi'_3, \dots, \pi_{m-1} = \pi'_{m-1}$$

Cela signifie que la spécificité ne change pas après l'augmentation de la valeur du seuil de probabilité.

Lors de l'utilisation de la TV, nous remarquons que si toutes les barres ont des valeurs égales, cas de l'ignorance totale, l'utilisation d'un seuil n'a pas de sens puisque $Se = 1$ permettra de conserver et amplifier toutes les barres. Ce cas correspond à une classe sans aucune zone de haute densité ou points aberrants. Pour les autres transformations, ce cas est traduit par une distribution nulle qui signifie une perte totale de l'informations.

Le mauvais choix de h détériore les performances de la méthode FPM mais également les performances de la TV, en terme d'élimination du bruit et en terme de détection des sous-classes [MOU02e].

5.3. Détermination automatique de k

Le choix du seuil p_{se} est primordial. Déterminer ce seuil automatiquement, en fonction de la répartition des points dans les classes, nous évite son choix subjectif. S'il n'existe pas de points aberrants ou de sous-classes, la distribution de possibilité obtenue par la TV accepte tous les points de la classe dans la base de données, en choisissant un seuil proche de zéro. Ce seuil augmente de manière à adapter la distribution de possibilité pour qu'elle élimine uniquement les points aberrants et qu'elle se concentre sur les sous-classes.

Nous proposons de calculer la valeur de k pour chaque barre m de chaque attribut i de chaque classe j de la manière suivante :

$$k_m^i = \left(\frac{p_{\max}^i}{p_m^i} \right)^{\left(\frac{p_{\max}^i - p_{\min}^i}{p_m^i} \right)} \quad m=1 \dots h, i=1 \dots a \quad (5.6)$$

Nous constatons que

- k augmente avec la diminution de la probabilité p_m de la barre m ,
- k augmente avec l'augmentation de la différence entre la barre ayant la probabilité la plus élevée et la barre ayant la probabilité la moins élevée.

Cette augmentation valorise les barres de haute densité et atténue l'effet des autres barres. k est beaucoup plus élevé dans le cas des points aberrants ou des sous-classes que dans le cas des classes de répartition uniforme. Cela permet à la TV de fonctionner comme un filtre adaptatif en fonction de la répartition des points dans la classe.

5.4. Exemples

5.4.1. Elimination du bruit

La Figure 5.1 montre un exemple d'une classe entourée de points de bruit ou points aberrants. Nous comparons les performances de FPM en terme de classification en utilisant les transformations précédentes.

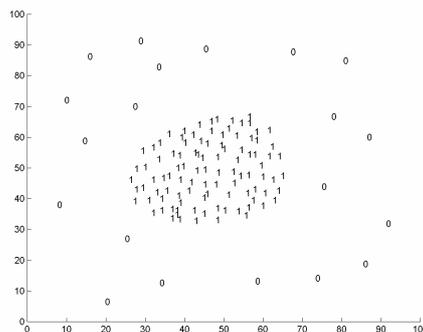


Figure 5.1. Classe entourée de points de bruit ou de points aberrants : 1 points appartenant à la classe 1 et 0 points de bruit.

Le seuil p_{se} peut être déterminé par l'utilisateur en utilisant les histogrammes de probabilité de la Figure 5.2. Nous choisissons $Se = 5$ pour l'attribut 1 et $Se = 4$ pour l'attribut 2, autrement dit nous choisissons les seuils p_5^1 et p_4^2 pour la bande passante de nos filtres lors de l'utilisation de la TV. En effet sur l'histogramme du premier attribut, nous constatons que 5 barres sont dominantes, le nombre de barres dominantes est de 4 pour le deuxième attribut.

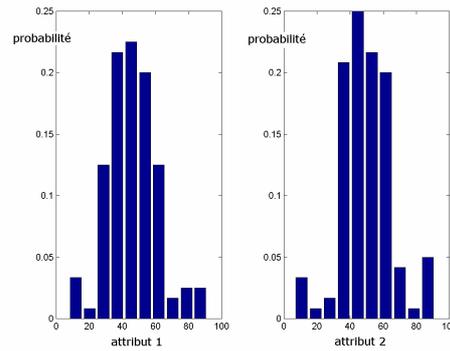


Figure 5.2. Histogrammes de probabilité pour l'exemple de la Figure 5.1.

La Figure 5.3 montre les courbes de densités de possibilité construites par FPM en utilisant les cinq transformations précédentes, avec les seuils p_5^1 et p_4^2 .

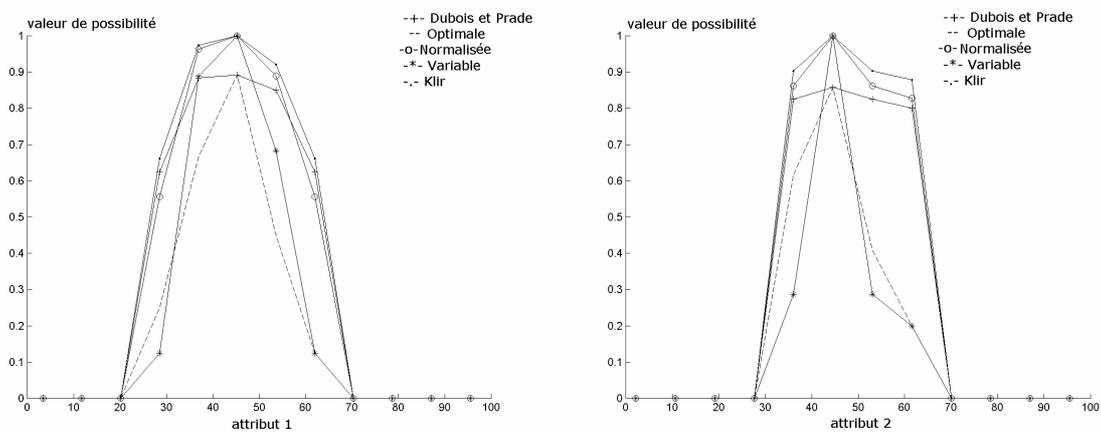


Figure 5.3. Densités de possibilité obtenues par les différentes transformations avec les seuils p_5^1 et p_4^2 .

La Figure 5.4 montre les courbes de niveau d'appartenance obtenues par la TV avec les seuils choisis.

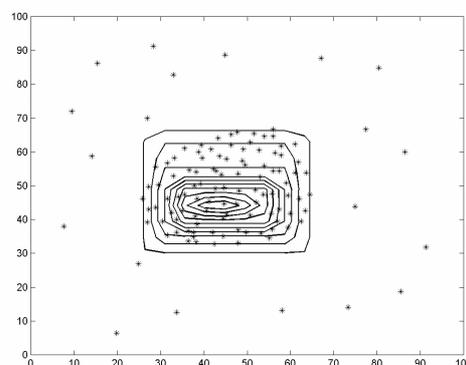


Figure 5.4. Courbes de niveau d'appartenance obtenues en utilisant la TV avec les seuils p_5^1 et p_4^2 : la courbe intérieure correspond au niveau d'appartenance le plus élevé.

Nous remarquons que, d'une part, les deux transformations de Dubois et Prade donnent des densités de possibilité sous-normalisées, contrairement aux autres. D'autre part, la densité de possibilité obtenue par la TV est plus spécifique que les densités obtenues par les transformations de Klir et normalisée. C'est donc la TV qui perd le moins d'information. En effet, la TV amplifie les probabilités dans la bande passante où se trouvent les points appartenant à la classe, et atténue les probabilités à l'extérieur de cette bande où se trouvent les points de bruits ou les points aberrants. Ces résultats sont vérifiés pour n'importe quel jeu de données.

Bien évidemment, si le bruit est tel que le point se retrouve au centre d'une classe qui n'est pas la sienne, la méthode FPM, quelle que soit la transformation utilisée, est incapable de décider que le point est mal classé.

5.4.2. Détection des zones de haute densité

5.4.2.1. Cas d'une seule zone de haute densité

Nous allons étudier l'aptitude des transformations précédentes à détecter les sous-classes. La Figure 5.5 montre un exemple d'une classe avec une zone de haute densité en son centre. Nous voulons détecter cette zone en y concentrant la densité de possibilité.

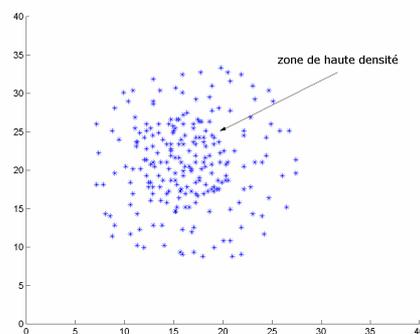


Figure 5.5. Exemple d'une classe avec une zone de haute densité en son centre.

En regardant les histogrammes de probabilité de la Figure 5.6, nous choisissons les seuils p_5^1 et p_6^2 pour les transformations précédentes. La Figure 5.7 montre les densités de possibilité correspondantes.

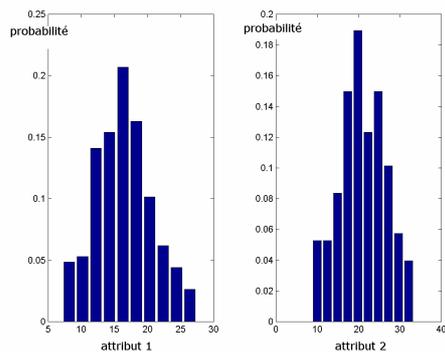


Figure 5.6. Histogrammes de probabilité pour l'exemple de la Figure 5.5.

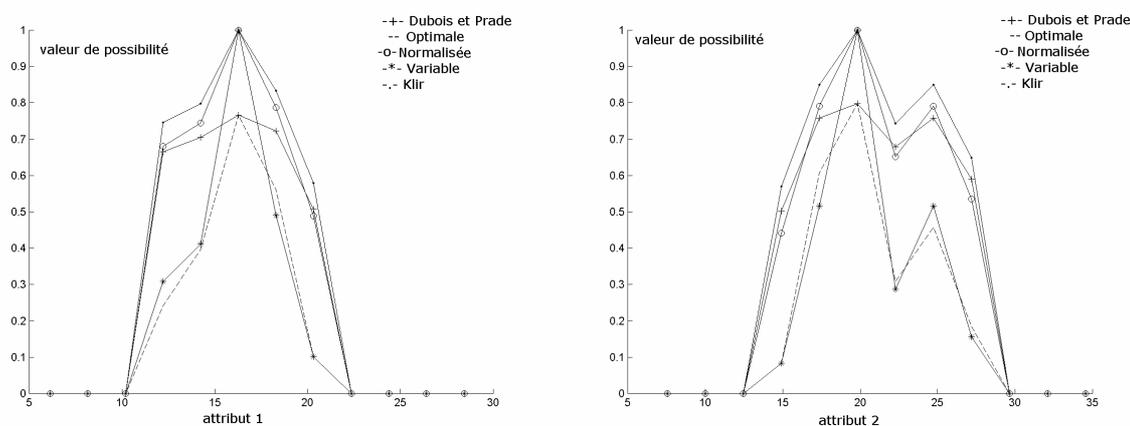


Figure 5.7. Densités de possibilité pour les différentes transformations avec les seuils p_5^1 et p_6^2 .

Les densités de possibilité obtenues par les transformations de Dubois et Prade classique et optimale sont toujours sous-normalisées contrairement aux autres. Nous remarquons également que la TV donne la densité de possibilité la plus concentrée sur la sous-classe. En regardant la Figure 5.8, nous remarquons que les courbes d'appartenance se concentrent sur la sous-classe.

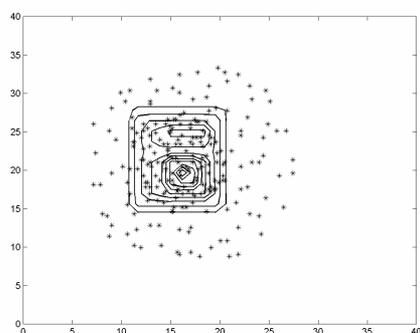


Figure 5.8. Courbes de niveau d'appartenance obtenues en utilisant la TV avec les seuils p_5^1 et p_6^2 : la courbe intérieure correspond au niveau d'appartenance le plus élevé.

Prenons l'exemple d'une zone de haute densité se trouvant déportée sur le côté de la classe (cf. Figure 5.9).

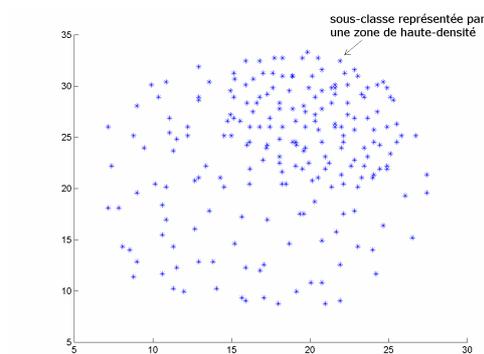


Figure 5.9. Exemple d'une classe avec une zone de haute densité déportée au côté droite de la classe.

Les histogrammes de probabilité de la Figure 5.10 nous conduisent à choisir les seuils p_5^1 et p_6^2 .

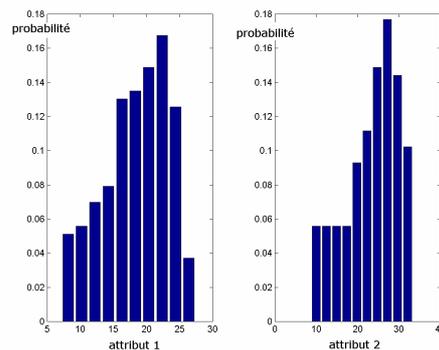


Figure 5.10. Histogrammes de probabilité pour l'exemple de la Figure 5.9

La Figure 5.11 montre que les densités de possibilité obtenues par la TV sont normalisées, contrairement à celles obtenues par les transformations symétrique et asymétrique de Dubois et Prade, et elles sont plus spécifiques que celles obtenues par les transformations de Klir et normalisée.

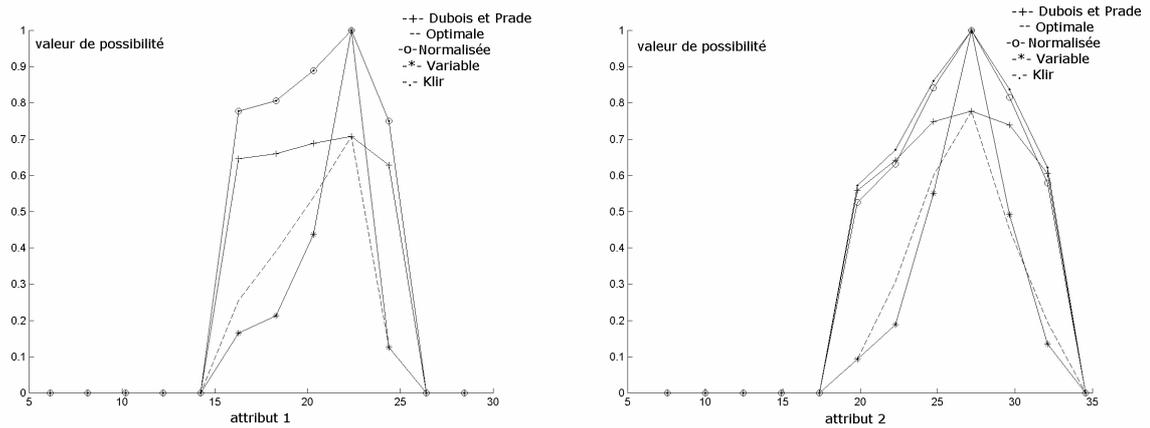


Figure 5.11. Densités de possibilité pour les différentes transformations avec les seuils p_5^1 et p_6^2 .

La Figure 5.12 nous montre que les courbes de niveau d'appartenance obtenues par la TV sont bien concentrées sur la zone de haute densité

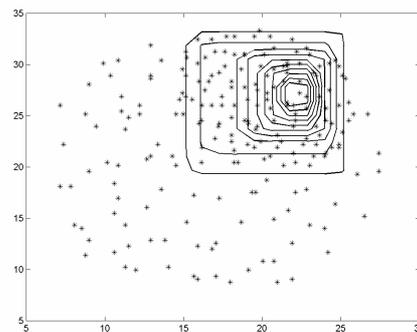


Figure 5.12. Courbes de niveau d'appartenance obtenues en utilisant la TV avec les seuils p_5^1 et p_6^2 : la courbe intérieure correspond au niveau d'appartenance le plus élevé.

5.4.2.2. Cas de plusieurs zones de haute densité

La Figure 5.13 montre une classe contenant 2 zones de haute-densité. Les histogrammes de probabilité de la Figure 5.14 nous conduisent à choisir les seuils p_4^1 et p_5^2 pour nos filtre.

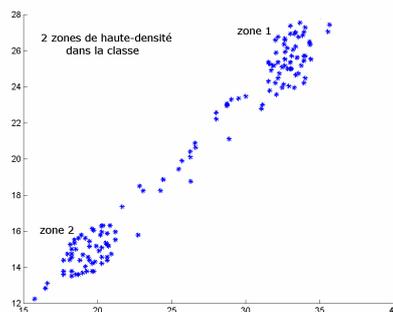


Figure 5.13. Exemple d'une classe avec 2 zones de haute densité.

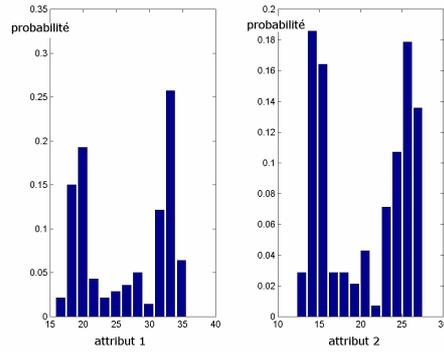


Figure 5.14. Histogrammes de probabilité pour l'exemple de la Figure 5.13.

La Figure 5.15 montre que les densités de possibilité obtenues par la TV sont les densités normalisées les plus spécifiques.

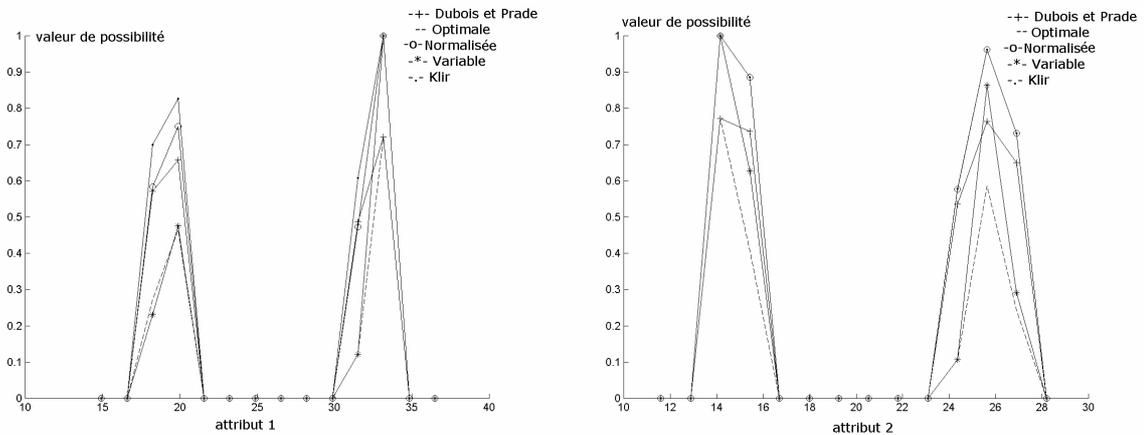


Figure 5.15. Densités de possibilité pour les différentes transformations avec les seuils p_4^1 et p_5^2 .

Les Figure 5.16.a et Figure 5.16.b montrent respectivement les courbes de niveau d'appartenance obtenues par la transformation de Dubois et Prade et la TV.

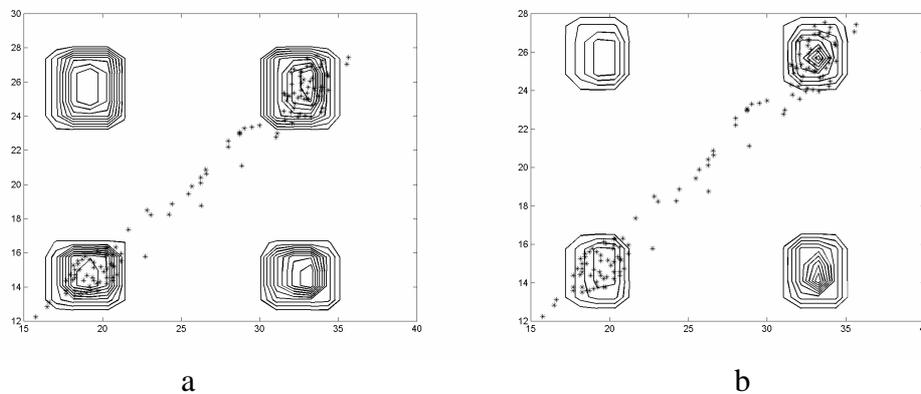


Figure 5.16. Courbes de niveau d'appartenance obtenues par a) la transformation de Dubois et Prade et b) par la TV avec les seuils p_4^1 et p_5^2 : la courbe intérieure correspond au niveau d'appartenance le plus élevé.

Nous constatons que :

- des courbes de niveau d'appartenance apparaissent dans des zones où il n'existe aucun point appartenant à la classe. Cela est dû au fait que la méthode FPM est inopérante dans le cas des classe de forme complexe ou oblique,
- la zone 1 contient des courbes de niveau d'appartenance plus fortes que celles de la zones 2 (cf. Figure 5.13). Cela signifie que la zone 1 est plus dense que la zone 2. Cette différence de densité entre les deux zones peut être vue plus clairement avec les courbes de niveau d'appartenance obtenues par la TV puisqu'elle est plus spécifique.

Nous pouvons résoudre le problème de la forme oblique ou complexe des classes, et donc éliminer les courbes de niveau d'appartenance apparaissant dans des zones vides de points, en utilisant la méthode FPMC. Les Figure 5.17.a et Figure 5.17.b montrent respectivement les courbes de niveau d'appartenance obtenues par la méthode FPMC avec les mêmes seuils que pour l'exemple de la Figure 5.13, et en utilisant la transformation de Dubois et Prade et la TV. Les courbes obtenues par la TV montrent mieux la différence de densité de points, dans les deux zones.

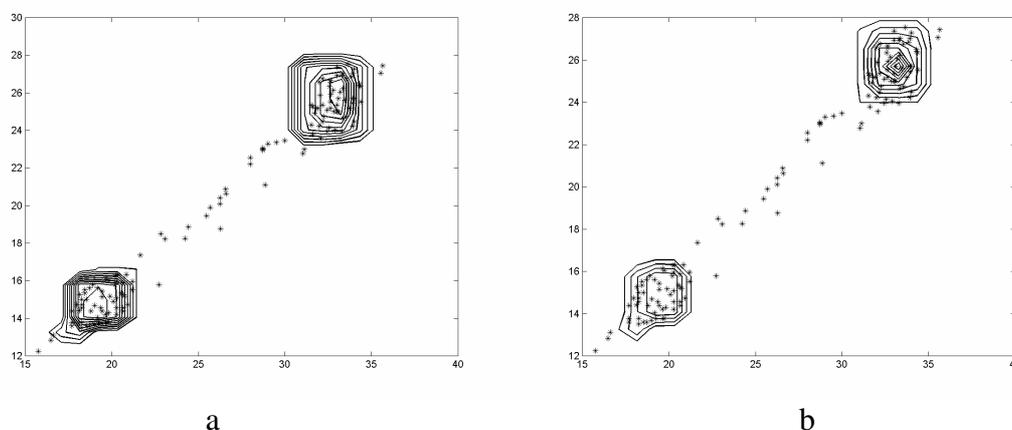


Figure 5.17. Courbes de niveau d'appartenance obtenues par la méthode FPMC avec a) la transformation de Dubois et Prade et b) avec la TV pour l'exemple de la Figure 5.13 et avec les seuils p_4^1 et p_5^2 .

5.5. Application industrielle

Cet exemple est une étude que nous avons réalisée pour la société Deville S.A. fabriquant des inserts de cheminée et des poêles à fuel. Pour atteindre ses objectifs en matière d'hygiène et d'environnement, cette société a remplacé le dégraissage en vapeurs de trichloroéthylène par un tunnel de dégraissage alcalin sans rejets aqueux et a remplacé la peinture à base de

solvants chlorés par une peinture hydrosoluble haute température. L'amélioration de la productivité nécessite de déposer une épaisseur minimale mais suffisante de peinture sur les pièces. Cela ne sera atteint qu'au travers la maîtrise complète du process de peinture et de tous ces paramètres.

L'objectif de notre travail consiste à montrer à la société Deville S.A. l'intérêt d'un module complet de supervision et de diagnostic de la ligne de peinture au trempé. Ce module doit identifier le mode de fonctionnement qui permet à la ligne de déposer une épaisseur minimale mais suffisante de peinture sur les pièces. Il s'agit donc d'un module prédictif, le but étant d'estimer la qualité des pièces peintes sans attendre leur sortie de la ligne.

Ce module permet de supprimer le contrôle de qualité systématique mis en place sur la plupart des procédés industriels de fabrication et par conséquent de diminuer le coût et de réduire le nombre de rebuts et de surplus. Pour que ce module remplisse son objectif, il doit être capable de :

- détecter, en temps réel, si le mode de fonctionnement dans lequel se trouve la ligne à l'instant t est un mode normal qui fournit des foyers bien peints, ou un mode anormal qui fournit des pièces de mauvaise qualité. Dans le deuxième cas, il doit le faire savoir à l'opérateur pour que ce dernier puisse prendre des mesures correctives c'est-à-dire régler les paramètres pour que la ligne revienne dans un mode normal,
- détecter, en temps réel, si le système dérive du mode normal vers un mode anormal. Il peut donc déterminer si la qualité du produit commence à se dégrader.

5.5.1. Fonctionnement de la ligne de peinture au trempé

La ligne de peinture au trempé est destinée à déposer un revêtement organique sur l'ensemble des pièces fabriquées dans les ateliers de tôlerie et de soudure, ainsi que sur les pièces en fonte achetées à l'extérieur. Elle se compose :

- d'une zone d'accrochage-décrochage permettant d'accrocher les différentes pièces brutes d'une carcasse ainsi que le décrochage des pièces peintes,
- d'une zone d'accrochage de carcasses brutes soudées, située à côté des machines à souder,
- d'un tunnel de dégraissage-rinçage qui se compose d'une zone d'aspersion de dégraissant et d'une zone de rinçage en 3 cascades : rinçage 1, rinçage 2 et rinçage 3.

Le rinçage 3 est alimenté à partir d'une station de fabrication d'eau osmosée et d'une réserve permettant d'assurer les régénérations de l'osmoseur. Le tunnel est équipé d'un extracteur de vapeur, l'eau extraite des vapeurs est réinjectée dans le rinçage 3 puis remonte l'installation par les différentes cuves. Le tunnel est donc à zéro rejet aqueux,

- d'une étuve pour sécher les pièces après rinçage. La présence d'une trace d'eau résiduelle, après étuvage, n'est pas critique,
- d'un refroidissement à l'air libre,
- d'une zone d'accrochage fonte,
- d'une cuve de trempe où l'on trouve la peinture hydrosoluble,
- d'une zone d'égouttage pour l'élimination de l'eau et du cosolvant,
- d'un four de polymérisation,
- d'une zone de refroidissement.

5.5.2. Paramètres de contrôle de fonctionnement de la ligne

Les paramètres de contrôle mettent la ligne de peinture en bonnes conditions de fonctionnement afin qu'elle puisse donner des pièces correctes. La relation entre la qualité des produits et ces paramètres est donc indirecte.

Il existe 26 mesures de contrôle distribuées sur la ligne de peinture, relevées par un enregistreur. Quelques capteurs sont munis d'une alarme avertissant le conducteur de ligne en cas de mesure anormale. Le Tableau 11 montre les 26 mesures ainsi que leur type et leurs seuils.

N° voie	Descriptif	Type	Seuil d'alarme
PEINTURE			
1	Température peinture	numérique	27°C 23°C
2	Pression sortie pompe	numérique	3bars 1bar
3	Pression sortie pompe	numérique	2bars 0.8bar
4	Niveau peinture	numérique	10Mm 200Mm
5	Vitesse pompe	numérique	50Hz 20Hz
6	Vitesse chaîne 1	numérique	1.7Mm 1.2Mm
7	Vitesse chaîne 2	numérique	1.7Mm 1.2Mm
FOUR CUISSON			
8	Température	numérique	250°C 215°C
9	Contrôleur 2 ventilateurs	Logique	
DÉGRAISSAGE			
10	Température dégraissage	numérique	70°C 50°C
11	Température étuve	numérique	135°C 90°C
12	Niveau 1 remplissage	Logique	
13	Niveau 2 remplissage	Logique	
14	Niveau 3 remplissage	Logique	
15	Pompe 1 marche sans défaut	Logique	
16	Pompe 2 marche sans défaut	Logique	
17	Pompe 3 marche sans défaut	Logique	
18	Electro vanne alimentation	Logique	
RINCAGE			
19	Pompe eau osmosée	Logique	
20	Circulation extraction dég	Logique	
21	Circulation extraction dég	Logique	
22	Eau déminéralisée niveau B	Logique	
23	Niveau 4 pompe	Logique	
24	Niveau 5 arrêt pompe	Logique	
25	Niveau 6 arrêt pompe	Logique	
26	Alarme osmosé	Logique	

Tableau 11. Paramètres de contrôle de la ligne au trempé.

5.5.3. Qualité du produit

L'épaisseur de peinture n'a pas pour but d'assurer un aspect esthétique aux pièces qui doivent être toutefois exemptes de coulures, boursouflures, dans leurs parties visibles. Ce revêtement doit protéger les pièces contre la corrosion dans des conditions normales de stockage. De plus, il doit résister aux températures les plus hautes constatées dans le fonctionnement des appareils. Si l'épaisseur du film est trop faible, la fonction protection n'est pas assurée. Si elle est trop importante, des défauts d'aspect apparaissent mais surtout le prix de revêtement de la pièce augmente. L'épaisseur minimale désirée est de 10 µm. On mesure l'épaisseur du produit en plusieurs points comme le montre la Figure 5.18.

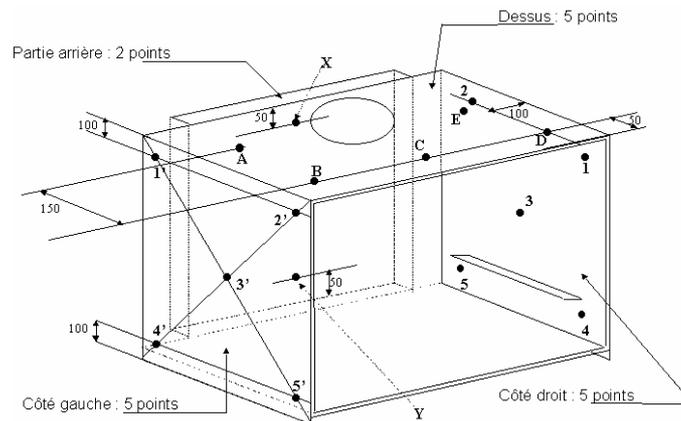


Figure 5.18. Points de mesure de l'épaisseur de peinture sur un foyer.

Le contrôle de qualité des pièces peintes est actuellement effectué par observation visuelle. Les critères d'acceptation des pièces peintes sont les suivants : homogénéité de la couleur, pas de coulure, pas de goutte, pas de rayure, pas d'écaillage, pas de manque dû à un point d'accrochage sur les faces et homogénéité de la peinture.

Des tests de dureté de la peinture, d'adhérence, d'épaisseur et de teinte sont réalisés périodiquement sur des plaquettes test passant par les conditions du fonctionnement de la ligne de peinture.

Le principe de la mesure de dureté, dit dureté crayon, consiste à déplacer des crayons à mines de duretés différentes, de 6 B à 7 H, sur la surface du revêtement à tester, sous charge constante de 7.5 N et suivant un angle de 45°. Pour chaque mine, la pénétration de la pointe détermine la dureté du film.

Le test d'adhérence consiste à diviser la surface peinte en 25 carrés de 1 mm² de surface chacun. L'essai du quadrillage sert à déterminer l'adhérence de la peinture sur son support. L'appareil pour inciser, dit ruban, est constitué d'un manche avec cylindre muni de 6 lames espacées de 1 mm. Le ruban incise la surface avec un angle de 90° d'un mouvement rapide et continu. Un adhésif est collé sur la surface puis décollé. On classe l'adhérence de la peinture selon l'importance de la zone décollée par rapport à la zone quadrillée. Chaque carré plein apparaissant sur l'adhésif représentera une perte d'adhérence de 10 %, aucun carré signifie que l'adhérence est de 100 %.

5.5.4. Paramètres de qualité des produits peints

Les expérimentations effectuées sur la qualité des pièces peintes, sur la ligne de peinture au trempé de Deville, ont permis de déterminer les paramètres qui contrôlent la qualité des produits. Ces paramètres sont : la viscosité du bain de peinture, son pH, et sa température.

5.5.4.1. Viscosité

La viscosité d'une peinture détermine l'épaisseur du film déposé sur la pièce. Une viscosité trop basse entraîne une épaisseur faible, une viscosité trop importante entraîne une épaisseur élevée. La viscosité idéale doit être comprise entre 19.8 et 21 secondes à 25° C

5.5.4.2. pH

Le pH a une incidence sur l'écoulement de la peinture, donc sur l'aspect du film. D'autre part, le pH assure la stabilité du bain de peinture. C'est pourquoi, il est très important de ne pas dépasser une valeur limite définie pour le type de produit utilisé. La valeur idéale de pH pour la ligne de peinture au trempé est de 9.1 ± 0.1 .

5.5.4.3. Température de la peinture :

La température joue un rôle important sur le film de peinture déposé et donc sur la qualité du produit. La température idéale est fixée à $25 \pm 1^\circ\text{C}$.

5.5.5. Construction du module de diagnostic de qualité des produits peints

5.5.5.1. Détermination et réduction de l'espace de représentation

Il s'agit de définir tous les paramètres nécessaires pour modéliser la qualité des produits peints. Les paramètres disponibles sont la viscosité, la température et le pH de la peinture. La qualité optimale est obtenue en fixant les paramètres précédents dans l'intervalle des valeurs idéales afin de donner l'épaisseur minimale nécessaire de peinture.

L'Analyse en Composantes Principales (ACP) montre que les paramètres viscosité et pH sont orthogonaux et que le pH est inversement corrélé avec la température. L'espace formé des deux paramètres viscosité et pH conserve 73.2 % de l'information totale du nuage de points. Nous conservons donc ces deux paramètres. La Figure 5.19 montre le cercle de corrélation des trois paramètres précédents.

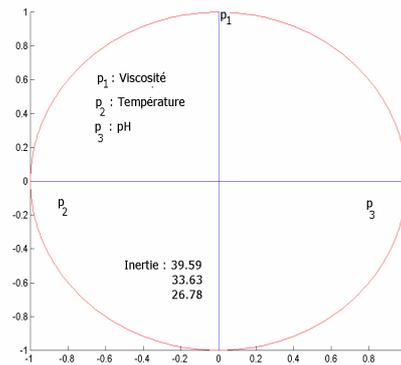


Figure 5.19. Cercle de corrélation des paramètres de qualités : viscosité, température et pH de la peinture.

5.5.5.2. Détermination de l'espace de décision

Il s'agit de déterminer l'ensemble de tous les modes de fonctionnement possibles. La Figure 5.20 montre la projection du nuage de points sur l'espace formé de la viscosité et du pH. Le nuage est divisé en trois zones : la zone 1 de haute densité de points, la zone 2 de densité plus faible et la zone 3 de points isolés.

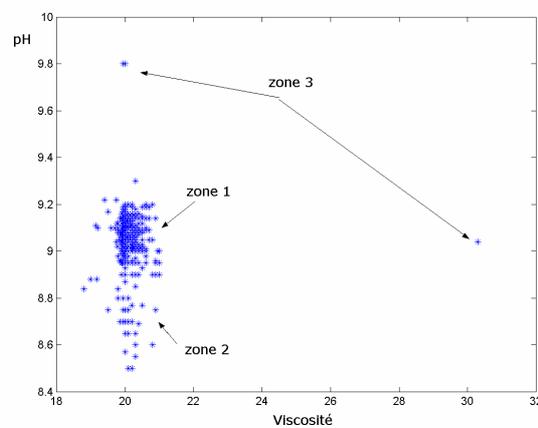


Figure 5.20. Nuage de points divisé en trois zones dans l'espace des deux paramètres : viscosité et pH.

La zone 1 est la zone où les points respectent les valeurs souhaitées de viscosité et de pH. Cette zone correspond aux produits à la fois de bonne qualité et de faible coût. Dans la zone 2, les points ont des valeurs de pH inférieures aux valeurs souhaitées. Cela conduit à déposer une épaisseur plus élevée que celle nécessaire pour la protection du produit, ce qui entraîne l'augmentation du coût. La zone 3 est la zone des points isolés correspondant à une épaisseur trop élevée ou trop fine, ce qui entraîne une dégradation importante de la qualité ou du prix. Cette zone est la zone des produits rebus.

5.5.5.3. Application de FPM avec la TV

Le système de diagnostic mis au point pour garantir une épaisseur de peinture minimale mais suffisante, utilise la méthode FPM associée à la Transformation Variable. Nous avons choisi les seuils p_1^1 et p_2^2 à partir des histogrammes de probabilité de la Figure 5.21. Cela permet d'obtenir des densités de possibilité concentrées sur la zone 1 correspondant à la zone souhaitée.

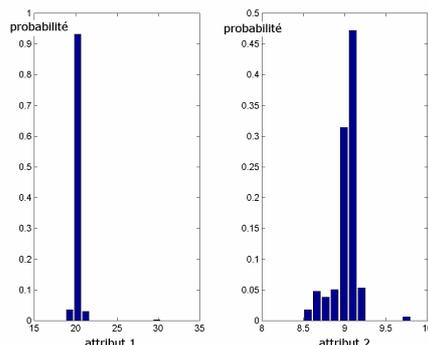


Figure 5.21. Histogrammes de probabilité pour les données de la ligne de peinture.

La Figure 5.22 montre les densités de possibilité pour les deux attributs : viscosité et pH en utilisant les transformations usuelles avec les seuils p_1^1 et p_2^2 .

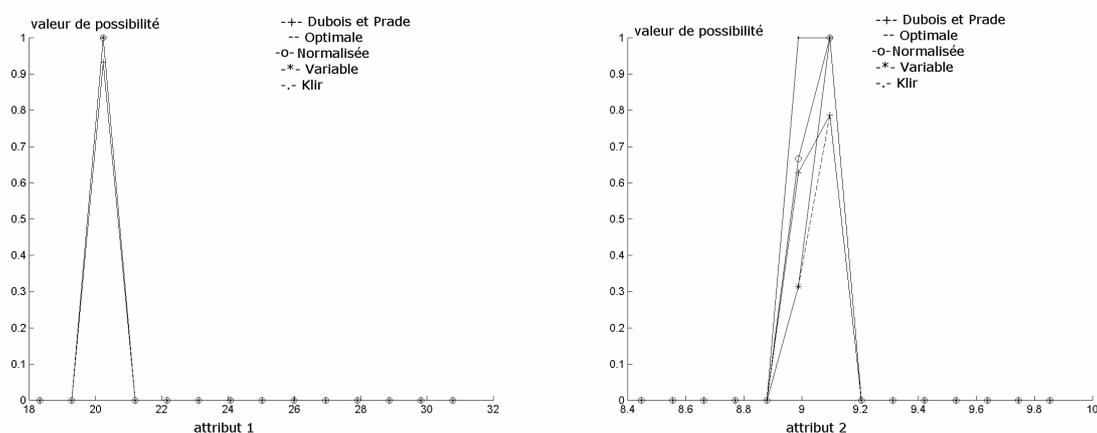


Figure 5.22. Densités de possibilité pour les différentes transformations avec les seuils p_1^1 et p_2^2 .

La Figure 5.23 montre les courbes de niveau d'appartenance obtenues en utilisant la TV avec ces seuils. Nous pouvons remarquer que seuls les points appartenant à la zone 1 sont classifiés, les autres points sont rejetés. Un point rejeté indique un état anormal de fonctionnement et nécessite un réglage de la ligne. De plus, nous pouvons prédire l'évolution de la ligne de peinture vers les zones 2 et 3 par la diminution des valeurs d'appartenance à la zone 1.

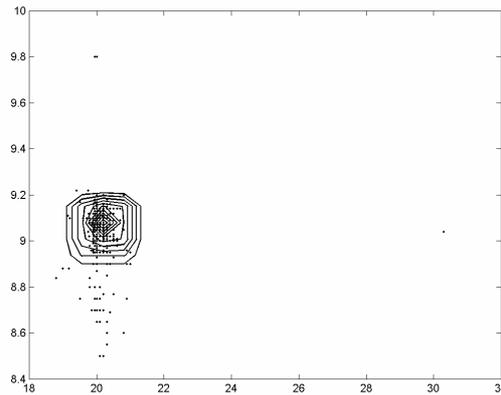


Figure 5.23. Courbes de niveau d'appartenance obtenues en utilisant la TV avec les seuils p_1^1 et p_2^2 .

L'utilisation de ce module de diagnostic devrait permettre d'éliminer la phase de contrôle qualité des produits et d'anticiper toute dérive éventuelle des paramètres.

5.6. Seuil adaptatif de la TV

L'application du seuil adaptatif défini en (5.6), sur les exemples précédents, donne des résultats très proches de ceux obtenus avec des seuils fixés par l'utilisateur. La TV avec le seuil adaptatif élimine les points aberrants et se concentre sur les noyaux ou les zones de haute densité des classes. Si la classe ne contient pas de zones de haute densité ou si elle n'est pas entourée de points aberrants, la TV adapte son seuil pour donner une bande passante assez large afin d'accepter tous les points dans la base de données. La Figure 5.24 montre un exemple de classe qui ne comprend pas de sous-classes et qui n'est pas entourée de points aberrants.

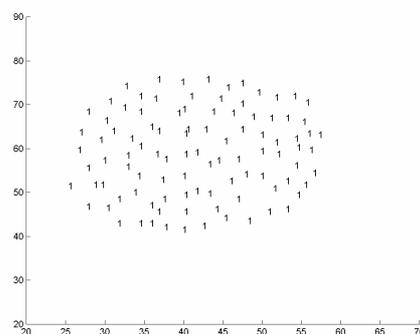


Figure 5.24. Classe qui ne contient ni zones de haute densité ni points aberrants.

Les histogrammes de probabilité de la Figure 5.25 montrent que les barres ont des hauteurs comparables.

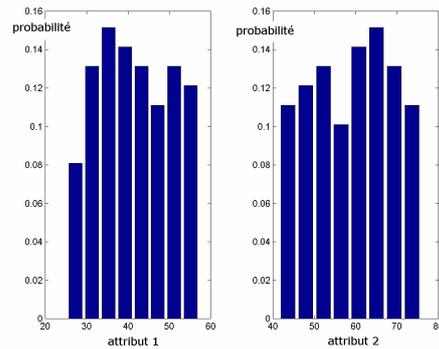


Figure 5.25. Histogrammes de probabilité pour l'exemple de la Figure 5.24.

La Figure 5.26 montre les densités de possibilité pour cet exemple et la Figure 5.27 montre les courbes de niveau d'appartenance. Nous remarquons à travers ces deux figures que la TV adapte son seuil pour avoir une bande passante acceptant tous les points de la classe dans la base d'apprentissage.

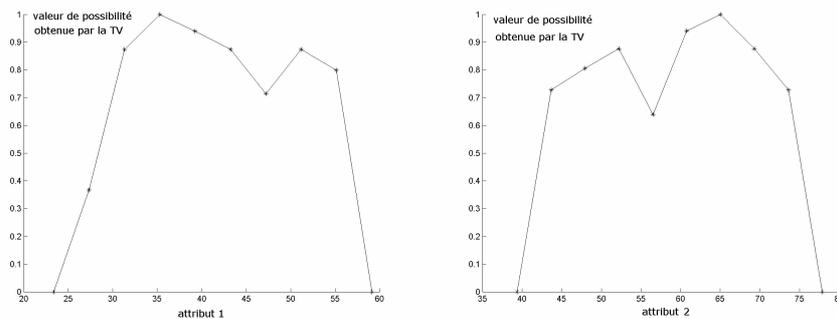


Figure 5.26. Densités de possibilité obtenue par la TV avec le seuil adaptatif pour l'exemple de la Figure 5.24.

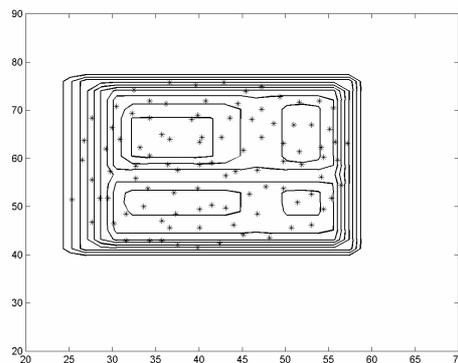


Figure 5.27. Courbes de niveau d'appartenance obtenues par la TV avec le seuil adaptatif pour l'exemple de la Figure 5.24.

5.7. Choix de h

Les performances pour la détection des zones de haute-densité ou pour l'élimination de points aberrants dépendent de la valeur de h . Dans le chapitre précédent, nous avons calculé la valeur optimale de h afin de maximiser la séparabilité entre les classes. En revanche, nous cherchons ici la valeur de h qui est capable de révéler l'existence des points aberrants ou des zones de haute densité. Pour cela nous utilisons l'écart type des probabilités de l'histogramme, que nous appelons l'indice de révélation :

$$I_d = \sqrt{\frac{1}{n} \cdot \sum_{m=1}^n |(p_m - \bar{p})|} \quad (5.7)$$

où \bar{p} est la valeur moyenne des probabilités de l'histogramme.

Pour une classe entourée de points aberrants ou comportant des zones de haute densité, la dispersion est importante et donc l'écart type est élevé. Dans ces conditions, la valeur optimale de h est celle qui conduit à l'écart type le plus élevé. Si I_d est toujours faible, quel que soit h , cela signifie qu'il n'y a pas de points aberrants ni de zones de haute-densité.

Les Figure 5.28.a, Figure 5.28.b, Figure 5.28.c et Figure 5.28.d montrent la relation entre I_d et h pour les exemples précédents.

Nous constatons que quand la classe contient des zones de haute densité ou des points aberrants, I_d a une valeur relativement élevée, c'est le cas des Figure 5.28.a, Figure 5.28.b, Figure 5.28.c. En revanche, sur la Figure 5.28.d, nous constatons que I_d a toujours une valeur faible ce qui signifie qu'il n'y a pas de points aberrants ou de zones de haute densité.

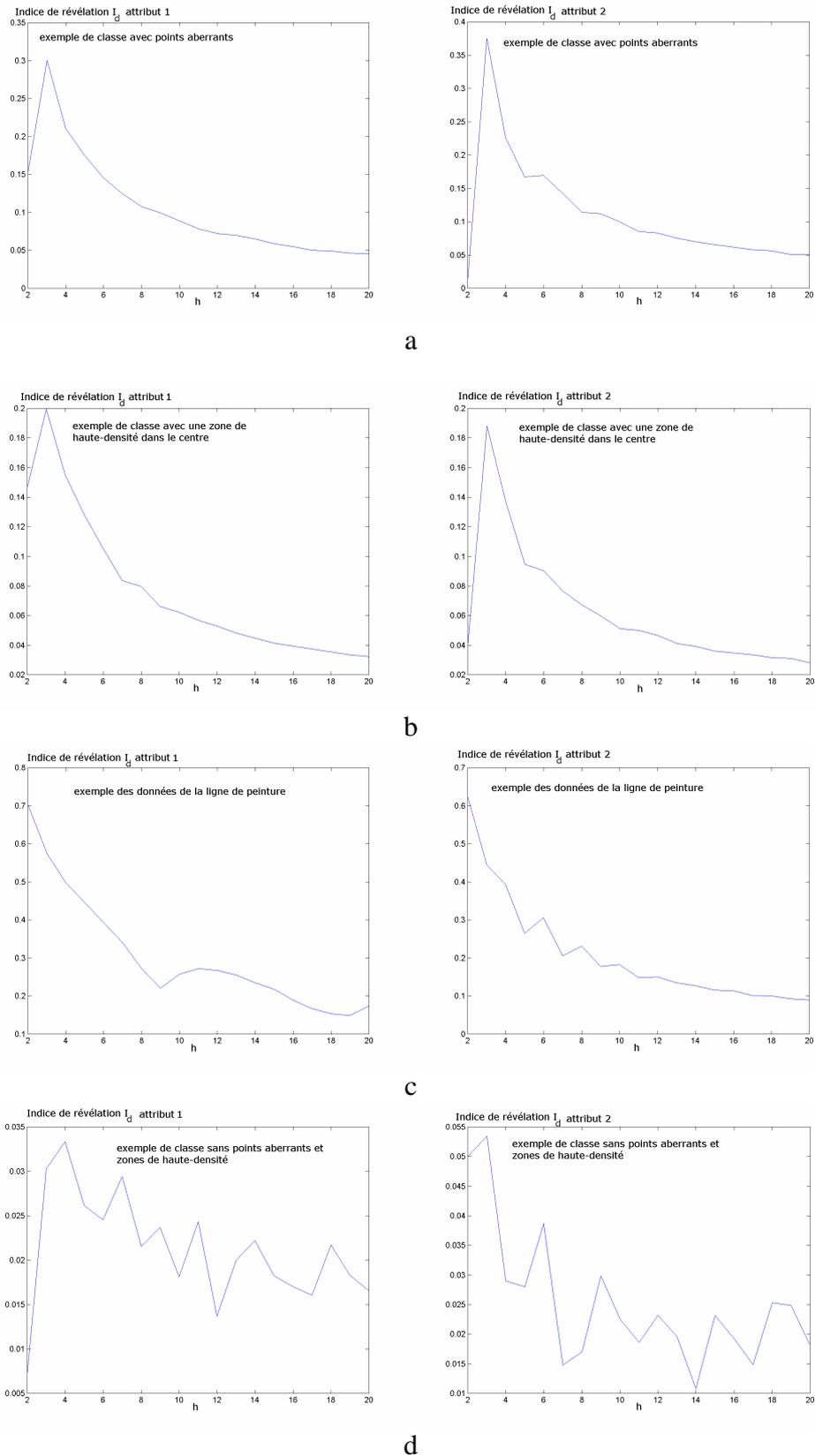


Figure 5.28. Relation entre h et l'indice de révélation pour les exemples précédents.

Si la valeur de h qui maximise l'indice I_d n'est pas assez grande, les courbes de niveau d'appartenance ne seront pas assez concentrées sur les zones souhaitées. Cela est dû à la

largeur de barre de l'histogramme. La valeur de h est un compromis entre la valeur donnée par la maximisation de I_d et la spécificité voulue. Cette valeur de compromis doit faciliter à l'utilisateur le choix des bornes de son filtre ou doit permettre au filtre adaptatif de bien détecter les zones de haute-densité ou les points aberrants. En prenant l'exemple des données de la ligne de peinture, $h = 2$ maximise l'indice de révélation. La Figure 5.29 montre les histogrammes de probabilité obtenus pour $h = 2$, en conséquence, nous choisissons les seuils p_1^1 et p_1^2 .

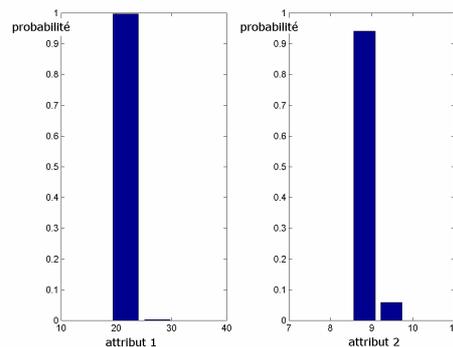


Figure 5.29. Histogrammes de probabilité obtenus pour les données de la ligne de peinture avec $h = 2$.

La Figure 5.30 montre les courbes de niveaux d'appartenance obtenues en utilisant ces seuils, nous constatons que ces courbes ne sont pas assez spécifiques pour se concentrer sur la zone de haute-densité.

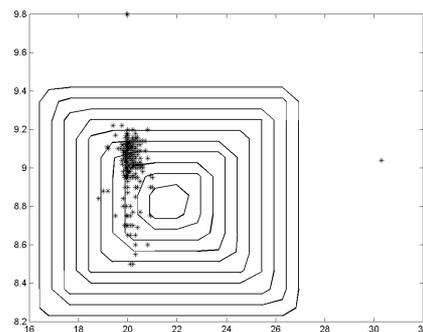


Figure 5.30. Courbes de niveaux d'appartenance obtenues pour les données de la ligne de peinture avec $h = 2$.

5.8. Conclusion

Dans ce chapitre, nous avons comparé l'aptitude de la Transformation Variable (TV) et des transformations les plus connues en terme d'élimination de bruit et de détection de zones de haute densité, en utilisant soit un seuil de possibilité, soit un seuil de probabilité. La TV facilite le choix d'un seuil de possibilité puisqu'elle est la plus spécifique et qu'elle valorise les probabilités les plus élevées. Dans le cas de l'utilisation d'un seuil de probabilité, la TV est la seule qui est à la fois normalisée et la plus spécifique.

La TV est également capable de déterminer automatiquement des seuils conformes aux différents jeux de données. Autrement dit, elle est capable d'adapter sa bande passante en fonction du système à étudier, de façon à éliminer uniquement les points aberrants ou à se concentrer sur les zones de haute densité, si elles existent. Enfin la TV est plus rapide que les transformations classique et optimale de Dubois et Prade. Elle est donc plus adaptée pour les applications en temps réel.

Les performances de la TV dépendent du choix du paramètre h . Il faut choisir une valeur de h suffisante d'une part pour révéler l'existence des points aberrants ou des zones de haute-densité, d'autre part pour donner des courbes de niveau d'appartenance assez concentrées sur les zones de points souhaitées. Cela peut être réalisé par un compromis entre la maximisation de l'indice de révélation et la spécificité de détection souhaitée.

6. Diagnostic adaptatif et prédictif en utilisant FPM

6.1. Introduction

L'application de la reconnaissance floue des formes en diagnostic se heurte souvent au problème de la connaissance a priori insuffisante du système étudié. Dans une base de connaissance incomplète, tous les modes de fonctionnement ne sont pas représentés. En effet, pour des raisons de coût ou de sécurité, certains états défectueux ou dangereux ne peuvent pas être provoqués et seront donc absents de la base. Un processus de diagnostic par RdF doit donc être adaptatif, c'est-à-dire capable de détecter les états inconnus et de les apprendre, afin d'enrichir sa connaissance a priori. De plus, il est moins onéreux d'anticiper l'évolution d'un mode de fonctionnement normal vers un mode anormal que d'attendre d'y arriver. Cela a pour but d'éviter les modes anormaux et leurs conséquences.

Il est nécessaire de concevoir des outils permettant d'acquérir cette connaissance manquante au cours de la supervision. Cette connaissance provient des observations qui ne sont affectées à aucune classe, c'est-à-dire qui sont rejetées. L'apprentissage incrémental est incapable d'obtenir cette connaissance manquante puisque les points sont rejetés. Dans ce chapitre, nous allons montrer que FPM est incapable d'anticiper l'évolution de l'état du système et de détecter l'apparition des nouveaux modes de fonctionnement. Nous allons introduire une nouvelle approche pour intégrer FPM dans un module de diagnostic prédictif et adaptatif [MOU01, MOU02c]. Enfin, nous allons évaluer les performances de notre approche sur différents exemples.

6.2. Diagnostic prédictif

Un système évolue d'un état normal vers un état anormal soit brutalement soit successivement ou lentement. Une évolution est brutale si le point à l'instant $t + 1$ n'est pas affecté à la même classe que le point à l'instant t . Dans ce cas, la prédiction de l'évolution est impossible. L'évolution successive se caractérise par l'apparition de plusieurs points intermédiaires entre un point affecté à la classe de départ et un point affecté à la classe d'arrivée. Ces points intermédiaires sont généralement rejetés. Dans ce cas d'évolution successive, il est préférable de prédire l'évolution plutôt que d'attendre d'arriver à un mode anormal afin de prendre immédiatement des démarches correctives et d'éviter les conséquences du dysfonctionnement. Une évolution désigne un état de transition entre deux

modes de fonctionnement. Prévoir la classe, ou le mode de fonctionnement, d'arrivée revient à effectuer un pronostic, ou diagnostic prédictif [PEL93].

La simple connaissance de l'appartenance à une classe ne suffit pas pour la prise en compte du phénomène d'évolution d'un mode de fonctionnement vers un autre. Il faut pouvoir quantifier la proximité ou l'éloignement d'une observation à une classe. Cette quantification traduit la représentativité de l'observation vis à vis des classes.

L'évolution s'appuie sur une suite chronologique d'observations décrivant le comportement d'un système sur un intervalle de temps. Cette évolution est caractérisée par un chemin entre les deux états comme le montre la Figure 6.1. Le nombre de points n_e nécessaire pour prédire l'évolution du système dépend de son type : lent ou rapide.

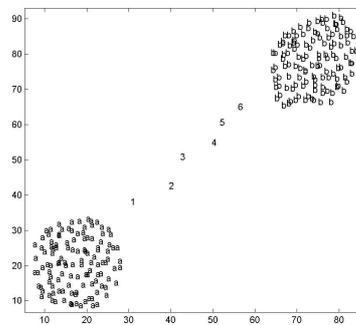


Figure 6.1. Evolution de la classe a vers la classe b.

FPM classe un nouveau point dans une des classes connues ou le rejette. La Figure 6.2 présente les courbes de niveaux d'appartenance pour l'exemple de la Figure 6.1. FPM divise donc l'espace de représentation en deux zones, la première est celle où les points seront affectés à la classe a ou à la classe b et la zone 2 est la zone où les points seront rejetés.

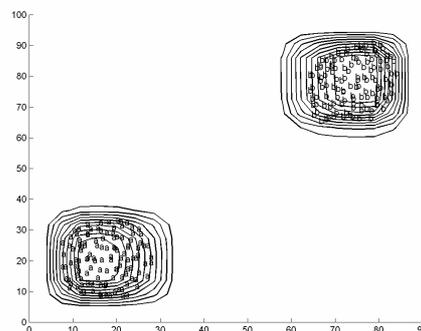


Figure 6.2. Division de l'espace de représentation en deux zones par FPM : la zone où les points sont affectés à la classe a ou b et la zone où les points sont rejetés.

La zone contenant le chemin d'évolution s'appelle la zone d'évolution. Puisque, lorsque

l'on utilise FPM, celle-ci correspond à la zone des points rejetés, FPM ne fournit aucune information de la représentativité des points vis à vis des classes. FPM n'est donc pas capable de prédire l'évolution entre les classes.

6.3. Diagnostic adaptatif

Un problème de diagnostic par RdF s'accompagne d'une base de connaissance incomplète où tous les modes de fonctionnement ne sont pas représentés. Un processus de diagnostic adaptatif doit être capable de détecter et d'inclure à sa base de connaissances les états inconnus.

Le rejet confère au système de décision son caractère adaptatif. En effet, détecter un état inconnu nécessite un nombre minimum de points rejetés et géographiquement proches les uns des autres. Une nouvelle classe ne peut donc apparaître que dans la zone d'évolution. FPM ne donne pas d'information sur la situation géographique d'un point dans cette zone. La Figure 6.3 montre un exemple d'apparition d'une nouvelle classe représentée par le rejet en appartenance d'un ensemble de points.

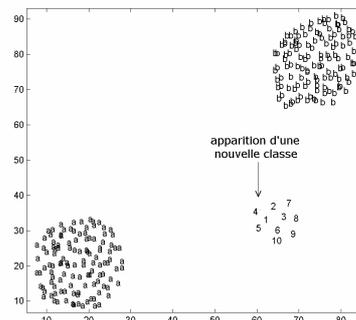


Figure 6.3. Apparition d'une nouvelle classe représentée par des points rejetés en appartenance.

Pour inclure un état inconnu dans la base de connaissance, FPM doit arrêter la phase de classification. Le système de diagnostic doit utiliser une méthode de coalescence floue pour introduire la nouvelle classe dans l'ensemble des classes. FPM apprend ensuite cette nouvelle classe. La Figure 6.4 montre une adaptation hors-ligne d'un système de diagnostic. Cette procédure lourde empêche l'utilisation du diagnostic adaptatif, basé sur FPM, en temps réel. De plus, ce traitement hors-ligne des points rejetés est critique lorsque le nouveau mode de fonctionnement est dangereux et doit être détecté aussi tôt que possible.

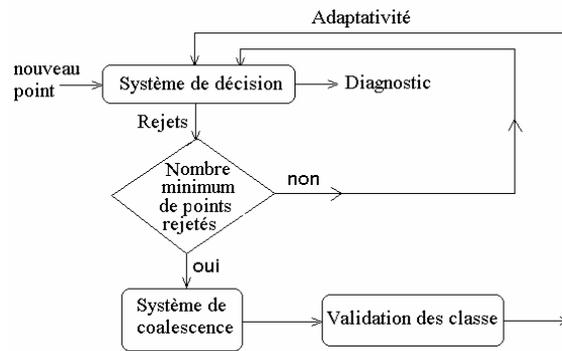


Figure 6.4. Adaptativité du système de diagnostic hors-ligne.

6.4. Intégration de FPM dans un module de diagnostic prédictif et adaptatif

Il existe deux types d'algorithmes pour prédire l'état futur du système [DUB90] :

- les algorithmes utilisant une approche floue qui introduisent la notion d'appartenance graduelle. Ils définissent des classes où les frontières ne sont pas brutales mais floues. Une partition floue d'un ensemble d'observations est préférable à une partition classique lorsqu'il existe des ponts entre les classes. En termes de modes de fonctionnement, ces ponts représentent un chemin d'évolution entre deux états d'où l'intérêt de cet algorithme,
- les algorithmes utilisant la géométrie de l'ensemble d'apprentissage pour lesquels l'espace peut être découpé en régions d'équi-appartenance aux classes.

Ces deux types d'algorithmes ont besoin d'informations dans la zone d'évolution, le premier nécessite la connaissance de la forme analytique de la fonction d'appartenance et le deuxième nécessite une base de connaissance parfaite. La fonction d'appartenance de FPM est une série de densités de possibilité calculées statistiquement et leurs formes analytiques sont inconnues. De plus, dans la plupart des cas réels, on dispose rarement d'informations complètes. En conséquence ces deux types d'algorithmes ne peuvent pas être utilisés pour réaliser le diagnostic prédictif et adaptatif en utilisant FPM.

Dans la littérature, de nombreux auteurs se sont intéressés au problème du diagnostic adaptatif et évolutif et ont apportés des solutions. Dans le cas du diagnostic prédictif, une première solution suppose la connaissance de la loi de densité de probabilité [FRE92, GAN87, GRE84, PEL93, ZIE95] et utilise donc des méthodes statistiques paramétriques de détection d'évolution comme la méthode CUSUM ou le filtre de Kalman. Une seconde

solution consiste à chercher à apprendre deux fonctions d'appartenance pour chaque couple de classes [BOU96, CRO02]. Ces deux fonctions d'appartenance épousent la forme de la trajectoire supposée au sein d'un couple. Cette solution nécessite la connaissance a priori du chemin d'évolution entre chaque couple de classes et la détermination d'un grand nombre de paramètres, pour construire la bonne forme de la trajectoire entre les classes. De plus, elle nécessite un temps de calcul relativement élevé.

Dans le cas du diagnostic adaptatif, les solutions proposées, moins nombreuses que dans le cas précédent, adaptent les systèmes soit hors-ligne soit en ligne. Le cas d'adaptation hors ligne est peu efficace lorsque le système dérive lentement vers un mode inconnu. De plus la détection tardive peut être catastrophique si ce mode est dangereux. L'adaptation en ligne peut être faite en utilisant : les critères de voisinage ou d'activation [BOU97], le filtre de Kalman [FRE92], ou la méthode potentielle [GRE84, PEL93]. L'utilisation des critères de voisinage et d'activation nécessite la détermination de plusieurs paramètres et donne des performances et une robustesse assez médiocres [BOU97]. L'utilisation du filtre de Kalman nécessite la vérification de l'hypothèse de la connaissance de la loi de densité de probabilité ce qui n'est pas souvent possible. Enfin l'utilisation de la méthode potentielle qui découpe l'espace de représentation en régions d'équi-appartenance aux classes, demande une connaissance complète de l'espace de décision et des temps de calcul élevés.

Dans notre cas, nous ne disposons d'aucune connaissance sur la loi de densité de probabilité. De plus nous cherchons, d'une part, à réaliser la prédiction et l'adaptation en ligne, et d'autre part, à avoir une méthode robuste et performante tout en respectant les particularités de la méthode FPM.

6.4.1. Prédiction de l'état futur

6.4.1.1. Fonction d'évolution

Supposons une série chronologique composée des dernières observations rejetées du système à surveiller, cette série porte des informations sur le comportement du système en terme d'éloignement ou de rapprochement des classes. Quand un système évolue progressivement de la classe 1 vers la classe 2, la possibilité d'appartenance, de la suite d'observations, à la classe 1 devrait diminuer au fur et à mesure de l'éloignement. En revanche, la possibilité d'appartenance à la classe 2 devrait augmenter. Nous allons montrer comment notre nouvelle approche suit cette évolution et prédit l'état futur du système.

Dans notre approche [MOU02c, MOU01], chaque axe ou attribut utilise deux informations

pour quantifier la représentativité d'une observation rejetée par lui dans la zone d'évolution : la valeur d'appartenance de son plus proche voisin et la distance entre eux. Nous allons construire une fonction quantifiant la représentativité de l'observation rejetée pour chaque axe en utilisant ces deux informations. Cette fonction, appelée fonction d'évolution, caractérise l'influence de la classe j sur le point \underline{x} par rapport à l'attribut i . Elle ne donne plus la valeur d'appartenance d'un point à une classe puisque ce point à été rejeté, mais elle donne sa représentativité. Cette fonction a la forme suivante :

$$v_j^i(\underline{x}) = (1 - d_n^i(\underline{x}, \underline{x}_{ppv})) \frac{\pi_j^i(\underline{x}_{ppv})}{c}, \quad i = 1 \dots a, \quad j = 1 \dots c \quad (6.1)$$

La représentativité de l'observation rejetée \underline{x} par rapport à la classe j et l'attribut i , $v_j^i(\underline{x})$, est une fonction décroissante de la distance normalisée $d_n^i(\underline{x}, \underline{x}_{ppv})$ entre l'observation \underline{x} et son plus proche voisin \underline{x}_{ppv} . Nous avons normalisé la distance par rapport à x_{\max}^i et x_{\min}^i , les valeurs maximum et minimum de l'ensemble d'apprentissage pour l'axe i , pour que la distance prenne des valeurs comprises dans l'intervalle $[0, 1]$ quel que soit le chemin d'évolution :

$$d_n^i(\underline{x}, \underline{x}_{ppv}) = \frac{d^i(\underline{x}, \underline{x}_{ppv})}{x_{\max}^i - x_{\min}^i} \quad (6.2)$$

$\pi_j^i(\underline{x}_{ppv})$ est la possibilité d'appartenance du plus proche voisin \underline{x}_{ppv} de l'observation \underline{x} , à la classe j par rapport à l'attribut i . $\pi_j^i(\underline{x}_{ppv})$ caractérise la pente de la fonction d'évolution par rapport à la distance normalisée, lorsque $\pi_j^i(\underline{x}_{ppv})$ augmente, la pente diminue afin d'indiquer une influence plus forte du plus proche voisin sur les points d'évolution.

proposition 1

Plus la valeur d'appartenance du plus proche voisin est faible, plus la fonction d'évolution prend des valeurs faibles :

$$\pi_j^i(\underline{x}_{ppv}) \Big|_1 = \pi_1 > \pi_j^i(\underline{x}_{ppv}) \Big|_2 = \pi_2 \Rightarrow \forall d_n^i(\underline{x}, \underline{x}_{ppv}) \in [0, 1] : v_j^i(\underline{x}) \Big|_{\pi_1} > v_j^i(\underline{x}) \Big|_{\pi_2} \quad (6.3)$$

Preuve :

Supposons que la possibilité d'appartenance du plus proche voisin \underline{x}_{ppv} du point rejeté \underline{x} , par rapport à la classe j et suivant l'attribut i , est $\pi_j(\underline{x}_{ppv}) = \pi_1$, la valeur d'évolution du point \underline{x}

est $v_j^i(\underline{x}) \Big|_{\pi_1} = (1 - d_n^i(\underline{x}, \underline{x}_{ppv}))^{\frac{c}{\pi_1}}$ Supposons maintenant que la valeur de possibilité du \underline{x}_{ppv}

est égale à $\pi_2 < \pi_1$, la valeur d'évolution du point \underline{x} est alors $v_j^i(\underline{x}) \Big|_{\pi_2} = (1 - d_n^i(\underline{x}, \underline{x}_{ppv}))^{\frac{c}{\pi_2}}$.

Comme $0 \leq (1 - d_n^i(\underline{x}, \underline{x}_{ppv})) \leq 1$ et $\frac{c}{\pi_1} < \frac{c}{\pi_2} > 1$ alors :

$(1 - d_n^i(\underline{x}, \underline{x}_{ppv}))^{\frac{c}{\pi_1}} > (1 - d_n^i(\underline{x}, \underline{x}_{ppv}))^{\frac{c}{\pi_2}}$, ce qui signifie que $v_j^i(\underline{x}) \Big|_{\pi_1} > v_j^i(\underline{x}) \Big|_{\pi_2}$.

Le nombre de classe influence également la fonction d'évolution, plus il est grand et plus la fonction d'évolution décroît rapidement. Cela est nécessaire puisque l'augmentation du nombre de classes diminue la zone d'évolution.

La fonction d'évolution vérifie les conditions suivantes :

$$- 0 \leq v_j^i(\underline{x}) \leq 1$$

$$- v_j^i(\underline{x}) = 1 \text{ si } d_n^i(\underline{x}, \underline{x}_{ppv}) = 0$$

$$- \lim_{d_n^i(\underline{x}, \underline{x}_{ppv}) \rightarrow 1} v_j^i(\underline{x}) = 0$$

$$- \forall \underline{x}_m, \underline{x}_k : \text{si } d_n^i(\underline{x}_m, \underline{x}_{ppv}) < d_n^i(\underline{x}_k, \underline{x}_{ppv}) \Rightarrow v_j^i(\underline{x}_m) > v_j^i(\underline{x}_k).$$

La Figure 6.5.a montre la décroissance de la fonction d'évolution par rapport à la distance normalisée pour différentes valeurs de $\pi_j^i(\underline{x}_{ppv})$ et la Figure 6.5.b montre cette décroissance pour différents nombres de classes.

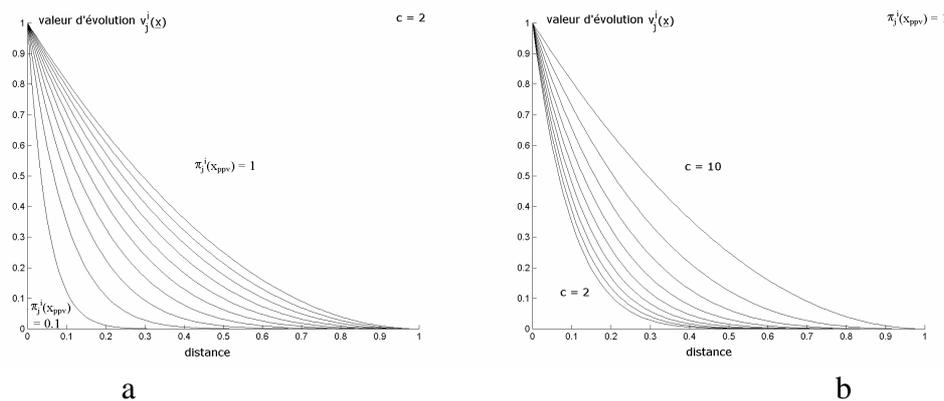


Figure 6.5. Décroissance de la fonction d'évolution par rapport à la distance normalisée a) pour différentes valeurs de $\pi_j^i(\underline{x}_{ppv})$, b) pour différents nombres de classes.

6.4.1.2. Matrice et vecteur d'évolution

Chaque observation rejetée possède une matrice d'évolution EV dont les composantes sont les représentativités de l'observation par rapport à chaque attribut de chaque classe. Cette matrice est de dimension (c,a) :

$$EV = \begin{pmatrix} v_1^1 & v_1^2 & \dots & v_1^i & \dots & v_1^a \\ v_2^1 & v_2^2 & \dots & v_2^i & \dots & v_2^a \\ \dots & \dots & \dots & \dots & \dots & \dots \\ v_j^1 & v_j^2 & \dots & v_j^i & \dots & v_j^a \\ \dots & \dots & \dots & \dots & \dots & \dots \\ v_c^1 & v_c^2 & \dots & v_c^i & \dots & v_c^a \end{pmatrix} \quad (6.4)$$

Il nous faut trouver un opérateur d'agrégation qui fusionne les différentes informations provenant des différentes sources ou attributs. En regardant la Figure 6.6, l'observation \underline{x} est loin de la classe 1 par rapport à l'attribut 1 et proche de cette classe par rapport à l'attribut 2. Lors de l'agrégation, on ne doit conserver que l'information provenant du premier attribut afin de tenir compte de l'éloignement de \underline{x}_{ppv} . Pour cela, il faut choisir l'attribut pour lequel la distance entre \underline{x} et \underline{x}_{ppv} est maximum, il faut donc utiliser l'opérateur minimum pour les valeurs d'évolution. L'agrégation de ces valeurs pour tous les attributs donne une seule valeur caractérisant la représentativité de \underline{x} par rapport à chaque classe j :

$$v_j = \min(v_j^1, v_j^2, \dots, v_j^i, \dots, v_j^a) \quad (6.5)$$

La matrice d'évolution devient alors un vecteur d'évolution de dimension (c,1) qui donne la représentativité de l'observation par rapport aux différentes classes :

$$ev = \begin{pmatrix} v_1 \\ v_2 \\ \dots \\ v_j \\ \dots \\ v_c \end{pmatrix} \tag{6.6}$$

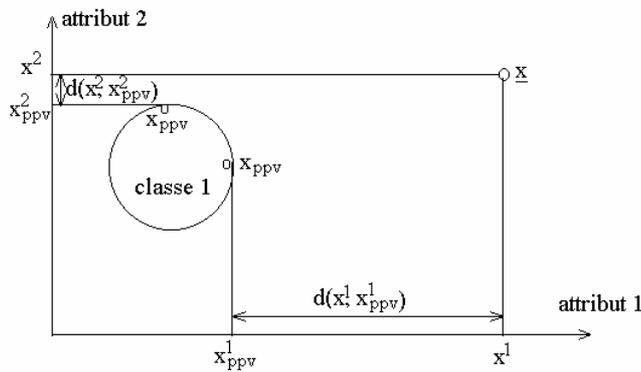


Figure 6.6. Illustration du choix de l'opérateur d'agrégation.

Un état d'évolution se caractérise par un chemin de transition entre les classes. Ce chemin est déterminé par une suite chronologique d'observations. Pour cette raison, la détection de l'état d'évolution nécessite un nombre minimum de points n_e . Soit la suite chronologique d'observations $\{ \underline{x}_k : k = 1.. n_e \}$ pour lesquelles la matrice chronologique d'évolution est $\{ ev_1, ev_2, \dots, ev_{n_e} \}$. Le vecteur chronologique VC_j formé des valeurs d'évolution de toutes les observations par rapport à la classe j , est calculé par :

$$VC_j = (v_j(x_1), v_j(x_2), \dots, v_j(x_{n_e})) \tag{6.7}$$

Chaque valeur du vecteur chronologique indique la situation du système vis-à-vis de la classe j à l'instant t_k , pour $k = 1 .. n_e$. Les différents VC_j , pour $j = 1 .. c$, forment des signaux temporels discrets monodimensionnels traduisant le comportement du système par rapport aux états connus. La Figure 6.7 montre un cas d'évolution entre deux classes a et b et la Figure 6.8 montre les signaux VC_a et VC_b . Le passage de la classe a à la classe b se traduit par une décroissance significative de VC_a . Cette décroissance s'accompagne éventuellement d'une croissance symétrique de VC_b . La pente de ces signaux traduit la vitesse d'évolution du système. Si le changement est très rapide, voire brutal, l'évolution se traduit par une rupture

de ces signaux aux instants de transition.

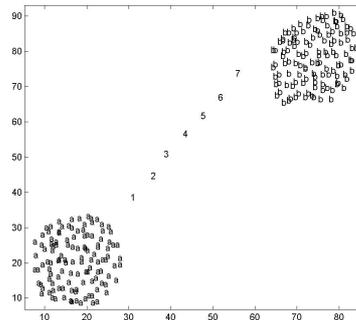


Figure 6.7. Evolution de la classe a vers la classe b.

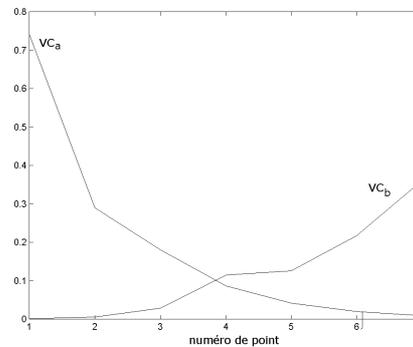


Figure 6.8. Signaux VC_a et VC_b pour l'exemple de la Figure 6.7.

6.4.1.3. Vitesse d'évolution

Il existe de nombreuses méthodes de détection de changement dans un signal [BAS93]. L'utilisation de ces méthodes nécessite la connaissance ou l'estimation de la loi de densité de probabilité, ce qui n'est pas possible avec FPM. Pour détecter cette évolution, nous introduisons une mesure de vitesse moyenne pour chaque classe :

$$VE_j = \frac{1}{n_e - 1} \sum_{k=1}^{k=n_e-1} (v_j(\underline{x}_{k+1}) - v_j(\underline{x}_k)) \quad (6.8)$$

VE_j prend une valeur négative si l'évolution part de la classe j et une valeur positive si l'évolution aboutit à la classe j . La vitesse moyenne d'évolution estime l'éloignement ou le rapprochement par rapport à une classe en fonction du plus proche voisin de la séquence d'évolution. Nous introduisons une règle de détection d'évolution en supposant que les observations dans la suite chronologique sont indépendantes :

$$\begin{aligned}
DE_j &= -1 \text{ si } VE_j < 0 \\
DE_j &= 1 \text{ si } VE_j > 0 \\
DE_j &= 0 \text{ sinon}
\end{aligned} \tag{6.9}$$

$DE_j = -1$ signifie que la suite chronologique s'éloigne de la classe j et $DE_j = 1$ signifie que la suite chronologique s'approche de la classe j . Détecter une évolution de la classe j vers la classe m nécessite donc deux décisions : $DE_j = -1$ et $DE_m = 1$. Pour l'exemple de la Figure 6.7, $VE_a = -0.12$ et $VE_b = 0.06$ ce qui signifie une évolution de la classe a vers la classe b .

Lorsqu'il y a rapprochement ou éloignement simultané de plusieurs classes, on essaie de trouver la classe sur laquelle le rapprochement ou l'éloignement est le plus rapide. Pour cela, on choisit la classe pour laquelle la vitesse d'évolution VE est la plus élevée. La prédiction de l'évolution fournit un couple de variables (f, t) , f est l'état de départ et t est l'état d'arrivée. Ce couple est calculé de la manière suivante :

$$(f, t) = \begin{cases} VE_f = \min(VE_1, VE_2, \dots, VE_c) \\ VE_t = \max(VE_1, VE_2, \dots, VE_c) \end{cases} \tag{6.10}$$

L'inconvénient de la vitesse moyenne est sa sensibilité au bruit ou points aberrants. Pour remédier à cet inconvénient, nous introduisons un seuil d'évolution se_j pour chaque classe j qui indique le taux maximum de points ayant le même sens d'évolution, éloignement ou rapprochement :

$$\begin{aligned}
se_j &= \max(psp_j, psn_j) \\
psp_j &= \frac{1}{n_e - 1} \cdot \sum_{k=1}^{n_e - 1} (\text{sign}((v_j(\underline{x}_{k+1}) - v_j(\underline{x}_k)) > 0)) \\
psn_j &= \frac{1}{n_e - 1} \cdot \sum_{k=1}^{n_e - 1} (\text{sign}((v_j(\underline{x}_{k+1}) - v_j(\underline{x}_k)) < 0)) = 1 - psp_j
\end{aligned} \tag{6.11}$$

où psp_j est le nombre moyen de points ayant une évolution positive caractérisant un rapprochement par rapport à la classe j et psn_j est le nombre moyen de points ayant une évolution négative caractérisant un éloignement. Une valeur de se_j égale à 0.5 signifie soit un cas de stabilisation ou autrement dit un cas d'apparition d'une nouvelle classe, soit un cas de bruit total autrement dit que la suite chronologique ne contient que des points aberrants. La distinction entre ces deux cas est faite par la vitesse moyenne qui est beaucoup plus petite en cas d'apparition d'une nouvelle classe. En effet le comportement aléatoire du bruit se caractérise par l'apparition des points dans des positions très différentes de l'espace de représentation. Pour l'exemple de la Figure 6.7, $(psp_b = 1, psn_b = 0)$ et $(psp_a = 0, psn_a = 1)$ ce

qui confirme une évolution de la classe a vers la classe b . Lorsque se_j s'approche de 1, la certitude d'un cas d'évolution augmente.

La confiance accordée à la prédiction des états de départ et d'arrivée augmente avec le nombre de points d'évolution n_e , mais le temps nécessaire pour la prédiction augmente également. Cette augmentation du temps est un handicap quand l'évolution est rapide. Lorsque la dynamique du système est faible, on pourra augmenter la longueur de la séquence d'observations. Enfin, la sous estimation de la longueur de la séquence réduit le retard de détection, et améliore donc la performance de détection, mais elle augmente les fausses alarmes.

6.4.1.4. Points de bruit dans la séquence d'évolution

Les points de bruit se distinguent par un changement brutal des valeurs d'évolution par rapport aux autres points de la séquence chronologique. La Figure 6.9 montre un exemple d'apparition de points aberrants pendant l'évolution de la classe a vers la classe b et la Figure 6.10 montre les vecteurs chronologiques VC_a et VC_b résultants. Nous remarquons clairement que le changement du sens de variation de l'évolution se produit à partir du point 4, premier point de bruit. L'évolution redevient normale à partir du point 5 : décroissance pour VC_a et croissance pour VC_b . Le même phénomène s'observe pour le point 7. Les vitesses moyennes d'évolution $VE_a = -0.08$ et $VE_b = 0.035$ montrent toujours une évolution de la classe a vers la classe b . Les nombres moyens de points ayant le même sens d'évolution, pour les deux classes a et b , sont ($psp_a = 0.22$, $psn_a = 0.78$) et ($psp_b = 0.78$, $psn_b = 0.22$) ce qui signifie que la majorité des points s'éloigne de la classe a et s'approche de la classe b .

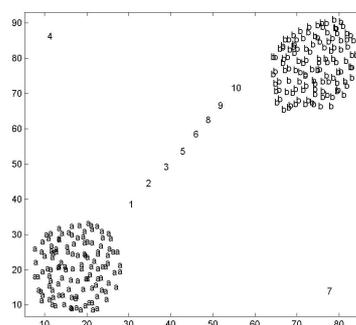


Figure 6.9. Apparition de deux points aberrants pendant l'évolution de la classe a vers la classe b .

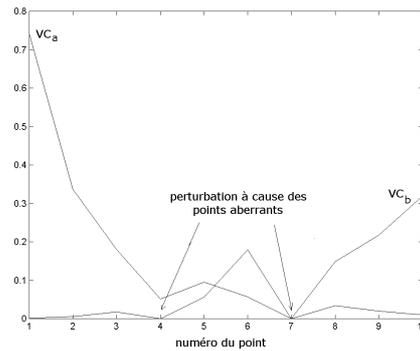


Figure 6.10. Vecteurs chronologiques VC_a et VC_b pour le cas d'apparition de deux points aberrant pendant l'évolution de la classe a vers la classe b .

6.4.2. Détection de nouveaux modes de fonctionnement

6.4.2.1. Détection d'un seul mode de fonctionnement

Un point rejeté peut correspondre à un des trois cas suivants :

- c'est un point de bruit,
- c'est un point caractérisant un état d'évolution,
- c'est un point d'une nouvelle classe.

Si la suite chronologique d'observations représente l'apparition d'une nouvelle classe, les observations consécutives doivent traduire des comportements voisins. Plus les observations sont proches dans la suite chronologique, plus elles sont susceptibles de traduire le même comportement. En effet, ces observations doivent se localiser dans une même zone restreinte de l'espace de représentation. Cela peut être traduit par la notion de stabilité qui peut être vue comme un, ou plusieurs, palier dans les signaux chronologiques VC avec des oscillations autour de la valeur moyenne de ce palier. Ces oscillations traduisent les mouvements éventuels du système au sein du nouvel état. La Figure 6.11 montre l'apparition d'une nouvelle classe et la Figure 6.12 montre les vecteurs chronologiques qui guident sa détection. Les vitesses moyennes d'évolution sont $VE_a = 0.004$ et $VE_b = -0.00008$, ces valeurs très faible signifient qu'il n'y a pas d'évolution et que les points sont très proches les uns des autres par rapport aux classes connues. De plus, les nombres moyens de points ayant le même sens d'évolution, $(psp_a = 0.67, psn_a = 0.33)$ et $(psp_b = 0.45, psn_b = 0.55)$, confirment cette détection.

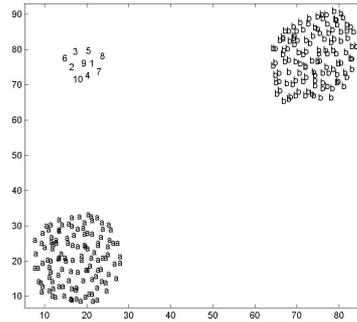


Figure 6.11. Apparition d'une nouvelle classe.

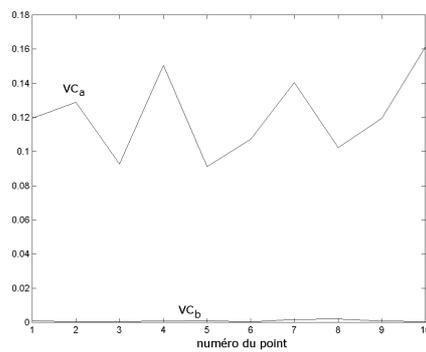


Figure 6.12. Signaux des vecteurs chronologiques VC des deux classes a et b pour l'exemple de la Figure 6.11.

Le bruit se traduit par des points apparaissant n'importe quand et n'importe où dans l'espace de représentation. Cela se caractérise aussi par des oscillations autour de la valeur moyenne des signaux des vecteurs chronologiques. L'amplitude de ces oscillations est beaucoup plus élevée que celle des points d'une nouvelle classe. La Figure 6.13 montre l'effet d'apparition de bruit pendant la formation d'une nouvelle classe et la Figure 6.14 montre les vecteurs chronologiques VC pour les deux classes. Nous pouvons bien distinguer les points de bruit des points de la nouvelles classe.

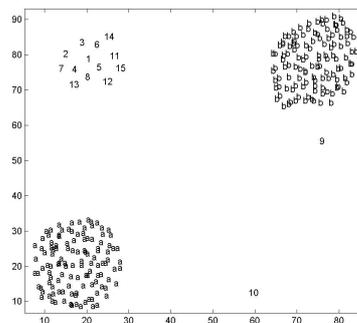


Figure 6.13. Apparition de deux points de bruit pendant la formation d'une nouvelle classe.

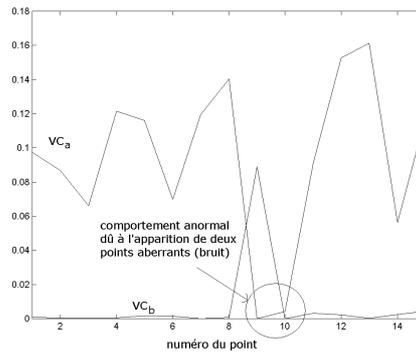


Figure 6.14. Vecteurs chronologiques pour le cas d'apparition de bruit pendant la formation d'une nouvelle classe pour l'exemple de la Figure 6.14.

6.4.2.2. Détection de deux modes de fonctionnement

Dans le cas d'apparition de plusieurs classes, les vecteurs chronologiques auront un palier ou une valeur moyenne pour chacune de celles-ci. La Figure 6.15 montre un cas d'apparition de deux nouvelles classes et la Figure 6.16 montre les vecteurs chronologiques correspondants. Pour les points 1 à 7, on constate que VC_a et VC_b forment deux paliers ce qui signifie l'apparition de deux nouvelles classes. VC_a , montre que la nouvelle classe $c1$ est proche de la classe a et éloignée de la classe b . Par contre VC_b montre que la deuxième nouvelle classe $c2$ est proche de la classe b et éloignée de la classe a .

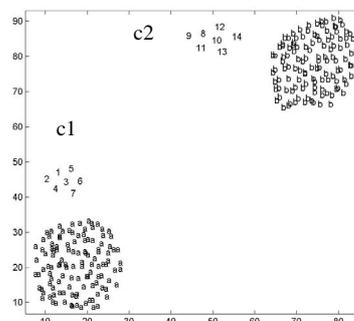


Figure 6.15. Apparition de deux nouvelles classes.

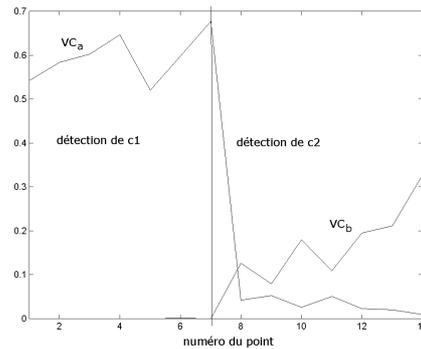


Figure 6.16. Détection de l'apparition de deux classes pour l'exemple de la Figure 6.15.

Les décisions finales peuvent être données sous formes symboliques, par exemple : le système évolue du mode de fonctionnement 1 vers le mode de fonctionnement 2 avec la vitesse moyenne d'éloignement VE_a et la vitesse moyenne de rapprochement VE_b ou le système se stabilise dans un nouveau mode de fonctionnement ou encore le système se stabilise dans deux, ou plusieurs, nouveaux modes de fonctionnement. Il faut noter que toute nouvelle classe créée devra être identifiée et validée par un expert ce qui entraîne une remise en cause de l'apprentissage.

Nous allons expérimenter le comportement de notre approche, pour prédire l'évolution vers des états futurs et détecter l'apparition des nouvelles classes, sur plusieurs exemples. Nous montrerons la robustesse de notre méthode vis à vis :

- d'un changement de la dynamique d'évolution : linéaire ou non,
- d'un changement des propriétés géométriques de la classe : forme sphérique ou allongée et différentes dispersions,
- d'un changement de la vitesse d'évolution,
- de l'existence de bruit.

6.4.3. Architecture d'un module de diagnostic prédictif et adaptatif en ligne

L'architecture du système de diagnostic adaptatif et prédictif en ligne, en utilisant FPM, est illustrée par la Figure 6.17. Nous remarquons que la décision est séquentielle d'ordre au moins 2, puisqu'on a besoin de mémoriser au moins 2 points rejetés successivement pour pouvoir détecter une évolution ou une apparition d'une nouvelle classe.

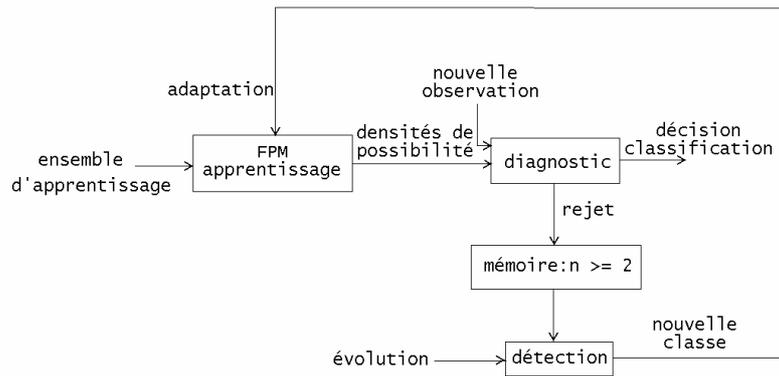


Figure 6.17. Architecture du système adaptatif et prédictif en ligne en utilisant FPM.

6.5. Expérimentation et validation

6.5.1. Exemple de simulation

6.5.1.1. Cas d'évolution

La Figure 6.18 montre un nuage comprenant trois classes repérées a , b et c dans un espace de représentation à deux attributs. Nous supposons que à l'instant t , le système commence à évoluer de la classe a vers la classe b . Cette évolution est représentée par la suite chronologique des points 1 à 8, rejetés en appartenance.

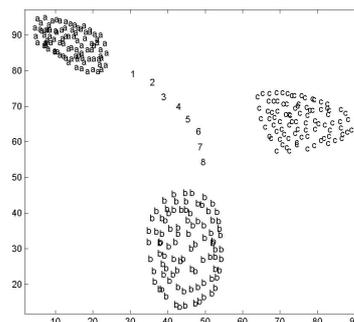


Figure 6.18. Evolution de la classe a vers la classe b représentée par la suite des points 1 à 8 rejetés en appartenance.

La Figure 6.19 montre les vecteurs chronologiques pour les trois classes. Les vitesses moyennes d'évolution sont : $VE_a = -0.066$, $VE_b = 0.064$ et $VE_c = 0.003$. La vitesse négative de la classe a signifie un éloignement. Les deux autres vitesses positives signifient un rapprochement vers les classes b et c . Puisque la vitesse moyenne d'évolution de la classe b

est plus élevée que celle de la classe c , nous prédisons donc l'évolution de la classe a vers la classe b . Les nombres moyens de points ayant le même sens d'évolution, pour les 3 classes, $(psp_a = 0, psn_a = 1)$, $(psp_b = 1, psn_b = 0)$ et $(psp_c = 1, psn_c = 0)$, confirment le cas d'évolution.

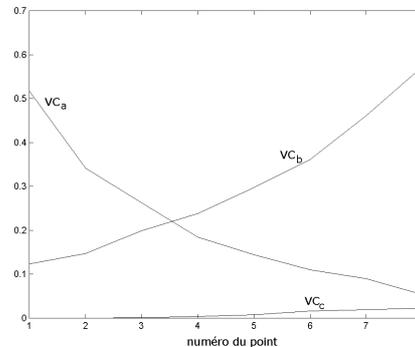


Figure 6.19. Vecteurs chronologiques VC pour les classes a , b et c , de la Figure 6.18.

L'évolution est déterminée par un comportement monotone sur une suite de points rejetés. Ce comportement est caractérisé par la diminution de la valeur d'évolution de la classe d'origine : éloignement, et par l'augmentation de la valeur d'évolution de la classe destinataire : rapprochement, pendant une durée de temps. Cette évolution a été détectée, pour l'exemple 1, en utilisant huit points. Le nombre de points d'évolution est déterminé par l'expert, en fonction : de la vitesse d'éloignement des classes, de la différence entre les vitesses de rapprochement et du nombre de points de bruit. Pour notre exemple, la vitesse de rapprochement vers la classe b est 20 fois plus grande que vers la classe c . La confiance en notre prédiction diminue quand cette différence diminue. A partir d'un certain niveau, on doit utiliser plus de points d'évolution pour accorder assez de confiance à notre prédiction.

6.5.1.2. Cas d'évolution accompagnée de bruit

Les points de bruit peuvent être détectés par l'instabilité du comportement des valeurs d'évolution des différentes classes. La Figure 6.20 montre un cas d'apparition de points de bruit pendant l'évolution de la classe a vers la classe b et la Figure 6.21 montre leurs vecteurs chronologiques. Nous pouvons voir le comportement anormal de ces vecteurs pendant l'apparition des points de bruit. Les vitesses moyennes d'évolution, $VE_a = 0.048$, $VE_b = 0.039$ et $VE_c = 0.0017$, et les nombres moyens de points ayant le même sens d'évolution pour les 3 classes, $(psp_a = 0.1, psn_a = 0.9)$, $(psp_b = 0.9, psn_b = 0.1)$ et $(psp_c = 0.8, psn_c = 0.2)$, confirment toujours un cas d'évolution de la classe a vers la classe b .

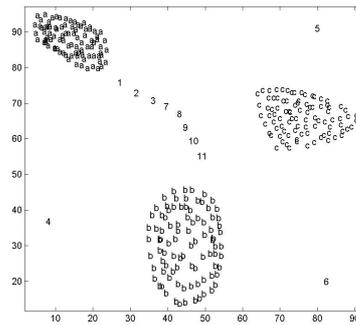


Figure 6.20. Apparition de trois points de bruit pendant l'évolution de la classe *a* vers la classe *b*.

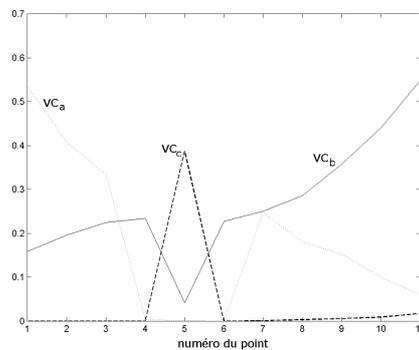


Figure 6.21. Détection de l'apparition des points de bruit pendant l'évolution de la classe *a* vers la classe *b* pour l'exemple de la Figure 6.20.

6.5.1.3. Cas d'apparition d'une nouvelle classe

La Figure 6.22 montre un cas d'apparition d'une classe et la Figure 6.23 montre sa détection. Nous remarquons que les vecteurs chronologiques sont stables avec des oscillations de faible amplitude. Les vitesses moyennes d'évolution, $VE_a = -0.002$, $VE_b = 0.013$ et $VE_c = 0.000$ très faibles, et les nombres moyens de points ayant le même sens d'évolution, $psp_a = 0.55$, $psn_b = 0.55$ et $psp_c = 0.55$, voisins de 0.5, montrent un cas d'apparition d'une classe.

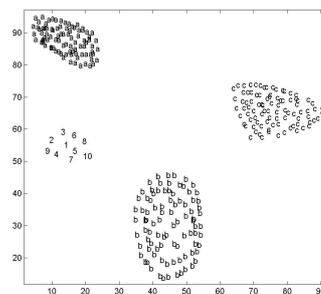


Figure 6.22. Apparition d'une nouvelle classe.

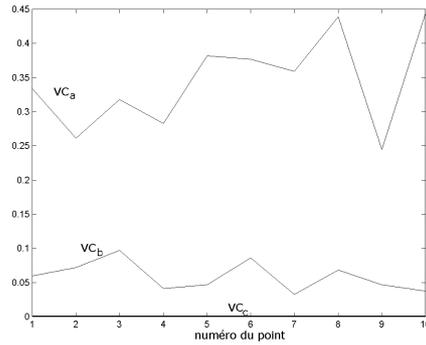


Figure 6.23. Détection de l'apparition d'une classe pour l'exemple de la Figure 6.22.

6.5.1.4. Cas d'apparition de deux nouvelles classes

La Figure 6.24 montre une suite d'observations formant deux nouvelles classes $c1$ et $c2$. La Figure 6.25 montre que les vecteurs chronologiques VC_a et VC_b prennent des valeurs importantes lors de l'apparition des points constituant la classe $c1$. Il en est de même pour VC_c lors de l'apparition des points de $c2$.

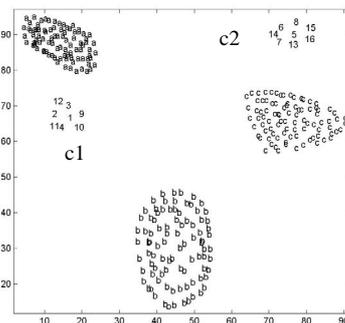


Figure 6.24. Apparition de deux nouvelles classes.

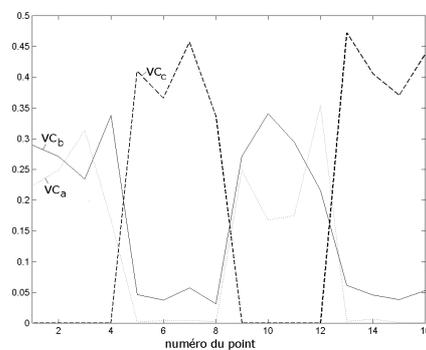


Figure 6.25. Détection de l'apparition de deux nouvelles classes pour l'exemple de la Figure 6.24.

6.5.1.5. Cas d'évolution vers une nouvelle classe

La Figure 6.26 montre une évolution depuis une classe connue : la classe a, suivie d'une suite d'observations formant une nouvelle classe. Notre méthode, essentiellement grâce au vecteur chronologique VC_a , détecte ce cas comme le montre la Figure 6.27.

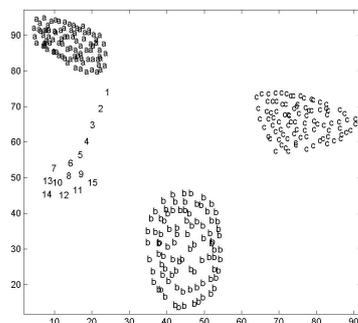


Figure 6.26. Suite d'observations comprenant une évolution suivie d'apparition d'une nouvelle classe.

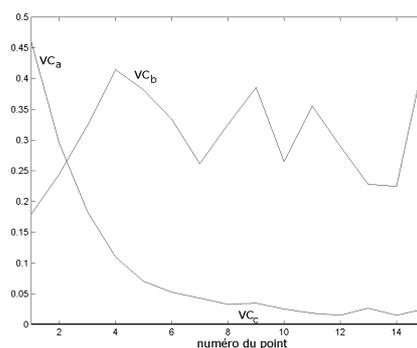


Figure 6.27. Détection de l'évolution de la classe a vers une nouvelle classe pour l'exemple de la Figure 6.26.

6.5.2. Exemple des données Matières Plastiques

Il se peut que dans le futur, un nouveau type de bouteille plastique apparaisse sur le marché. Si les deux attributs utilisés sont suffisamment discriminants, ce nouveau type sera représenté par une nouvelle classe. La Figure 6.28 montre un exemple d'apparition d'une nouvelle classe et la Figure 6.29 montre sa détection. On peut remarquer que la classe *b* est la plus proche de cette nouvelle classe alors que la classe *c* est la plus distante. Cela signifierait que les propriétés de transmission infra-rouge de ce nouveau type de plastique sont proches de celles du PET et éloignées de celles du PEHD, pour les deux longueurs d'onde sélectionnées. Les vitesses moyennes, $VE_a = 0.0008$, $VE_b = -0.0087$ et $VE_c = -0.0004$, et les nombres

moyens de points ayant le même sens d'évolution, ($psp_a = 0.57$, $psn_a = 0.43$), ($psp_b = 0.43$, $psn_b = 0.57$) et ($psp_c = 0.71$, $psn_c = 0.29$), montrent un cas d'apparition d'une nouvelle classe.

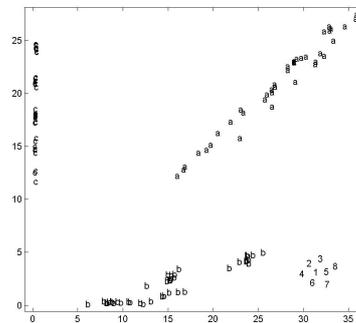


Figure 6.28. Apparition d'une nouvelle classe ou d'un nouveau type de matière plastique.

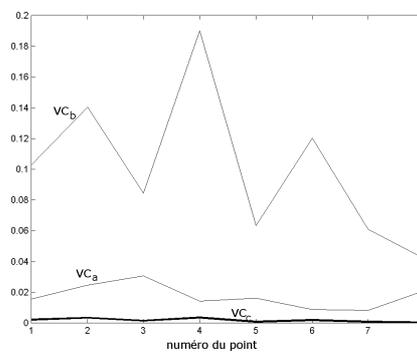


Figure 6.29. Détection de l'apparition d'une classe pour l'exemple de la Figure 6.28.

Supposons maintenant que les deux classes a et b sont connues a priori et essayons de détecter l'apparition de la troisième classe. La Figure 6.30 montre les deux classes connues a priori et la suite chronologique, déterminée aléatoirement, des points de la classe c qui sont tous rejetés. La Figure 6.31 montre les signaux chronologiques utilisés pour détecter l'apparition de la troisième classe. Nous pouvons voir que les vitesses moyennes d'évolution, par rapport aux deux classes, $VE_a = -0.00002$ et $VE_b = -0.00004$, sont très petites ce qui signifie le cas d'apparition d'une classe bien concentrée. Les nombres moyens de points ayant le même sens d'évolutions, ($psp_a = 0.49$, $psn_a = 0.51$) et ($psp_b = 0.54$, $psn_b = 0.46$), confirment l'apparition d'une nouvelle classe.

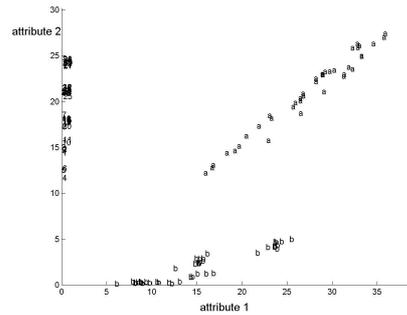


Figure 6.30. Classes *a* et *b* connues a priori et classe *c* inconnue dont tous les points sont rejetés.

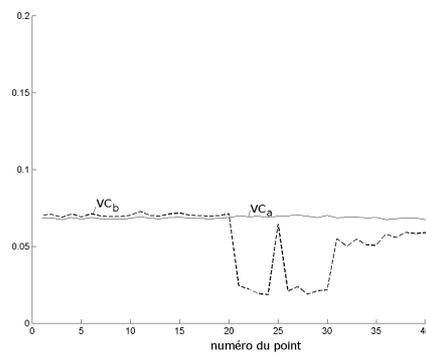


Figure 6.31. Détection de l'apparition de la classe *c* de la Figure 6.30.

Supposons maintenant que la classe *a* est la classe inconnue et les deux classes *b* et *c* sont connues a priori comme le montre la Figure 6.32. La Figure 6.33 montre la détection de la classe *a*, nous pouvons voir que les vitesses moyennes d'évolution, $VE_a = -0.0004$ et $VE_b = -0.0004$ sont beaucoup plus élevées que celles de la Figure 6.30, ce qui signifie que la classe détectée est beaucoup plus allongée. Les nombres moyens de points ayant le même sens d'évolution, ($psp_a = 0.54$, $psn_a = 0.46$) et ($psp_b = 0.54$, $psn_b = 0.46$), confirment le cas de détection d'une nouvelle classe.

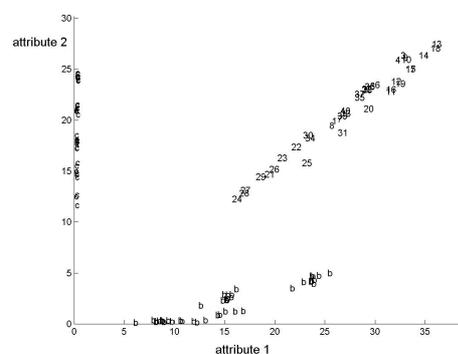


Figure 6.32. Classes *b* et *c* connues a priori et classe *a* inconnue, tous ses points sont donc rejetés.

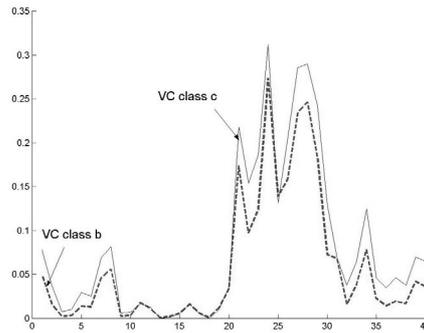


Figure 6.33. Détection de l'apparition de la classe a de la Figure 6.32.

6.5.2.1. Cas d'apparition de deux nouvelles classes

Si la suite d'observations comprend deux nouveaux types de matière plastique : Figure 6.34, notre approche les détecte comme le montre la Figure 6.35. On peut constater que le deuxième type, correspondant à c_2 , est beaucoup plus proche du PEHD que le premier. Le vecteur chronologique VC_c est le plus discriminant entre les deux nouvelles classes puisque la classe c est très compacte par rapport aux deux autres classes. En revanche, VC_a est le vecteur le moins discriminant puisque la classe a est la plus allongée. L'apparition simultanée de deux classes donne une valeur moyenne des deux vitesses d'évolution pour ces deux classes. Nous constatons l'apparition de plusieurs nouvelles classes en regardant les vecteurs chronologiques, on peut remarquer deux niveaux de vitesses d'évolution, un pour chaque nouvelle classe. La détection d'apparition de plusieurs classes peut être faite visuellement pour n'importe quel nombre d'attributs puisque les signaux chronologiques sont toujours représentés dans un espace à deux dimensions. Le seuil d'évolution, quant à lui, ne peut pas nous informer sur le nombre de nouvelles classes apparues.

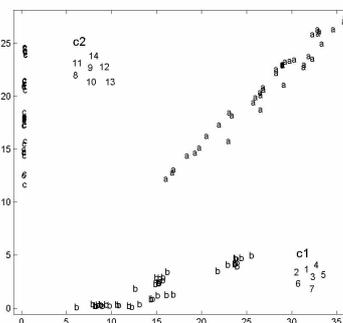


Figure 6.34. Apparition de deux nouvelles classes ou de deux nouveaux types de matière plastique.

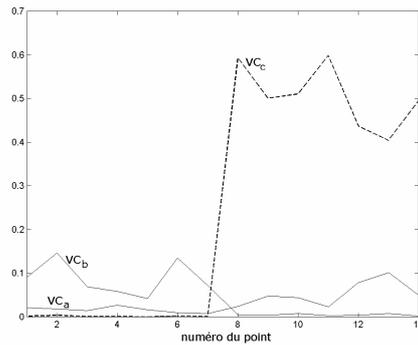


Figure 6.35. Détection de l'apparition de deux nouvelles classes pour l'exemple de la Figure 6.34.

6.5.2.2. Cas des points aberrants

La Figure 6.36 montre une suite d'observations où les points ne signifient ni un cas d'évolution ni une apparition d'une nouvelle classe. La Figure 6.37 montre que notre approche considère ce cas comme un cas de bruit. L'amplitude des oscillations est très élevée, autrement dit les vitesses moyennes d'évolution, $VE_a = -0.013$, $VE_b = 0.019$, $VE_c = -0.122$ sont relativement élevées par rapport à un cas d'apparition d'une nouvelle classe. Puisque la classe c est très compacte, elle distingue le mieux le cas de bruit puisqu'elle est la plus sensible à la distance, contrairement à la classe a qui est très allongée.

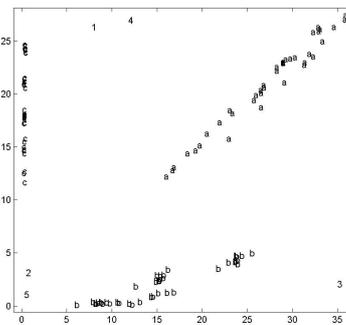


Figure 6.36. Cas de bruit.

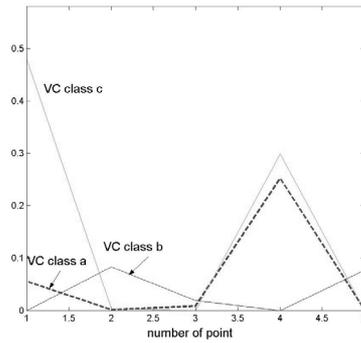


Figure 6.37. Détection du cas de bruit de la Figure 6.36.

6.5.2.3. Cas d'évolution linéaire

Nous allons générer une séquence d'évolution en utilisant une loi linéaire de la forme $y = y_0 + a_0 \cdot x$, où a_0 est une constante et y_0 indique l'état initial de l'évolution. La Figure 6.38.a montre un exemple d'une séquence d'évolution linéaire de la classe c vers la classe b et la Figure 6.39.a montre une séquence linéaire d'évolution de la classe c vers la classe a . La Figure 6.38.b et la Figure 6.39.b montrent les vecteurs chronologiques résultants. Pour le premier exemple les vitesses d'évolution sont $VE_a = 0.012$, $VE_b = 0.055$ et $VE_c = -0.081$, les nombres moyens de points ayant le même sens d'évolution sont $(psp_a = 1, psn_a = 0)$, $(psp_b = 1, psn_b = 0)$ et $(psp_c = 0, psn_c = 1)$. Pour le deuxième exemple les vitesses moyennes d'évolution sont : $VE_a = 0.04$, $VE_b = 0.003$ et $VE_c = -0.043$, les nombres moyens de points ayant le même sens d'évolution sont : $(psp_a = 1, psn_a = 0)$, $(psp_b = 1, psn_b = 0)$ et $(psp_c = 0, psn_c = 1)$. En conséquence, notre méthode détecte bien cette évolution pour les deux exemples.

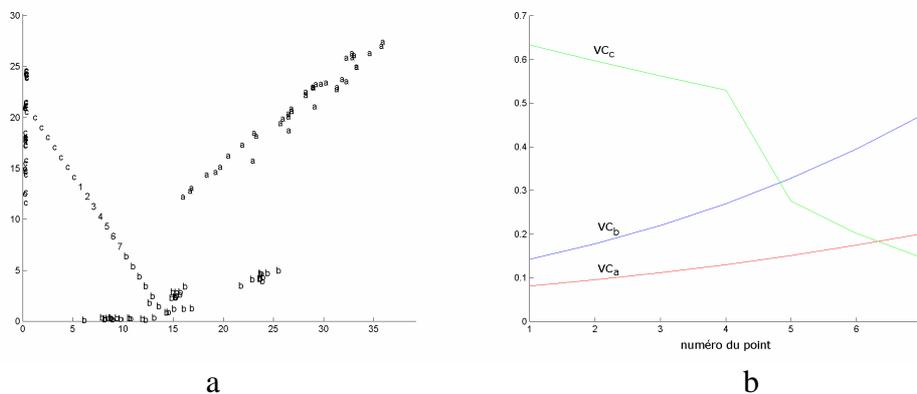


Figure 6.38. a) Séquence d'évolution linéaire de la classe c vers la classe b , b) signaux chronologiques résultants.

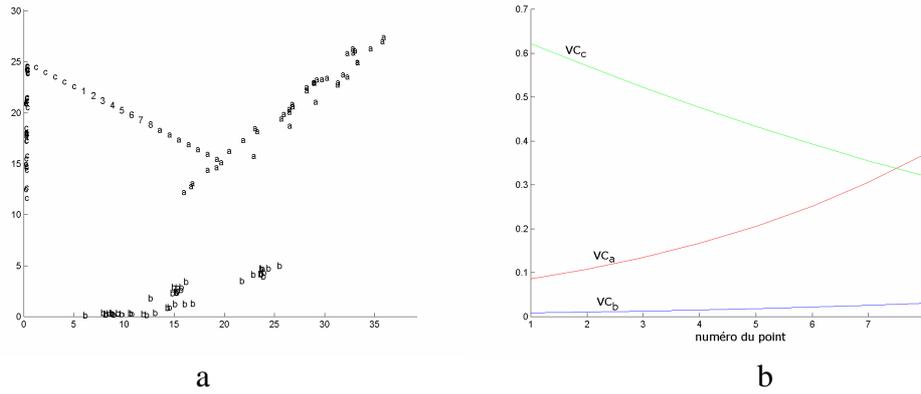


Figure 6.39. a) Séquence d'évolution linéaire de la classe c vers la classe a, b) signaux chronologiques résultants.

6.5.2.4. Cas d'évolution non linéaire

Nous générons une séquence d'évolution quadratique qui suit la loi $y = y_0 + a_0 \cdot x^2$ ou a_0 est une constante et y_0 indique l'état initial de l'évolution. La Figure 6.40.a montre un exemple d'une séquence d'évolution quadratique de la classe c vers la classe b et la Figure 6.41.a montre une séquence quadratique d'évolution de la classe c vers la classe a. La Figure 6.40.b et la Figure 6.41.b montrent les vecteurs chronologiques résultants. Pour le premier exemple les vitesses moyennes d'évolution sont $VE_a = 0.017$, $VE_b = 0.068$ et $VE_c = -0.117$, les nombres moyens de points ayant le même sens d'évolution sont : $(psp_a = 1, psn_a = 0)$, $(psp_b = 1, psn_b = 0)$ et $(psp_c = 0, psn_c = 1)$. Pour le deuxième exemple nous avons : $VE_a = 0.04$, $VE_b = 0.006$, $VE_c = -0.043$, $(psp_a = 1, psn_a = 0)$, $(psp_b = 1, psn_b = 0)$ et $(psp_c = 0, psn_c = 1)$. En conséquence, notre méthode détecte bien cette évolution pour les deux exemples.

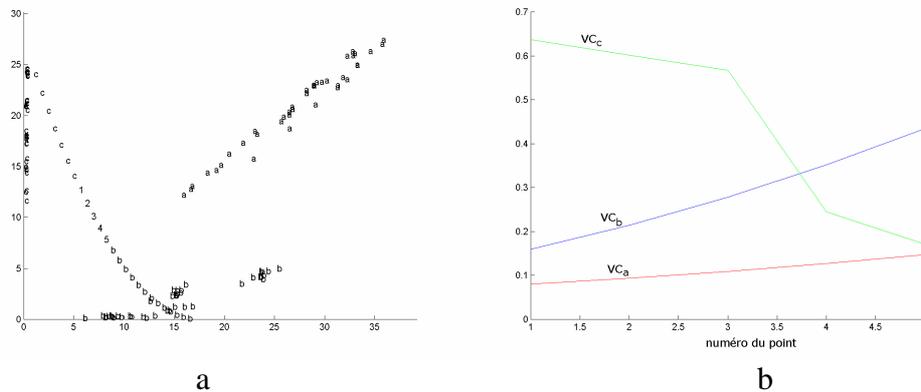


Figure 6.40. a) Séquence d'évolution quadratique de la classe c vers la classe b, b) vecteurs chronologiques résultants.

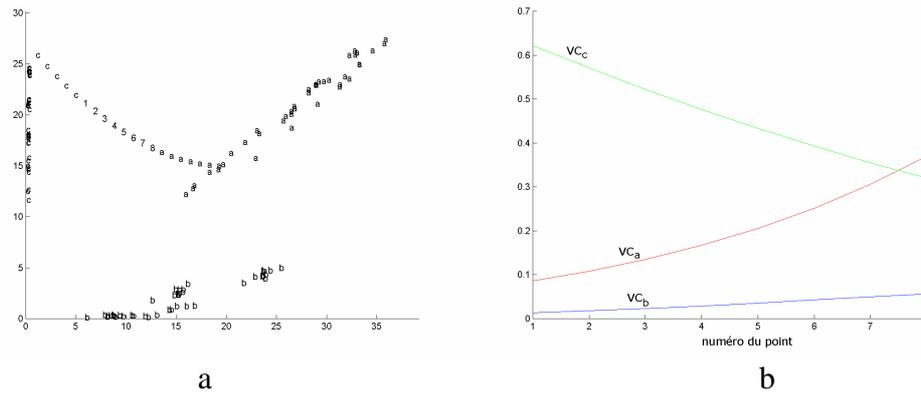


Figure 6.41. a) séquence d'évolution quadratique de la classe c vers la classe a, b) vecteurs chronologiques résultants.

6.5.3. Exemple des données Machine à Laver

6.5.3.1. Cas d'apparition d'une nouvelle classe

Supposons maintenant que les classes *a*, *b* et *c*, des données Machine à Laver, sont connues a priori, alors que la classe *d* est inconnue comme le montre la Figure 6.42. La Figure 6.43 montre les vecteurs chronologiques *VC* pour la détection de l'apparition de la nouvelle classe. Les vitesses moyennes d'évolution sont $VE_a = 0.0014$, $VE_b = 0.0000$, $VE_c = 0.0000$ et les nombres moyens de points ayant le même sens d'évolution sont : $(psp_a = 0.44, psn_a = 0.56)$, $(psp_b = 0.44, psn_b = 0.56)$ et $(psp_c = 0.38, psn_c = 0.62)$. En conséquence, notre méthode détecte bien l'apparition de la classe *d*.

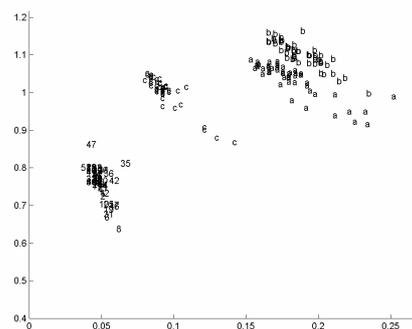


Figure 6.42. Classes *a*, *b* et *c* de l'exemple Machine à Laver connues a priori et classe *d* inconnue dont tous les points sont rejetés.

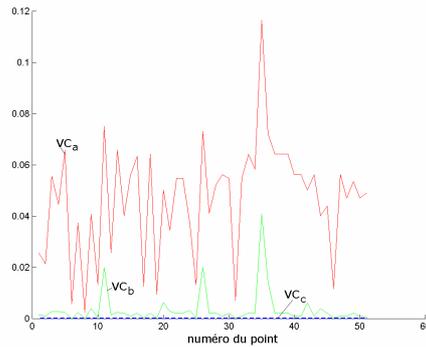


Figure 6.43. Détection de l'apparition de la classe d , de la Figure 6.42.

6.5.3.2. Cas d'apparition de deux nouvelles classes

La Figure 6.44 montre le cas où les deux classes c et d sont inconnues. La Figure 6.45 montre la détection de ces deux classes, nous pouvons distinguer les deux classes par les deux niveaux différents des vecteurs chronologiques des deux classes connues. Les vitesses d'évolution mesurées dans ce cas sont les vitesses moyennes pour les deux classes détectées. Cela nous empêche d'avoir une idée sur la vitesse moyenne d'évolution de chaque classe. les nombres moyens de points ayant le même sens d'évolution sont : ($psp_a = 0.46$, $psn_a = 0.54$) et ($psp_b = 0.44$, $psn_b = 0.56$), confirment le cas d'apparition d'une nouvelle classe mais ils ne déterminent pas le nombre de classes apparues.

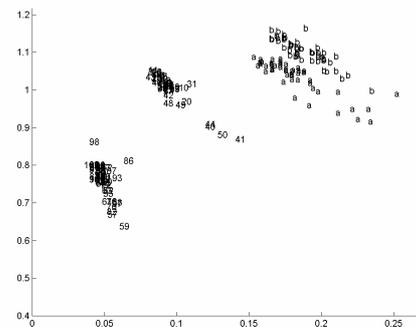


Figure 6.44. Classes a et b de l'exemple Machine à Laver connues a priori et classes c et d inconnues dont tous les points sont rejetés.

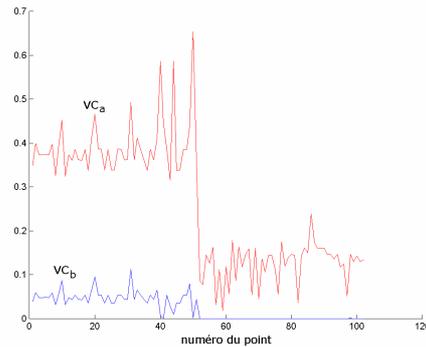


Figure 6.45. Détection de l'apparition des deux classes c et d , de la Figure 6.44.

6.5.4. Exemple des données Injection Plastique

L'intérêt d'appliquer notre approche sur les données Injection Plastique est de montrer que notre méthode est capable de prédire l'évolution et de détecter l'apparition des nouvelles classes pour des données représentées dans un espace de dimension supérieure à 2, en analysant les vecteurs chronologiques toujours dans un espace à 2 dimensions.

6.5.4.1. Cas d'apparition d'une nouvelle classe

Supposons que la classe e n'est pas connue a priori : Figure 6.46. La stabilité et les faibles oscillations des vecteurs chronologiques de la Figure 6.47 montrent que notre méthode permet de détecter cette apparition.

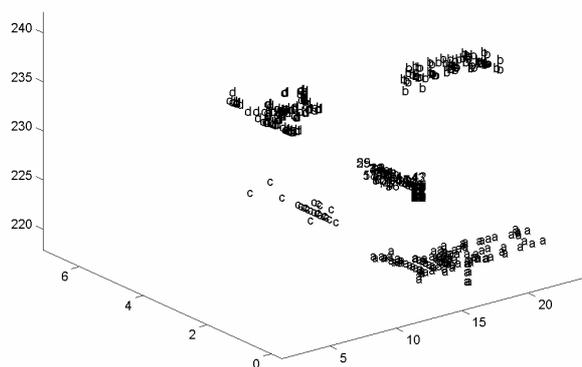


Figure 6.46. Classes a , b , c et d des données Injection Plastique connues a priori et classe e inconnue dont tous les points sont rejetés.

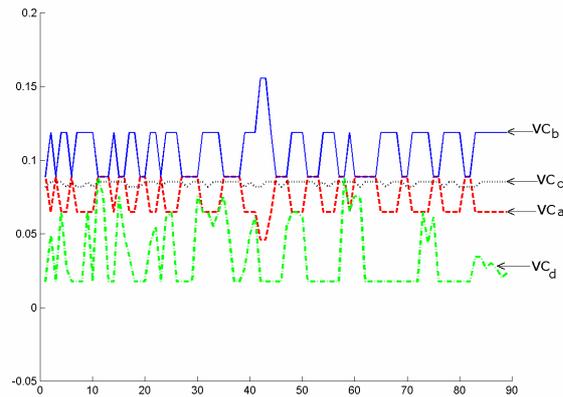


Figure 6.47. Détection de la classe e de la Figure 6.46.

6.5.4.2. Cas d'apparition de deux nouvelles classes

Supposons maintenant que les deux classes d et e ne sont pas connues a priori : Figure 6.48, la Figure 6.49 montre les vecteurs chronologiques résultants de l'application de notre méthode sur l'ensemble d'apprentissage contenant les trois autres classes. VC_a et VC_c montrent que la classe d est loin par rapport aux deux classe a et c , en revanche la classe d est proche de la classe b , VC_b ayant des valeurs plus élevées. Les oscillations importantes de VC_b sont dues à l'ordre d'arrivée des points de la classe d , en effet cet ordre étant aléatoire, la distance entre deux points successifs peut être importante.

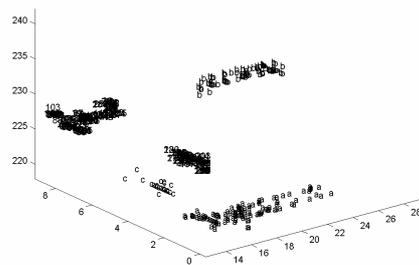


Figure 6.48. Classes a , b , et c des données Injection Plastique connues a priori et classes d et e inconnues, tous leurs points sont donc rejetés.

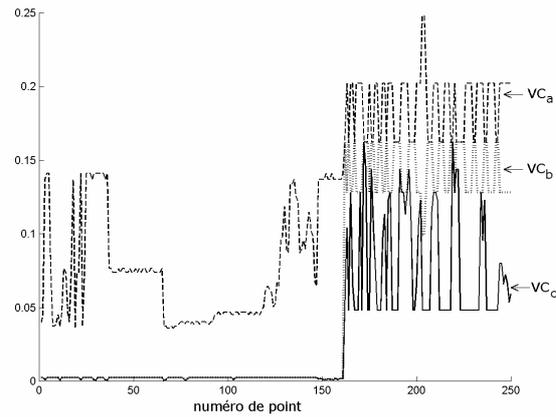


Figure 6.49. Détection des deux nouvelles classes de la Figure 6.48.

6.5.4.3. Cas d'apparition de quatre nouvelles classes

Ici, nous supposons que l'ensemble d'apprentissage des données Injection Plastique ne contient que la classe *a*. Notre méthode est capable, à partir du vecteur chronologique d'une seule classe, de détecter l'apparition de quatre nouvelles classes comme le montre la Figure 6.50. Nous constatons qu'il y a une confusion possible entre les deux classes *b* et *d* qui sont très proches par rapport à la classe *a*, notre méthode peut les considérer comme une seule classe surtout si elles apparaissent l'une après l'autre.

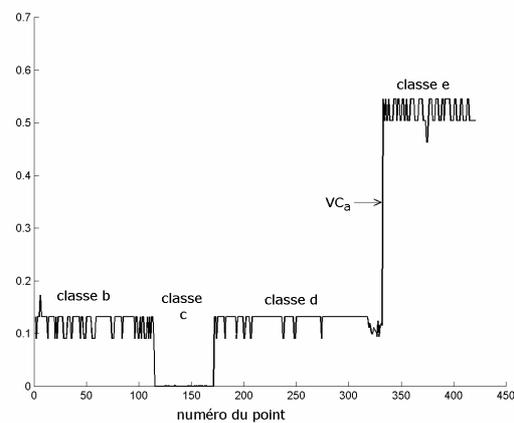


Figure 6.50. Détection de l'apparition de quatre nouvelles classes, *b*, *c*, *d* et *e*.

6.5.5. Exemple des données Iris

6.5.5.1. Cas d'apparition d'une seule classe

En supposant que la troisième classe des données Iris est inconnue a priori, l'application de notre méthode sur le reste du jeu de données donne les vecteurs chronologiques affichés de la Figure 6.51. Nous remarquons que le vecteur chronologique VC_b de la deuxième classe présente des oscillations de grande amplitude ce qui signifie l'apparition d'une classe allongée et très proche de la classe b . C'est ce qui se passe en réalité. Les vitesses d'évolution $VC_a = 0$ et $VC_b = 0.0031$ et les nombres moyens de points ayant le même sens d'évolution, ($psp_a = 0.52$, $psn_a = 0.48$) et ($psp_b = 0.52$, $psn_b = 0.48$) confirment cette détection.

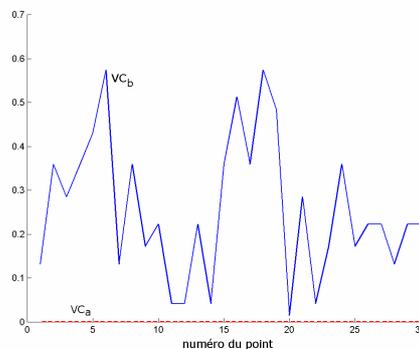


Figure 6.51. Détection de l'apparition de la classe c des données Iris.

6.5.5.2. Cas d'apparition de deux classes

En supposant que seulement la classe b est connue a priori, notre méthode reste capable de détecter l'apparition des deux classes a et c . La Figure 6.52 montre que la classe a est bien concentrée et est séparable ou éloignée de la classe b , en revanche la classe c est allongée et très proche de la classe b .

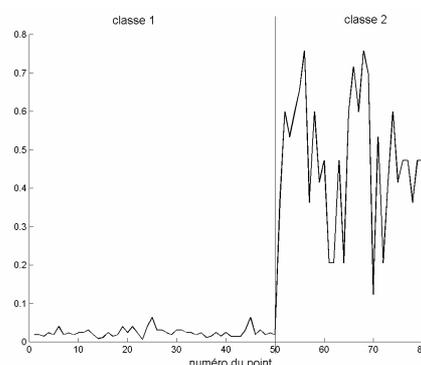


Figure 6.52. Détection de l'apparition de deux classes a et c des données Iris.

6.6. Conclusion

L'application de la RdF au diagnostic se heurte souvent au problème de la connaissance a priori insuffisante du système étudié. Dans une base de connaissance incomplète, tous les modes de fonctionnement ne sont pas représentés. En effet, pour des raisons de coût ou de sécurité, certains états défailants ou dangereux ne peuvent pas être provoqués. Un processus de diagnostic par RdF doit donc être adaptatif, c'est à dire capable de détecter les états inconnus et capable de les apprendre afin d'enrichir sa connaissance a priori.

De plus, il est moins onéreux d'anticiper l'évolution d'un mode de fonctionnement normal vers un mode anormal que d'attendre d'arriver à celui-ci. Cela évite les conséquences des dysfonctionnement.

L'évolution s'appuie sur une suite chronologique d'observations décrivant le comportement d'un système sur un intervalle de temps. Une évolution se traduit par des points s'éloignant de l'état origine et s'approchant de l'état destinataire. Cette évolution est caractérisée par un chemin entre les deux états. L'évolution a donc lieu dans la zone entre les classe, que nous appelons zone d'évolution. Cette zone contient les points rejetés par FPM, autrement dit les points qui ont une valeur d'appartenance nulle. Pour cette raison FPM n'est pas capable de prédire l'évolution entre les classes.

De plus, détecter un état inconnu nécessite un nombre minimum de points rejetés qui ne sont affectés à aucune classe. L'apparition d'une nouvelle classe se passe donc dans la zone d'évolution. Pour la même raison que précédemment FPM ne donne toujours pas d'information sur la situation géographique du point. En conséquence FPM n'est pas capable de détecter l'apparition de nouvelles classes.

Pour inclure un état inconnu dans sa base de connaissance, le module de diagnostic doit arrêter la phase de classification et utiliser une méthode de coalescence floue pour trouver la nouvelle structure des classes. Il est ensuite nécessaire de refaire toute la phase d'apprentissage. Cette procédure empêche le diagnostic adaptatif en temps réel. Il n'existe, à ce jour, aucune autre approche pour utiliser FPM dans un processus de diagnostic en ligne prédictif et adaptatif.

Nous proposons une nouvelle approche qui consiste à construire une fonction d'évolution caractérisant la situation géographique ou la représentativité d'une observation, vis à vis des classes connues, dans la zone d'évolution.

Chaque observation rejetée possède une matrice d'évolution EV dont les composantes sont la représentativité de l'observation par rapport à chaque attribut de chaque classe. Cette

matrice permet de déterminer un vecteur chronologique VC par classe connue. Chaque vecteur chronologique forme un signal temporel discret monodimensionnel traduisant le comportement du système par rapport aux états connus.

L'évolution d'une classe 1 vers une classe 2 se traduit par une décroissance significative des valeurs d'évolution de la classe 1 et d'une croissance symétrique des valeurs d'évolution de la classe 2. La pente VE de ces signaux traduit la vitesse d'évolution du système. Si le changement est très rapide, voire brutal, l'évolution se traduit par une rupture des signaux aux instants de transition. La confiance accordée à la prédiction des états de départ et d'arrivée dépend du nombre de points d'évolution n_e . Cette confiance augmente avec n_e mais le temps nécessaire pour la prédiction augmente également.

Si la suite chronologique d'observations représente l'apparition d'un nouvel état, les vecteurs chronologiques VC forment un palier avec des oscillations autour de la valeur moyenne. Ces oscillations traduisent les mouvements éventuels du système au sein du nouvel état.

Le cas de bruit se traduit par des points apparaissant n'importe quand et n'importe où dans l'espace de représentation. Cela se caractérise par des oscillations autour de la valeur moyenne des signaux chronologiques. Ces oscillations sont beaucoup plus élevées que celles observées pour des points d'une nouvelle classe.

Notre méthode ne prédit pas les valeurs futures des observations mais prédit la tendance future du système à diagnostiquer. Elle détecte un cas d'évolution ou d'apparition d'une nouvelle classe à partir des observations rejetées. L'interprétation est aisée car les vecteurs chronologiques sont toujours représentés dans un espace à deux dimensions.

Conclusion générale

Les recherches menées au cours de cette thèse nous ont permis d'élaborer un module de diagnostic en ligne adaptatif et prédictif pour des systèmes évolutifs. Ce module est basé sur la méthode de classification Fuzzy Pattern Matching (FPM). Cette méthode a été sélectionnée pour sa rapidité, sa simplicité et sa possibilité de traiter des données étant à la fois incertaines et imprécises. Toutefois cette méthode ne peut pas être intégrée dans ce module de diagnostic dans sa version classique pour plusieurs raisons :

- l'ensemble d'apprentissage ne contient pas souvent d'information suffisante sur la forme réelle des classes. La méthode FPM obtient l'information manquante en réalisant un apprentissage continu après la classification de chaque nouvelle observation. Le temps de calcul, augmentant avec la taille de l'ensemble d'apprentissage, empêche l'apprentissage continu en temps réel,
- la méthode FPM est inopérante pour des classes de formes non-sphériques en raisons de la représentation des classes attribut par attribut. Les classes des données provenant des systèmes industriels sont souvent de forme non-sphérique,
- les performances de FPM se dégradent en cas de présence de points aberrants,
- il n'existe aucune méthode pour déterminer les valeurs des paramètres de FPM afin d'optimiser ses performances. Ces paramètres sont déterminés heuristiquement ou expérimentalement,
- l'ensemble d'apprentissage contient rarement tous les modes de fonctionnement du système. FPM est incapable de détecter l'apparition des nouveaux modes et de les inclure à sa base de connaissance,
- l'anticipation de l'évolution d'un mode de fonctionnement normal vers un mode anormal est très utile pour les systèmes industriels afin d'éviter les conséquences des modes anormaux, voire dangereux, ce dont est incapable FPM.

En conséquence, nous avons proposé plusieurs adaptations afin que FPM soit apte à être utilisée dans ce module de diagnostic.

La première adaptation est la réalisation de l'apprentissage continu en utilisant une approche incrémentale. Cette dernière permet à FPM d'ajouter, à la base de connaissance, l'information apportée par chaque nouvelle observation classifiée sans répéter toute la phase

d'apprentissage.

La deuxième adaptation consiste à inclure une information sur la forme des classes dans la phase d'apprentissage afin de permettre à FPM d'être opérante dans le cas des classes de forme non sphérique. Cette information est représentée par une corrélation entre les différents attributs de l'espace de représentation. Nous avons appelé FPM utilisant cette adaptation : Fuzzy Pattern Matching Corrélée (FPMC).

Nous avons proposé une transformation probabilité-possibilité dont la spécificité varie par l'intermédiaire d'un paramètre, nous l'avons appelé la Transformation Variable (TV). Nous avons cherché à régler la spécificité de la TV afin de permettre à FPM d'éliminer les points aberrant et de détecter les zones de haute densité ou les sous-classes. Par conséquent FPM devient plus résistante au bruit et plus précise.

Nous avons proposé une mesure de séparabilité des classes adaptée pour FPM, nous l'avons utilisé comme indicateur des performances de FPM. Puis, nous avons cherché itérativement les valeurs optimales des paramètres de FPM qui maximisent cette mesure de séparabilité. De plus, nous avons suggéré d'utiliser FPM avec des barres de largeur Adaptative en fonction de l'étendue des classes par rapport à chaque attribut. Le but est d'améliorer la séparabilité en diminuant le temps de calcul. Un critère a été proposé pour déterminer si une nouvelle observation est bien ou mal classifiée. Cela permet à FPM de décider quand il faut ajouter l'information apporté par une nouvelle observation à la base de connaissance.

Enfin, nous avons proposé une approche pour analyser les observations rejetées afin de déterminer, en temps réel, si elles constituent de nouvelles classes ou s'il s'agit d'une évolution vers une classe connue ou inconnue.

Nous n'avons pas pu tester l'aspect prédictif de notre module de diagnostic sur des données réelles par manque d'une base de données recueillie sur un processus industriel.

Nous nous sommes limité à l'application de notre module, pour la supervision des systèmes industriels. Il est envisageable et intéressant d'utiliser ce module pour d'autres applications comme dans le domaine biomédical ou le domaine de la reconnaissance de la parole ou de l'image. De plus ce module utilise des connaissances de type numérique, il est envisageable d'adapter ce module pour qu'il utilise des connaissances de type symbolique. Cela pourrait être intéressant pour la construction des systèmes experts fonctionnant dans un environnement évolutif et avec une base de connaissance incomplète.

Bibliographie

[ANI00] ANIL K.J., ROBERT P.W., MAO J., Statistical Pattern Recognition: A review, *IEEE Transactions on pattern analysis and machine intelligence* 1 (22), (2000) 4-36.

[BAL67] BALL G. H., HALL D. J., A clustering technique summarizing multivariate data, *Behavioral Science* 12, (1967) 153-155.

[BEZ74] BEZDEK J. C., Cluster validity with fuzzy sets, *Journal of cybernetics* 3 (3), (1974) 58-73.

[BEZ78] BEZDEK J. C., HARRIS J. D., Fuzzy relations and partitions : an axiomatic basis for clustering, *Fuzzy Sets and Systems* 1, (1978) 11-27.

[BEZ81] BEZDEK J. C., Pattern recognition with fuzzy objective function algorithms, New-York, Plenum Press, 1981.

[BIL98] BILLAUDEL P., DEVILLEZ A., VILLERMAIN LECOLIER G., An unsupervised fuzzy classification algorithm for the non elliptic classes, *EURISCON'98*, (1998), Athens, Greece.

[BIL99] BILLAUDEL P., DEVILLEZ A., VILLERMAIN LECOLIER G., Unsupervised fuzzy classification method based on a proximity graph and on a graduated hierarchy, *8th IEEE International conference on fuzzy systems FUZZ-IEEE'99*, (1999), Seoul, Korea.

[BOU97] BOUDAUD N., *Conception d'un système de diagnostic adaptatif en ligne pour la surveillance des systèmes évolutifs*, Thèse présentée devant l'Université de Technologie de Compiègne, France, 1997.

[BOU96] BOUTLEUX E., Diagnostic et suivi d'évolution de l'état d'un système par reconnaissance des formes floues. Application au modèle du réseau téléphonique français, Thèse présentée devant l'Université de Technologie de Compiègne, 1996.

[CHE76] CHEN C. H., On information and distance measures, error bounds and feature

selection, *Inform. Sci.* 10, (1976) 159-173.

[CHO57] CHOW C. K., An optimum character recognition system using decision, *R.E. Transaction Electronic Computer* (1957) 247-254.

[COR02] CORN G., DUBUISSON B., Pattern characteristics of an evolution between two classes, *Fuzzy Sets and Systems* 126, (2002) 293-310.

[COV67] COVER T. M., HART P. E., Nearest neighbour pattern classification, *IEEE Trans. On information Theory* 13, (1967) 21-27.

[DEM93] DEMARTINE P., HERAULT J., Representation of nonlinear data structures through fast VQP neural network, *Neuronîmes*, (1993) 411-424.

[DEV93] DEVEUGHELE S., *Etude d'une méthode de combinaison adaptative d'informations incertaines dans un cadre possibiliste*, Thèse présentée devant l'Université de Technologie de Compiègne, France, 1993.

[DEV99] DEVILLEZ A., Contribution à la classification floue de données comportant des classes de forme quelconque Application au développement d'un module d'aide à la décision, Thèse présentée devant l'Université de Reims Champagne Ardenne, France, 1999.

[DID71] DIDAY E., La méthode des nuées dynamiques, *Revue de Statistiques Appliquées* 19, (1971) 19-34.

[DOC81] DOCTOR P. G., HARRINGTON T. P., DAVIS T. J., MORRIS C. J., FRALEY D. W., « Pattern recognition methods for classifying and sizing flaws using eddy current data », *Eddy current characterization of materials and structures, ASTM STP 722*, George Birnbaum and George Free Eds., American society for testing and materials, (1981) 464-483.

[DUB87] DUBOIS D. et PRADE H., Théorie des possibilités Application à la représentation des connaissances en informatique, Deuxième édition, Masson, 1987.

[DUB88] DUBOIS D., PRADE H., TESTEMALE C., Weighted Fuzzy Pattern Matching,

Fuzzy Sets and Systems 28, (1988) 313-331.

[DUB93] DUBOIS D., PRADE H., On possibility/probability transformations, *Fuzzy Logic*, (1993) 103-112.

[DUB90] DUBUISSON B., *Diagnostic et reconnaissance des formes*, Traité des Nouvelles Technologies, série Diagnostic et Maintenance, HERMES, 1990.

[FIS36] FISHER R., The use of multiple measurements in taxonomic problems, *Ann. Eugenics* 7, (1936) 179-188.

[FRE92] FRELICOT C., *Un système adaptatif de diagnostic prédictif par reconnaissance des formes floue*, Thèse présentée devant l'Université de Technologie de Compiègne, France, 1992.

[FRI96] FRIGUI H., KRISHNAPURAM R., A robust algorithm for automatic extraction of an unknown number of clusters from noisy data, *Pattern Recognition Letters* 17, (1996) 1223-1232.

[GAT89] GATH I., GEVA B., Unsupervised optimal fuzzy clustering, *IEEE Transactions on pattern analysis and machine intelligence* 11 (7), (1989).

[GOU95] GOUTIER J.M., Optimisation de la tête de détection d'un capteur à balance d'induction. Identification de la réponse d'un code métallique, Thèse de doctorat en Génie Automatique, Université de Reims, 1995.

[GUS79] GUSTAFSON D. E., KESSEL W. C., Fuzzy clustering with a fuzzy covariance matrix, *Proceeding of IEEE-CDC*, San Diego (1979).

[HE97] HE K., MEEDEN G., Selecting the number of bins in a histogram: A decision theoretic approach, *Journal of Statistical Planning and inference* (61), (1997) 1-14.

[ISH95] ISHIBUCHI H., NOZAKI K., TANAKA H., Selecting Fuzzy If-Then Rules for Classification Problems Using Genetic Algorithms, *IEEE Transactions on fuzzy systems* 3,

(1995).

[IZE91] IZENMAN A. J., Recent developments in nonparametric density estimation, *Journal of the American Statistical Association* 86 (413), (1991) 205-224.

[KLE98] KLEIN F., *Etude comparative de méthodes de classification floues*, Thèse présentée devant l'Université de Reims Champagne Ardenne, France, 1998.

[KLI92] KLIR G. J., Probability-possibility transformations: a comparison, *Int. J. General Systems* 21, (1992) 291-310.

[KLI93] KLIR G. J., Information-preserving probability-possibility transformations: recent developments, *Fuzzy Logic*, (1993) 417-428.

[KOH90] KOHONEN T., The self-organizing map, *Proceedings of the IEEE* 78 (9), (1990) 1464-1480.

[KRI93] KRISHNAPURAM R., KELLER J.M., A possibilistic approach to clustering, *IEEE transactions on fuzzy systems* 1(2), (1993) 98-109.

[LAS99] LASSERRE V., *Modélisation floues des incertitudes de mesures de capteurs*, Thèse présentée devant l'Université de Savoie, France, 1999.

[LEB95] LEBART L., MORINEAU A., PIRON M., *Statistique exploratoire multidimensionnelle*, Dunod, Paris, 1995.

[MAS96] MASSON M. H., DUBUISSON B., FRELICOT C., Conception d'un module de reconnaissance de formes floues pour le diagnostic, *RAIRO-APII-JESA* 30 (2,3), (1996) 319-341.

[MED98] MEDASANI S., JAESEOK K., KRISHNAPURAM R., An overview of membership function generation techniques for pattern recognition, *International Journal of Approximate Reasoning* 19, (1998) 391-417.

[MIL93] MILLGRAM M., Reconnaissance de formes, méthodes numériques et connexionnistes, Armand Colin, Dunod, Paris, 1993.

[MOU01] MOUCHAWEH S. M., BILLAUDEL P., LECOLIER V. G., Following the evolution of a system using Fuzzy Pattern Matchnig : *13th International Conference on Control Systems and Computer Science* (2001) 187-191, Bucharest, Romania.

[MOU02a] MOUCHAWEH S. M., DEVILLEZ A., LECOLIER V. G., BILLAUDEL P., Incremental learning in Fuzzy Pattern Matching, *Fuzzy Sets and Systems* 132 (1), (2002) 49-62.

[MOU02b] MOUCHAWEH S. M., BILLAUDEL P., LECOLIER V. G., Variable probability-possibility transformation, *Fuzzy Sets and Systems* (2002), à paraître.

[MOU02c] MOUCHAWEH S. M., BILLAUDEL P., LECOLIER V. G., Diagnostic prédictif et adaptatif pour la supervision en temps réel d'un système évolutif en utilisant Fuzzy Pattern Matching, *Rencontres Francophones de la Logique Floue et ses Applications* (2002), Montpellier, France.

[MOU02d] MOUCHAWEH S. M., DEVILLEZ A., BILLAUDEL P., Correlated Fuzzy Pattern Matching method : a classification method for classes of complex shape, *Proceedings of International conference on Circuits, Systems, Communications and Computers* (2002), WSEAS, Crete, Greece.

[MOU02e] MOUCHAWEH S. M., BILLAUDEL P., LECOLIER V. G., Elimination du bruit et détection des zones de haute densité, *la Revue d'Intelligence Artificielle* 16 (3), (2002) 279-306.

[MOU02f] MOUCHAWEH S. M., BILLAUDEL P., Influence of the choice of histogram parameters at Fuzzy Pattern Matching performance, *Int. Journal of WSEAS Transactions on System* 2 (1), (2002) 260-266.

[MOU02g] MOUCHAWEH S. M., BILLAUDEL P., Histogram with Adaptive Bin Width to increase class separability measure, *Int. Journal of Approximate Reasoning* (2002), à paraître.

[OTN72] OTNES R. K., ENOCHSON L., *Digital time series analysis*, Wiley-Interscience Publication, New York, 1972.

[OUS00] OUSSALAH M., On the probability/possibility transformations: a comparative analysis, *Int. J. General Systems* 29 (5), (2000) 671-718.

[PEL 93] PELTIER M. A., Un système adaptatif sur la reconnaissance des formes floues. Application au diagnostic du comportement d'un conducteur automobile, Thèse présentée devant l'Université de Technologie de Compiègne, France, 1993.

[RAU91] RAUDYS S. J., ANIL K. J., Small sample size effects in statistical pattern recognition: Recommendations for practitioners, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 3 (13), (1993).

[REI82] REILLY D. L., COOPER L. N., ELBAUM C., A neural model of category learning, *Biological Cybernetics* (45), (1982) 35-41.

[RUD82] RUDEMO M., Empirical choice of histograms and kernel density estimates, *Scandinavian Journal of Statistics* 9, (1982) 65-78.

[SAN00] SANCHO J. L., PIERSON W. E., ULUG B., FIGUEIRAS A. R., AHALT S. C., Class separability estimation and incremental learning using boundary methods, *Neurocomputing* 35, (2000) 3-26.

[SAN91] SANDRA S., *La combinaison de l'information incertaine et ses aspects algorithmiques*, Thèse présentée devant l'Université Paul Sabatier de Toulouse, France, 1991.

[SCO79] SCOTT D. W., On optimal and data-based histograms, *Biometrika* (66), (1979) 605-610.

[SCO92] SCOTT D. W., *Multivariate density estimation*, Wiley, New York, 1992.

[SHA76] SHAFER G., *A mathematical theory of evidence*, Princeton University Press, Princeton, N.J., 1976.

[SUG74] SUGENO M., *Theory of fuzzy integrals and its applications*, PhD Thesis, Tokyo

institute of Technology, 1974.

[TER85] TERREL G. R., SCOTT D. W., Oversmoothed non parametric density estimates, *Journal American Statistics Association* (80), (1985) 209-214.

[WAN96] WAND M. P., Data-based choice of histogram bin width, *Australian Graduate School of Management*, University of New South Wales, (1996).

[XIE91] XIE X. L., BENI G., A validity measure for fuzzy clustering, *IEEE Transactions on pattern analysis and machine intelligence* 13, (1989) 841-847.

[ZAD65] ZADEH L. A., Fuzzy sets, *Informations and control* 8, (1965) 338-353.

[ZAD78] ZADEH L. A., Fuzzy sets as a basis for a theory of possibility, *Fuzzy sets and systems* 1, (1978) 3-28.

[ZOU97] ZOUHAL L., *Contribution à l'application de la théorie des fonctions de croyance en reconnaissance des formes*, Thèse présentée devant l'Université de Technologie de Compiègne, France, 1997.