



HAL
open science

Diagnostic par techniques d'apprentissage floues: concept d'une méthode de validation et d'optimisation des partitions

Claudia Victoria Isaza Narvaez

► **To cite this version:**

Claudia Victoria Isaza Narvaez. Diagnostic par techniques d'apprentissage floues: concept d'une méthode de validation et d'optimisation des partitions. Automatique / Robotique. INSA de Toulouse, 2007. Français. NNT: . tel-00190884

HAL Id: tel-00190884

<https://theses.hal.science/tel-00190884>

Submitted on 23 Nov 2007

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

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

THÈSE

Préparée au

Laboratoire d'Analyse et d'Architecture des Systèmes du CNRS

En vue de l'obtention du titre de

*Docteur de l'Université de Toulouse, délivré par l'Institut National
des Sciences Appliquées de Toulouse*

Spécialité : Systèmes Automatiques

Par CLAUDIA VICTORIA ISAZA NARVAEZ

DIAGNOSTIC PAR TECHNIQUES D'APPRENTISSAGE FLOUES : CONCEPTION
D'UNE METHODE DE VALIDATION ET D'OPTIMISATION DES PARTITIONS

Soutenue le 25 d'Octobre 2007 devant le jury :

Président :	G. ROUX
Rapporteurs :	N. PERROT R. LOPEZ DE MANTARAS
Directeurs de thèse :	J. AGUILAR-MARTIN M. V. LE LANN
Examineurs :	J. WAISSMAN VILANOVA F. RIVAS
Invitée :	T. KEMPOWSKY

Table des matières

Introduction Générale.....	1
1. Surveillance et diagnostic de systèmes complexes.....	5
1.1 Introduction.....	5
1.2 Définition des systèmes complexes.....	7
1.3 La surveillance et le diagnostic dans le cadre de la supervision.....	8
1.3.1 Principes de la supervision.....	9
1.3.2 Théorie générale du diagnostic et de la surveillance.....	10
1.4 Techniques employées pour le diagnostic automatique.....	12
1.4.1 Méthodes basées sur des modèles.....	14
1.4.1.1 Méthodes basées sur des modèles quantitatifs.....	14
1.4.1.2 Méthodes basées sur des modèles qualitatifs.....	15
1.4.2 Méthodes basées sur des données historiques (fouille de données).....	15
1.5 Diagnostic en utilisant des méthodes de classification.....	17
1.5.1 Phase d'entraînement ou apprentissage.....	19
1.5.2 Phase de reconnaissance.....	20
1.5.3 Méthodes d'apprentissage et de reconnaissance de formes utiles pour le diagnostic.....	20
1.6 Conclusion.....	23
2. Comparaison de méthodes de classification.....	25
2.1 Introduction.....	25
2.2 Contexte général des méthodes de classification basées sur la fouille de données	26
2.3 Description des différents Méthodes.....	27
2.3.1 Méthodes avec Apprentissage Supervisé.....	27

2.3.1.1 Discriminant Linéaire – ACP	27
2.3.1.2 Les K plus proches Voisins (<i>K- Nearest Neighbors</i>)	27
2.3.1.3 Réseaux de Neurones	27
2.3.1.4 Arbres de Décision – CART	29
2.3.2 Méthodes avec Auto-Apprentissage	30
2.3.2.1 Méthodes de coalescence (<i>clustering</i>).....	30
2.3.3 LAMDA (<i>Learning Algorithm for Multivariate Data Analysis</i>)	32
2.3.3.1 Degré d'Appartenance Marginale - Fonctions MADs.....	34
2.3.3.2 Paramètres d'apprentissage.....	36
2.3.3.3 Degré d'Appartenance Globale – Fonction d'Agrégation Floue GAD	38
2.3.3.4 Similitude entre LAMDA et le Réseau de Neurones de type Perceptron	40
2.4 Comparaison des méthodes.....	41
2.4.1 Exemples	41
2.4.1.1 Critère d'évaluation	42
2.4.1.2 Résultats	43
2.4.2 Étude particulière des données qualitatives sur un exemple.	45
2.4.2.1 Résultats	47
2.5 Conclusion.....	50
3. Critères d'évaluation des partitions de l'espace de données	53
3.1 Introduction.....	53
3.2 Critères pour évaluer une partition de l'espace des données	54
3.3 Mesures pour évaluer les critères de qualité des partitions.	55
3.3.1 Mesure de distance entre classes	58
3.3.2 Mesure de dispersion des classes.....	62
3.4 Indices d'évaluation de la qualité d'une partition.....	65
3.4.1.1 Indice de dissemblance entre partitions.....	66
3.4.1.2 Indices de qualité des partitions floues	68
3.4.1.3 Indice de qualité des partitions floues basé uniquement sur les degrés d'appartenance floue.....	70
3.5 Méthode d'optimisation des partitions	71
3.6 Conclusion.....	73
4. Applications de la méthode d'évaluation et d'optimisation des partitions floues	75
4.1 Introduction.....	75
4.2 Cas illustratifs avec données artificielles	78

4.2.1 Cas d'étude dans l'espace 2D.....	78
4.2.2 Cas d'étude dans l'espace 3D.....	85
4.3 Processus de production de propylène glycol.	87
4.3.1 Description du procédé.	87
4.3.2 Application de la méthode d'optimisation avec tous les descripteurs.....	89
4.3.3 Application de la méthode d'optimisation avec les descripteurs sélectionnés. ...	95
4.4 Réaction d'Estérification utilisant la technologie "Open Plate Reactor"	100
4.5 Réaction du thiosulfate de sodium utilisant la technologie "Open Plate Reactor" .	105
4.6 Conclusion	110
5. Conclusions et perspectives	111
Bibliographie	117
Annexes.....	123

Liste des figures

Figure 1.1. Schéma général des tâches d'un système de supervision.	9
Figure 1.2. Alarmes Traditionnelles [BOI03]	12
Figure 1.3. Le problème d'utiliser des diagrammes de diagnostic indépendants [KOU02]	13
Figure 1.4. Organisation des Méthodes de Diagnostic [VEN03]	14
Figure 1.5. Schéma de la supervision utilisant des techniques de classification	18
Figure 1.6. Schéma du diagnostic utilisant des techniques de classification	19
Figure 2.1. Structure du réseau de type RBF pour le diagnostic	28
Figure 2.2. Schéma de la méthode de classification LAMDA	33
Figure 2.3. Algorithme d'apprentissage de LAMDA	34
Figure 2.4. Partition de l'espace des données selon le type de fonction MAD et GAD	39
Figure 2.5. Résultats obtenus avec les différentes méthodes	44
Figure 2.6. Nombre Moyen d'Itérations	44
Figure 2.7. Profil des Classes obtenue avec LAMDA (Cas Qualitatif)	49
Figure 2.8. Profil des Classes obtenu avec LAMDA (Cas Quantitatif)	49
Figure 3.1. Représentation des classes comme ensembles flous	56
Figure 3.2. Représentation de l'actualisation de la matrice de partition floue	72
Figure 4.1. Schéma général de la méthode proposée par Orantes A. [ORA07a]	77
Figure 4.2. Ensemble de données 1	79
Figure 4.3. Ensemble de données 2	79
Figure 4.4. Classification Initiale Ensemble de Données 1	79
Figure 4.5. Classification Initiale Ensemble de Données 2	79
Figure 4.6. Evolution des partitions ensemble de données 2, Algorithme d'évaluation et d'optimisation de partitions floues	80
Figure 4.7. Classification Optimale Ensemble de Données 1	81
Figure 4.8. Classification Optimale Ensemble de Données 2	81
Figure 4.9. Comparaison des indices CV et Icc, ensemble de données 1	82
Figure 4.10. Comparaison des indices CV et Icc, ensemble de données 2	82
Figure 4.11. Comparaison de la dispersion indices CV et Icc, ensemble de données 1	83
Figure 4.12. Comparaison de la dispersion indices CV et Icc, ensemble de données 2	83

Figure 4.13. Comparaison distance minimale pour indices CV et Icc, ensemble de données 1	84
Figure 4.14. Comparaison distance minimale pour indices CV et Icc, ensemble de données 2	84
Figure 4.15. Evolution des partitions ensemble de données 3, Algorithme d'évaluation et d'optimisation de partitions floues	85
Figure 4.16. Comparaison des indices CV et Icc, ensemble de données 3	86
Figure 4.17. Comparaison de la dispersion indices CV et Icc, ensemble de données 3	86
Figure 4.18. Comparaison de la distance minimale indices CV et Icc, ensemble de données 3	87
Figure 4.19. Schéma de l'unité de production de propylène glycol [ORA05]	88
Figure 4.20. Classification initiale -21 descripteurs, Procédé de production du propylène glycol	91
Figure 4.21. Indice de qualité des partitions CV -21 descripteurs, Procédé de production du propylène glycol	91
Figure 4.22. Classification optimale -21 descripteurs, Procédé de production du propylène glycol	92
Figure 4.23. Classification proposée par l'expert -21 descripteurs, Procédé de production du propylène glycol	92
Figure 4.24. Classification de référence -21 descripteurs, Procédé de production du propylène glycol	93
Figure 4.25. Comparaison des classifications optimale et de référence -21 descripteurs, Procédé de production du propylène glycol	95
Figure 4.26. Classification initiale -7 descripteurs, Procédé de production de propylène glycol	96
Figure 4.27. Indice de qualité des partitions CV -7 descripteurs, Procédé de production de propylène glycol	97
Figure 4.28. Classification optimale -7 descripteurs, Procédé de production de propylène glycol	97
Figure 4.29. Classification de référence -7 descripteurs, Procédé de production du propylène glycol	98
Figure 4.30. Comparaison des classifications optimale et de référence -7 descripteurs, Procédé de production du propylène glycol	99
Figure 4.31. L' "Open plate reactor"	100
Figure 4.32. Classification initiale -8 descripteurs, Réaction d'estérification	103
Figure 4.33. Indice de qualité des partitions CV -8 descripteurs, Réaction d'estérification	103
Figure 4.34. Classification optimale -8 descripteurs, Réaction d'estérification	104
Figure 4.35. Classification de référence -8 descripteurs, Réaction d'estérification	104
Figure 4.36. Comparaison des classifications optimale et de référence -8 descripteurs, Réaction d'estérification	105
Figure 4.37. Classification initiale -9 descripteurs, Réaction du Thiosulfate de Sodium	107
Figure 4.38. Indice de qualité des partitions CV -9 descripteurs, Réaction du Thiosulfate de Sodium	107

Figure 4.39. Classification optimale -9 descripteurs, Réaction du Thiosulfate de Sodium	108
Figure 4.40. Classification de référence -9 descripteurs, Réaction du Thiosulfate de Sodium	108
Figure 4.41. Comparaison des classifications optimale et de référence -9 descripteurs, Réaction du Thiosulfate de Sodium.	109

Remerciements

Ce mémoire constitue la conclusion de mes travaux de thèse. Ma thèse a été réalisée au Laboratoire d'Analyse et d'Architecture de Systèmes du Centre National de la Recherche Scientifique (LAAS-CNRS) au sein du groupe Diagnostic, Supervision et Conduite qualitatifs (DISCO). Je remercie les directeurs successifs du LAAS, Messieurs Malik Ghallab et Raja Chatila ainsi que Madame Luise Travé-Massuyès, responsable du groupe DISCO, pour m'avoir accueillie pendant ces trois années de thèse.

Je veux adresser de manière spéciale ma gratitude à mes directeurs de thèse Madame Marie Veronique Le Lann et Monsieur Joseph Aguilar-Martin, qui m'ont guidé avec une disponibilité et amitié immense pendant les travaux de ma thèse, et qui m'ont apporté non seulement dans le cadre scientifique mais aussi dans le cadre personnel. Joseph s'est toujours montré prêt à discuter avec moi sur les travaux de la thèse et sur de nombreux sujets de la vie courante. Marie Véronique a été à tout instant surveillante du bon développement de ma thèse dans le cadre scientifique et personnel.

Je suis très reconnaissante de l'honneur que m'ont fait Madame Nathalie Perrot, chercheur à l'INRA-Thiverval Grignon, et Monsieur Ramon Lopez de Mantaras, Professeur à l'Université Autonoma de Barcelona-Espagne, en acceptant l'ardue tâche d'évaluer en qualité de rapporteurs les travaux présentés. J'exprime également ma reconnaissance à Messieurs Gilles Roux, Professeur à l'Université Paul Sabatier de Toulouse, Francklin Rivas, Professeur à l'Université des Andes, Merida-Venezuela et Julio Weissman, enseignant-chercheur à l'Université Autonoma de Hidalgo-Mexique, d'avoir accepté de prendre part au jury. Je remercie également Madame Tatiana Kempowsky, Docteur de l'INSA-Toulouse, d'avoir été disponible pendant ces années pour discuter les résultats de ma recherche et enfin pour avoir accepté de prendre part au jury. Je les remercie tous pour l'intérêt et les commentaires très constructifs qu'ils ont fait à ma thèse.

Je tiens à remercier les personnes qui font partie du Programme de Coopération Post Graduate-PCP entre la France et le Venezuela. Le fait d'avoir participé au PCP m'a permis de partager et de compléter mes travaux avec l'Université de Mérida-Venezuela. Dans le cadre de ce programme j'ai eu la possibilité de travailler avec Monsieur Jose Aguilar-Castro qui a apporté des idées très intéressantes pour le développement de la thèse. Je voudrais aussi remercier particulièrement la personne en France en charge de l'organisation des stages, Madame Martine Arroyo, qui réalise son travail de manière très aimable et efficace.

Ce document de thèse pour moi constitue l'atteinte d'un objectif que je m'étais fixé quelques années avant d'arriver au LAAS. Je voudrais adresser mes remerciements aux deux personnes qui m'ont guidé en Colombie dans mes premiers étapes de la recherche, Monsieur Victor Hugo Grisales qui m'a montré à travers son exemple que réaliser une thèse est quelque chose de faisable et à Monsieur Alain Gauthier qui a cru en moi pour développer la recherche pendant mes années du master à l'Université des Andes Bogotá-Colombie et grâce à qui j'ai connu le LAAS.

Finally, there is a large number of people who despite the distance have brought me a very strong emotional support. First, my family: my mother, my father and my sister who throughout my life have been by my side to push me to achieve my academic and personal goals. I also thank my friend Manuel and my two big friends Paola and Gabriel who have shared with me all my experiences during these years.

During my life in Toulouse, I have met very special people that I will keep in my heart: Gustavo who has always been ready to listen to me even when he is very busy, Tatiana who is an expert in administrative procedures for Colombians in Toulouse and who has helped me a lot during my installation in Toulouse, her help has been invaluable. Corinne who has encouraged me regarding the language, her encouragement has been very useful in particular for the end of the thesis.

Finally, I would like to say that with the finalization of my thesis a new stage of my life begins for which each of the people who have been close to me have brought me, thank you to all.

INTRODUCTION GENERALE

Dans le cas de processus complexes (e.g. chimiques et biochimiques) qui présentent des dynamiques hautement non linéaires et changeantes au cours du temps, il est parfois difficile d'obtenir un modèle mathématique simple et précis qui permette de détecter des états anormaux. Ces procédés se caractérisent généralement par le fait qu'ils possèdent un grand nombre de variables, qui doivent être prises en considération au moment d'effectuer la supervision. Dans ce cas, la connaissance sur le système doit être extraite à partir des signaux issus de la surveillance en ligne du système, tout en y incluant des mécanismes d'apprentissage ainsi que la connaissance de l'expert.

Les méthodes d'apprentissage et de reconnaissance de formes permettent à partir des données du processus, d'obtenir l'information sur les états du système utiles pour la surveillance. Parmi les techniques d'apprentissage, celles incluant de la logique floue ont l'avantage d'exprimer les appartenances d'un individu à plusieurs classes, ce qui permet de mieux connaître la situation réelle du système et de prévoir des changements vers des états de défaillance. Toutefois leurs performances sont fortement liées à la bonne sélection des paramètres de l'algorithme et au nombre de classes lorsque celui-ci est nécessaire (e.g. algorithmes de coalescence).

Cette thèse a pour objectifs l'amélioration de ces techniques, tout d'abord en diminuant la dépendance des algorithmes d'apprentissage aux paramètres de synthèse propres à chacun d'eux et le développement d'une procédure quasi-automatique permettant de déterminer la meilleure classification qui caractérise les états fonctionnels, ainsi que le nombre de classes lorsque celui-ci doit être connu *a priori*. Nous présentons une méthode qui donne aux experts du procédé une orientation pour établir les états fonctionnels lors de l'implantation d'une technique de surveillance.

Notre objectif est l'élaboration de cette méthode permettant de valider et d'adapter automatiquement la partition de l'espace de données obtenue par une technique de classification floue. Elle permet de trouver automatiquement une partition optimale en

termes de compacité et de séparation des classes, à partir uniquement de la matrice d'appartenance obtenue par une classification antérieure. Cette méthode est donc une aide importante donnée à l'expert du processus pour établir les états fonctionnels dans l'implémentation d'une technique de surveillance d'un procédé complexe.

Le document de thèse est structuré en quatre chapitres :

Le premier chapitre présente les concepts généraux pour la supervision et le diagnostic. Nous présentons un tour d'horizon de différentes techniques de diagnostic. L'objectif de ce chapitre est de mettre en contexte l'utilisation de méthodes de classification pour le diagnostic de systèmes complexes.

Le deuxième chapitre précise le contexte général des méthodes de classification basées sur la fouille de données. Nous présentons une comparaison entre différentes techniques de classification pour mettre en lumière leurs avantages et limitations. Nous mettons en évidence que ces techniques sont adéquates pour faire le diagnostic automatique des systèmes, quand il est considéré comme la différenciation d'états en prenant en compte les états de défaillance; néanmoins il existe une forte dépendance aux paramètres de synthèse de chaque algorithme. D'ailleurs, en général le choix du nombre de classes est critique dans les cas où on n'a pas d'idée *a priori* de la distribution de l'espace des données. Toutefois, avec des paramètres adéquats, bien que non optimaux, il est possible d'avoir une classification initiale avec plusieurs classes qui peut être améliorée ensuite. En conséquence, nous considérons particulièrement intéressant de proposer une technique applicable aux méthodes de classification de type flou en général, qui permette d'améliorer automatiquement la partition initiale obtenue.

Dans le troisième chapitre, nous analysons les indices de qualité des partitions existants dans la littérature pour évaluer une partition en termes de compacité et de séparation des classes. Nous proposons des modifications pour que leurs calculs ne dépendent plus des valeurs des données, mais uniquement de la matrice des degrés d'appartenance aux différentes classes. Finalement, nous proposons la méthode de validation et d'optimisation automatique des partitions. Pour pouvoir implanter cette amélioration dans les différents algorithmes de classification floue, les mesures proposées sont basées uniquement sur le résultat commun délivré par ces méthodes de classification, à savoir la matrice de partition obtenue. Cette proposition permet d'obtenir un meilleur entraînement pour un système de surveillance basé sur la fouille de données quand le nombre des classes est inconnu.

L'analyse de la qualité des partitions, en fonction de la compacité et de la séparation des classes, et le regroupement des classes semblables sont la base de la méthode que nous proposons. Cette méthode constitue une aide pour l'entraînement des méthodes de reconnaissance de formes en diminuant la dépendance aux paramètres à choisir par l'utilisateur et en donnant automatiquement, à partir d'une partition correcte bien que non optimale, le nombre de classes. La méthode est

applicable à tous les algorithmes de classification floue, n'ajoute pas de paramètres à choisir par l'utilisateur et elle ne nécessite pas de relancer l'algorithme d'apprentissage.

Nous terminons le document avec un chapitre dédié d'abord aux résultats obtenus lors de l'application de la méthode proposée sur un cas d'étude académique dont la structure est typiquement employée dans le contexte de l'évaluation de la qualité des partitions. Dans une deuxième partie nous présentons les résultats de l'application de la méthode proposée à 3 cas issus du domaine industriel (chimique et pharmaceutique):

-Procédé de production du propylène glycol.

-Nouveau type de réacteur chimique développé par le Laboratoire de Génie Chimique de Toulouse en partenariat avec ALFA-LAVAL qui utilise la technologie « *Open Plate Reactor* »[DEV03]. Il s'agit d'un nouveau concept de réacteurs avec échange thermique amélioré. La mise en œuvre de deux réactions chimiques a été étudiée : une réaction d'estérification et la réaction du thiosulfate de sodium sur l'eau oxygénée.

A travers des exemples issus du domaine industriel nous montrons comment la méthode proposée peut être une aide très utile dans la conception d'un système de surveillance basé sur des méthodes de classification floue non supervisées.

Chapitre 1

1. SURVEILLANCE ET DIAGNOSTIC DE SYSTEMES COMPLEXES

1.1 Introduction

Dans le cas des processus industriels, une grande quantité des pannes et une diminution de la production sont en général le résultat d'une défaillance dans le procédé qui n'a pas été corrigée à temps. Actuellement, il y a plusieurs intérêts, pas seulement économiques, à avoir un bon fonctionnement de ces processus. Si une défaillance est détectée au tout début de son apparition ou même anticipée, de grands dommages voire même des catastrophes peuvent être évitées. La détection et le diagnostic des défaillances pour ces processus industriels sont donc d'un intérêt capital.

Aujourd'hui, le concept d'automatisation des processus inclue plusieurs aspects globaux. L'automatisation de nos jours ne s'attache pas seulement au maintien de quelques variables à des valeurs de référence, elle s'intéresse au procédé dans sa globalité et dans ses divers modes (de fonctionnement, de démarrage, d'arrêt, de fonctionnement normal et de défaillance). L'automatisation s'effectue grâce à un système numérique de contrôle commande et peut avoir des objectifs divers. Une première catégorie d'objectifs concerne les performances du système de production (garantir la qualité du produit et diminuer les coûts de production), une deuxième catégorie qui prend de plus en plus d'importance, est la sûreté du fonctionnement. Notre société n'est plus prête à accepter des accidents industriels faisant des victimes humaines ou des dégâts considérables sur l'environnement [GEN07].

L'inclusion de tâches pour prévenir des accidents est donc devenu un autre des objectifs du concept global de l'automatisation des processus, et de manière plus spécifique de la surveillance et du diagnostic de systèmes afin d'empêcher de tomber dans des états dangereux. Savoir détecter un mode de fonctionnement anormal

suffisamment tôt peut permettre de produire une commande susceptible de revenir à un mode de fonctionnement plus adapté à l'objectif pour lequel ce système a été conçu ; ce qui est d'un grand intérêt pour l'industrie. Le diagnostic automatique est donc maintenant un élément essentiel d'un système de production ou d'un système conçu pour être utilisé par un tiers [DUB01].

L'automatisation a donc évolué depuis la régulation et l'optimisation vers la garantie de la sécurité, fondée sur la surveillance, le diagnostic en ligne et la maintenance préventive. Ces tâches font désormais partie intégrante de la supervision d'une installation industrielle.

Ces constatations mettent en évidence le besoin d'inclure un système de surveillance et de diagnostic global des processus pour vérifier le bon fonctionnement et prévoir des arrêts et éviter des accidents industriels. Ce type de diagnostic qui doit inclure des concepts globaux du procédé, doit être basé sur la connaissance la plus précise possible du système.

Les processus complexes (chimiques et biochimiques par exemple) ont des dynamiques hautement non linéaires et changeantes au cours du temps. Ils se caractérisent par le fait de posséder un grand nombre de variables, qui doivent être prises en considération au moment d'effectuer la supervision. Pour ces systèmes l'incertitude et l'imprédictibilité rendent les décisions difficiles. L'automatisation apparaît comme le plus important facteur de développement de ces processus.

D'un autre côté, dans un système complexe, il est cependant impossible de tout anticiper et planifier. C'est pourquoi des opérateurs humains sont encore souvent en charge des grandes installations industrielles. Les humains ont en effet pour principale qualité de savoir gérer l'imprévu et l'incertain, et donc de pouvoir créer des stratégies adaptées quand surviennent des situations de dysfonctionnement.

Il est donc clair que pour les processus complexes, il est essentiel de disposer de systèmes de surveillance et de diagnostic qui permettent de travailler avec un niveau d'incertitude élevé et d'inclure la connaissance humaine. Pour les processus complexes, il est très compliqué ou impossible d'obtenir un modèle mathématique du procédé. La connaissance sur le comportement du procédé peut être obtenue alors des mesures récupérées pendant des situations normales et anormales en utilisant des mécanismes d'apprentissage et en incluant la connaissance de l'expert du procédé. Cette connaissance peut être organisée comme une partition en classes de l'ensemble des données historiques représentant les états fonctionnels du procédé.

Les méthodes de reconnaissance de formes permettent d'extraire, à partir de données et des historiques du processus, l'information qui caractérise les états normaux et de défaillances du système, en identifiant les classes de fonctionnement que peut présenter le processus. L'utilisation de techniques d'intelligence artificielle et de la logique floue pour le diagnostic de systèmes complexes permet d'obtenir des

résultats interprétables et offre une information utile pour la décision d'actions à faire en cas de défaillance [HUL05].

En prenant en compte la forte relation entre l'homme et la machine qui doit exister dans les systèmes de supervision, nous considérons comme important de proposer des techniques ou d'améliorer celles existantes afin de faciliter l'implantation des automatismes et l'interprétation des résultats pour une prise de décisions plus aisée. Dans cette thèse, nous proposons une méthode qui se veut être une aide utile à l'expert du procédé pour bien identifier les états du système en utilisant des techniques basées sur l'analyse de données.

Au cours de ce chapitre, nous présentons les concepts généraux de la surveillance et du diagnostic automatique de systèmes. Nous présentons un tour d'horizon de différentes techniques de diagnostic. L'objectif de ce chapitre est de mettre en contexte l'utilisation de méthodes de classification pour le diagnostic de systèmes complexes.

1.2 Définition des systèmes complexes

Les processus complexes se caractérisent comme étant des processus en environnement ouvert, où l'incertitude, l'imprédictibilité et la dynamique des phénomènes rendent les décisions difficiles [GEN07]. Dans l'industrie, on trouve des systèmes complexes de manière courante, dans les systèmes embarqués, les systèmes de transport (avion, train), les procédés chimiques ou biochimiques.

Les processus chimiques généralement ont des comportements non linéaires, avec des variables fortement couplées, des mesures bruitées et des paramètres non connus [AKB01].

Dans un système complexe, il y a divers modes (nous les appellerons états fonctionnels) incluant les défaillances. Notons qu'un certain nombre de procédés fonctionnent en mode discontinu ou semi continu et comportent donc *a priori* plusieurs modes de fonctionnement normal, ce type de procédés est parfois qualifié de procédés ou systèmes hybrides [GEN07].

Pour les systèmes complexes, les modes de défaillance sont spécialement difficiles à prévoir. Ces modes particuliers peuvent nécessiter soit l'arrêt immédiat du procédé, soit de basculer la commande vers un mode dégradé qui consiste à modifier les consignes des boucles locales ou même carrément l'architecture de la commande ou du procédé. Donc pour ces systèmes, il est tout particulièrement intéressant, de pouvoir détecter la tendance qui les fait dériver vers un des états anormaux pour prendre les décisions pertinentes et éviter l'arrêt du procédé ou une réparation coûteuse.

Etant donné que la connaissance de la dynamique du système est importante pour interpréter convenablement les situations, et que ces systèmes ont des dynamiques très complexes, la localisation des défaillances est spécialement difficile en utilisant des techniques classiques.

Les caractéristiques des systèmes complexes compliquent leur modélisation avec des techniques où le procédé est représenté par un ensemble d'équations. Néanmoins, l'utilisation de techniques issues du domaine de l'intelligence artificielle apparaît comme la principale alternative pour aborder ces problèmes lorsqu'il est nécessaire de prendre en compte l'intervention des experts du domaine ou de traiter de l'information de nature qualitative. Dans ce cadre, les techniques qui permettent de caractériser les états en utilisant l'information extraite des données sont envisageables comme une aide très précieuse pour le diagnostic des systèmes.

1.3 La surveillance et le diagnostic dans le cadre de la supervision

De manière générale, la supervision correspond à l'action de surveiller un système, afin de prendre des actions nécessaires si le système est hors de l'objectif de commande. De manière simple : la surveillance consiste à détecter le comportement présent du système en différenciant entre plusieurs états (normal et défectueux) du processus et le diagnostic est l'identification de la nature d'un dysfonctionnement, d'une difficulté.

On peut considérer aussi le diagnostic comme l'identification des causes des divers modes des processus, on inclut non seulement les modes de fonctionnement normal mais aussi les modes de défaillances, qui correspondent aux différents états que peut prendre un procédé affecté par une panne ou un dysfonctionnement, états qui sont beaucoup plus difficiles à prévoir [GEN07].

Dans ce contexte, il est important de définir quelques concepts associés à la supervision des processus [HER06] :

- *Fonctionnement normal d'un système.* Un système est dit dans un état de fonctionnement normal lorsque les variables le caractérisant demeurent au voisinage de leurs valeurs nominales. Le système est dit défaillant dans le cas contraire.

- Une *défaillance* est la cause d'une anomalie.

- Une *dégradation* d'un procédé caractérise le processus qui amène à un état défaillant du procédé.

- Un *défaut* se définit comme une anomalie du comportement d'un système sans forcément remettre en cause sa fonction.

- Une *panne* caractérise l'inaptitude d'un dispositif à accomplir une fonction requise. Un système est toutefois généralement considéré en panne dès l'apparition d'une défaillance.

- Un *symptôme* est l'événement ou l'ensemble de données au travers duquel le système de détection identifie le passage du procédé dans un fonctionnement anormal. C'est le seul élément dont a connaissance le système de surveillance au moment de la détection d'une anomalie.

Le défaut est constaté grâce à un indicateur de défaut, il révèle la présence d'une défaillance. L'indicateur de défaut est en pratique élaboré grâce aux variables mesurées de l'installation [GEN07]. Cet indicateur permet de détecter l'état présent du système et dans le cas de défaillance permet d'estimer les causes du problème.

1.3.1 Principes de la supervision

Le rôle de la supervision est de contrôler l'effet de la commande sur le procédé. Il s'agit alors de vérifier (surveiller) que le procédé est en fonctionnement normal. S'il ne l'est pas, il agit au moins pour éviter des dommages sur l'installation, pour limiter les arrêts intempestifs et au mieux pour garantir la continuité de la production.

Un schéma de supervision inclut les tâches pour détecter l'état du système (surveillance), estimer les causes (diagnostic) et les tâches de corrections si le système s'écarte du point de fonctionnement (maintenance) [GEN07].

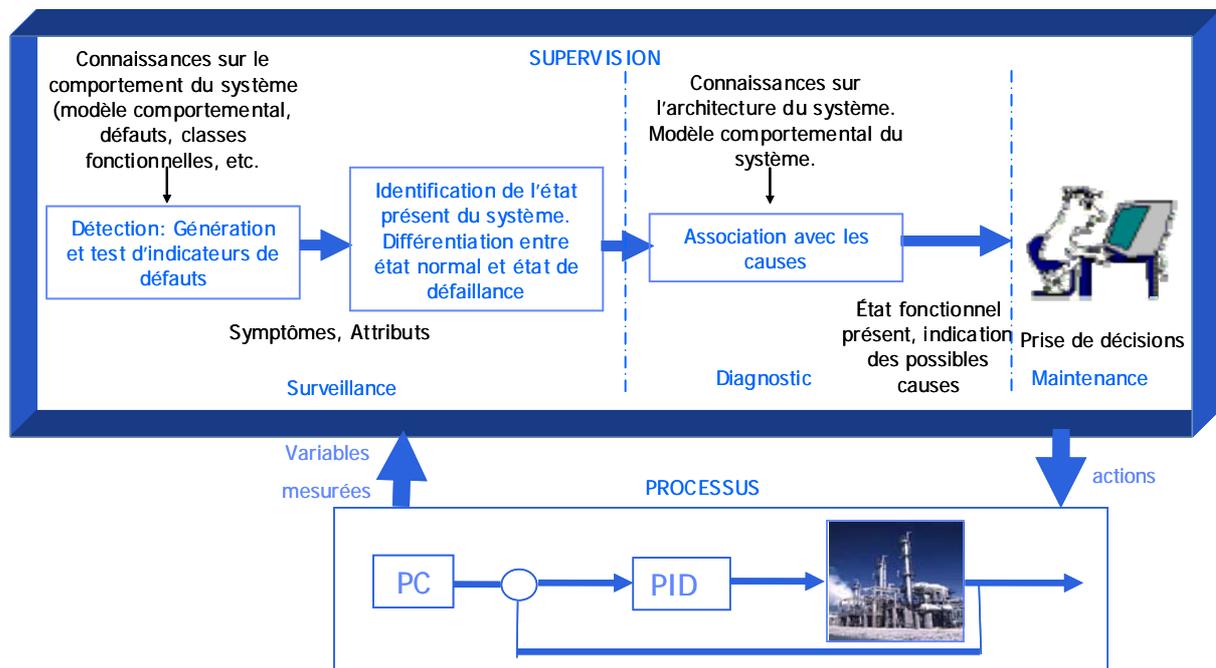


Figure 1.1. Schéma général des tâches d'un système de supervision.

En effet, la supervision des systèmes correspond à établir l'état présent du système pour pouvoir prendre des décisions et pouvoir maintenir des critères de performance (qualité, rentabilité, sûreté, etc.) du système. On peut dire qu'elle consiste aussi à vérifier que le système ne s'écarte pas du point de fonctionnement choisi. La

surveillance (ou monitoring) se réfère à la capacité de reconnaître un comportement anormal et de le signaler.

Les méthodes avancées de surveillance et de diagnostic de défauts sont nécessaires, pour répondre aux exigences issues de la supervision comme : l'anticipation de la détection de défauts, le diagnostic de pannes d'actionneurs ou de capteurs, de défauts de comportement du procédé, la détection de défauts dans les boucles fermées et la supervision de procédés lors d'états transitoires. L'objectif de l'anticipation de la détection de défauts et du diagnostic est d'avoir assez de temps pour traiter des actions, comme la reconfiguration du processus ou la maintenance.

Pour les systèmes complexes, la surveillance doit être établie en tenant compte de l'information reçue des différents variables. La supervision classique basée sur une variable donnée peut conduire à des résultats incorrects dans plusieurs cas [AKB01]. Dans la section consacrée aux méthodes de diagnostic, nous montrerons l'importance de prendre en compte les différentes variables ainsi que les relations qui les lient.

1.3.2 Théorie générale du diagnostic et de la surveillance

L'objectif de la surveillance est de vérifier l'état courant du procédé en incluant la détection de l'apparition d'un défaut et le diagnostic permet d'en trouver la cause puis d'en déduire la marche à suivre afin d'assurer la sûreté de fonctionnement du procédé. En généralisant, on peut dire que le diagnostic consiste à identifier les états du système en utilisant des indicateurs qui sont élaborés grâce aux variables du procédé. Dans la plupart des cas, le système de diagnostic cherche à identifier surtout les états à éviter.

Dans la littérature, deux étapes très importantes pour la supervision sont généralement distinguées [GEN07]:

- La détection, qui consiste à utiliser un jeu de mesures (indicateurs de défauts) qui permettent d'engendrer des symptômes. On parle aussi d'attributs quand on fait référence à l'utilisation de méthodes de classification pour la détection.
- La localisation ou la classification, qui consiste à remonter d'un ensemble de symptômes à un ensemble minimal de composants matériels dont le dysfonctionnement est suffisant pour expliquer l'apparition des symptômes observés.

Il est évidemment souhaitable que cet ensemble de composants suspectés soit le plus réduit possible et contienne avec certitude le(s) composant(s) réellement défaillant(s) pour que le diagnostic soit précis et fiable. Des travaux sur ce sujet ont été abordés dans le groupe DISCO du LAAS [ORA05].

Une difficulté importante pour trouver la cause des défaillances, concerne le lien qu'il faut établir entre le modèle comportemental d'un procédé et son modèle structurel (connaissance de l'ensemble des composants qui le constitue et de leurs interconnexions). Communément est appelé support d'un indicateur, l'ensemble des

composants dont le modèle est utilisé pour construire cet indicateur. Dans la section 1.5, nous montrerons comment en utilisant les techniques de classification, l'association entre l'indicateur de défaillance (paramètres des classes) et le support de l'indicateur est plus directe que dans le cas d'autres méthodes.

Une des parties les plus importantes de la boucle de supervision est la partie diagnostic. Cette tâche permet d'identifier des causes de défaillance pour que l'expert puisse prendre une décision. Pour le diagnostic il est donc très important d'inclure l'expertise de l'expert du système, et les techniques employées doivent permettre une « communication » homme-machine efficace.

Si le procédé est supervisé par un opérateur humain, on peut comprendre aisément qu'il lui est difficile de surveiller en même temps un grand nombre de variables. Une technique de diagnostic est alors nécessaire pour faire l'analyse des variables pertinentes et donner une information plus claire, concise et utile pour que l'opérateur puisse prendre des décisions.

Comme on peut le constater la surveillance a besoin fortement de l'interaction homme-machine, c'est pourquoi il est important de développer et d'améliorer les techniques existantes pour faciliter cette relation et enrichir les tâches automatiques avec l'expertise humaine sur le procédé.

Un système de supervision avancé devrait intégrer toute sorte d'aides pour diminuer la charge de travail et la fréquence des erreurs humaines, sans toutefois viser à éliminer l'opérateur de la fonction de décision. Il y a une difficulté particulière des tâches liées à la détection des conditions anormales, l'identification de la situation (état) (diagnostic) en cours et l'anticipation de son évolution, il est donc important de développer et d'améliorer les outils pour ce travail en se basant sur une supervision fondée sur une coopération homme/machine où l'homme se voit assisté non seulement dans ses tâches routinières mais aussi dans ses tâches de raisonnement [GEN07].

Vu la difficulté existante de bien interpréter les alarmes à partir de l'analyse de variables individuelles, des outils permettant de relier des défauts aux différentes variables sont extrêmement utiles pour le diagnostic et la surveillance.

Nos travaux se situent dans le domaine de l'identification automatique des états (défaillants ou non), tout en donnant une aide à l'expert du procédé pour identifier les états et les inclure dans un système de surveillance et de diagnostic en vue d'effectuer une reconnaissance automatique des états en ligne. Nous recherchons des algorithmes qui permettent une interprétation facile des résultats pour effectuer, de manière plus directe l'identification des causes de l'état identifié automatiquement.

Pour accroître l'application des systèmes de surveillance, il est nécessaire d'introduire des techniques facilement utilisables par les experts des processus. Il y a un besoin de techniques qui soient facilement interprétables dans le contexte même du procédé où l'expert a une vaste connaissance. Nous prétendons en effet donner une

aide à l'expert du procédé, qui n'est généralement pas un expert des techniques de surveillance, afin qu'il trouve une caractérisation adéquate des états, quand l'identification est basée sur des méthodes classification floues.

Un autre aspect important pour le développement des techniques de diagnostic est d'éviter les fausses alarmes. Dans la construction d'un système utile pour le diagnostic, il faut tenir compte de cet aspect. Des études pour enlever les fausses alarmes ont été implantées comme améliorations aux techniques de diagnostic [ISA07] mais ne seront pas expliquées de manière détaillées dans ce document parce qu'elles ne sont pas considérées comme une partie intégrante de la thèse.

1.4 Techniques employées pour le diagnostic automatique

Il y a plusieurs méthodes de diagnostic connues et de nouvelles techniques sont proposées constamment dans la littérature. La plupart des nouvelles techniques peuvent être vues comme des variantes ou des améliorations des précédentes. Par conséquent, il est important de faire un schéma des tendances générales de ces méthodes de diagnostic et de situer les méthodes de classification basées sur la fouille de données dans ce contexte.

Auparavant dans un système de supervision traditionnel, la tendance a été de surveiller simplement chaque signal pour vérifier que les variables restent à l'intérieur d'un domaine de valeurs caractéristiques du fonctionnement normal. Un seuil haut SH et un seuil bas SB sont fixés pour chaque variable et l'on vérifie qu'à chaque acquisition la variable appartient à l'intervalle [SB,SH]. Plusieurs seuils peuvent être établis pour générer des pré-alarmes au fur et à mesure que la variable s'écarte des limites établies (voir Figure 1.2).

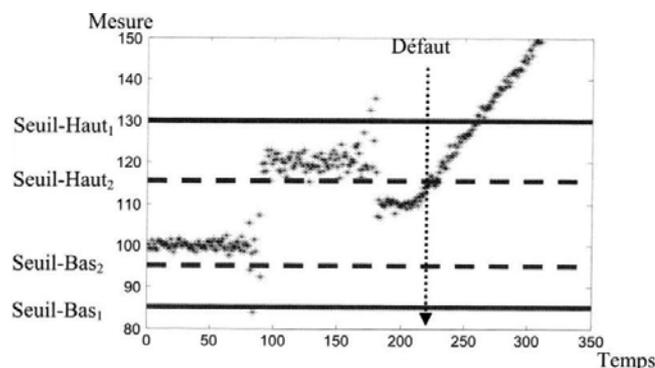


Figure 1.2. Alarmes Traditionnelles [BOI03]

Pour concevoir les alarmes traditionnelles, il n'est pas facile de choisir *a priori* les seuils de chaque variable. Avec cette technique, un mauvais paramétrage des seuils peut conduire à des alarmes répétitives non fondées dans par exemple 95% des cas ;

la conséquence de ce type de biais entraîne à terme un acquittement systématique de la détection par les opérateurs, même pour les 5% de bonnes détections [GEN07].

De plus, les variables n'étant pas indépendantes mais liées par des relations d'influence, en toute logique, leurs seuils devraient être reliés par les mêmes relations. Cela suppose que l'on utilise un modèle du procédé pour gérer la cohérence des seuils (le problème est qu'il est très difficile de posséder un modèle précis pour les systèmes complexes) [GEN07][AKB01][KOU02].

Pour mettre en évidence l'intérêt de prendre en compte plusieurs variables et leur relation pour faire le diagnostic, nous présentons dans ce qui suit un exemple simple. Dans la plupart des cas, l'utilisation de l'analyse monovariante qui traite de manière indépendante chaque variable, génère la non identification des états anormaux. Ceci est dû à la non indépendance des variables qui sont normalement corrélées. La Figure 1.3. est une illustration classique du fait d'utiliser des diagrammes de diagnostic indépendants (*separate control charts*) pour deux variables (y_1 et y_2).

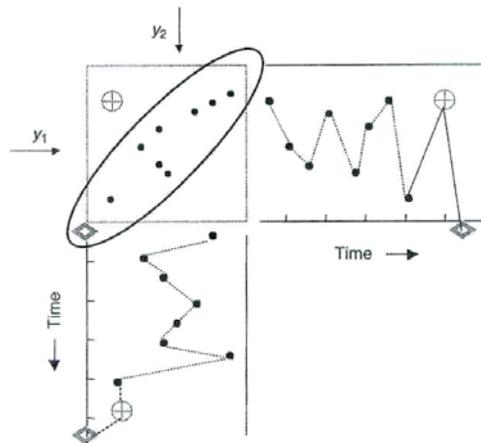


Figure 1.3. Le problème d'utiliser des diagrammes de diagnostic indépendants [KOU02]

Sur cette figure, les deux variables ont été tracées sur les axes (en haut à gauche), pour constituer les points (y_1, y_2) de la figure. Sur la même figure, le comportement des variables dans le temps a été tracé pour chacune d'elles (les graphiques vertical et horizontal), en incluant les seuils établis pour identifier les défaillances. Le graphique de chaque variable au cours du temps en y incluant les seuils correspond aux diagrammes de diagnostic indépendants (*separate control charts*).

Quand une variation normale est présentée, les points constitués par les paires (y_1, y_2) , correspondent à une distribution qui peut être associée à l'ellipse dessinée sur la figure tout en lui ajoutant une région de confiance. Le point marqué avec un « + », est clairement en dehors de la région de confiance établie par les caractéristiques statistiques des paires (y_1, y_2) en état normal, et pourtant lorsque l'on regarde les diagrammes de diagnostic indépendants (*separate control chart*) aucune des variables au moment de la défaillance (marquées aussi avec des « + ») n'est en dehors du seuil établi. De plus, ces valeurs sont plus basses que d'autres valeurs pour lesquelles le

système a été dans un état normal. Comme conséquence, si le diagnostic est fait uniquement à partir des diagrammes de diagnostic indépendants (*separate control chart*), plusieurs défaillances ne seront pas détectées. D'autre part, les mesures marquées avec un \diamond , sont en dehors des seuils minimaux établis pour chaque variable, mais en regardant le graphique qui symbolise la relation entre les deux variables, cet échantillon est dans l'état normal [KOU02].

En conséquence, plusieurs méthodes de diagnostic qui permettent d'analyser les relations entre plusieurs variables ont été proposées et il y a constamment des méthodes nouvelles ou des améliorations proposées.

Avoir une idée générale de l'organisation des méthodes utiles pour le diagnostic, permet de mieux choisir une méthode pour l'implanter dans un système de supervision. Une organisation des méthodes de diagnostic qui permet d'avoir une vision générale a été présentée par Venkatasubramanian et al. [VEN03]. La Figure 1.4 présente le schéma de classification des méthodes de diagnostic auxquelles nous avons ajouté les méthodes de classification par coalescence (*clustering*).

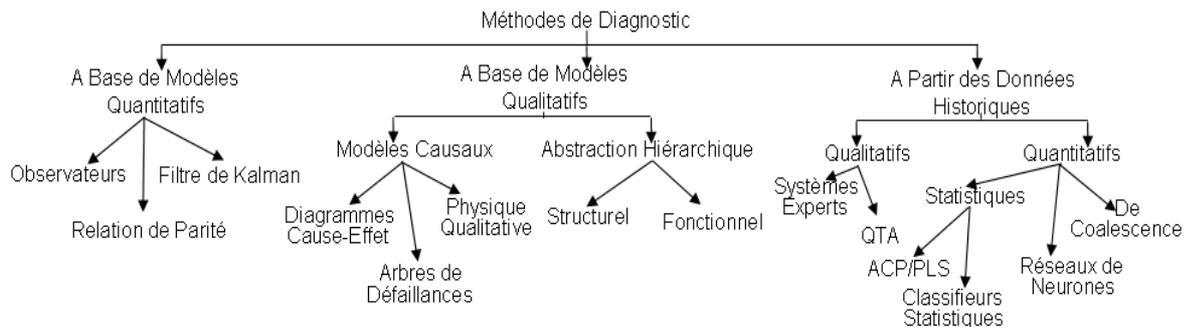


Figure 1.4. Organisation des Méthodes de Diagnostic [VEN03]

1.4.1 Méthodes basées sur des modèles

Traditionnellement, les problèmes de diagnostic ont été résolus par la construction de modèles très complets du procédé. La détection de défaillance en utilisant ces modèles, consiste à identifier l'écart entre le modèle et le système réel.

Néanmoins, quand les processus sont devenus plus complexes, l'obtention de modèles « classiques » est plus difficile en impliquant du temps et des efforts vraiment importants, qui ne sont pas justifiables pour l'industrie [KOU02].

1.4.1.1 Méthodes basées sur des modèles quantitatifs

Les méthodes basées sur des modèles quantitatifs comparent les valeurs des variables réelles avec les valeurs calculées en utilisant un modèle qui met en rapport les variables du processus. La comparaison des valeurs réelles à celles calculées donne des résidus. Si les résidus dépassent un seuil ou si leur valeur n'est pas très proche de zéro, la présence d'une défaillance est alors détectée.

Dans le cas de processus très complexes, multi-variables ou non linéaires, il n'est pas facile de construire un modèle suffisamment précis et/ou complet pour être utilisé lors de la comparaison avec le comportement du système réel afin de déterminer l'existence d'une défaillance.

Les équations théoriques représentent seulement une part du système (qui est la plus familière, masse, énergie, etc.). Plusieurs paramètres dans ce type de modèles doivent être fixés en utilisant des valeurs données par la littérature ou estimées en utilisant des historiques du fonctionnement du procédé. D'autre part, les tâches de diagnostic incluent l'analyse des actionneurs et des capteurs. Ces composants ne sont pas normalement pris en compte dans la modélisation du processus, et les inclure dégrade le modèle en une version très complexe difficile à utiliser pour des fins de diagnostic [KOU02].

Toutefois, s'il y a un modèle précis du système, les performances de ce type de méthodes sont très bonnes.

1.4.1.2 Méthodes basées sur des modèles qualitatifs

Les méthodes basées sur des modèles qualitatifs fonctionnent sur le même principe que celles basées sur des modèles quantitatifs, mais la relation entre les descripteurs est exprimée par des fonctions qualitatives. Ces méthodes ne sont pas basées sur l'écriture de lois physiques, par conséquent leurs performances dépendent fortement de la connaissance de l'expert qui formalise la relation entre variables.

1.4.2 Méthodes basées sur des données historiques (fouille de données)

Les automatismes récents incluent le stockage de plusieurs variables. C'est pourquoi, à l'heure actuelle, les industries disposent en général d'un vaste historique des processus en y incluant les comportements des variables en présence de défaillances. L'utilisation de ces données pour en extraire de la connaissance sur le procédé est donc amplement justifiable.

Les méthodes d'apprentissage et de reconnaissance de formes permettent à partir des données du processus, d'obtenir de l'information des états du système [KEM04a] [WAI00][LUR03][MARI03][CAS03].

Les méthodes basées sur la fouille de données effectuent le diagnostic à partir du traitement d'une grande quantité d'échantillons. Ces méthodes ont beaucoup de propriétés très utiles pour le diagnostic des processus. Elles peuvent travailler de manière indifférente avec des processus non linéaires parce qu'aucun modèle de processus n'est nécessaire, l'information est extraite des données d'entraînement (notion d'apprentissage) [KIO94].

Les techniques d'obtention d'information à partir de données, permettent de résoudre les problèmes d'extraction de la connaissance. Ces techniques sont intéressantes pour le développement de lois de supervision et de commande de systèmes pour lesquels il est difficile d'obtenir un modèle précis et utile pour le diagnostic. Ainsi, la connaissance est automatiquement extraite des données, où les classes obtenues correspondent à des situations de fonctionnement normal ou de défaillance [CHI01].

Il est important de remarquer que la caractérisation des systèmes complexes pour le diagnostic, doit inclure la connaissance de l'expert tout spécialement pour les méthodes de type flou.

Le grand inconvénient qui existe avec ces méthodes est la dépendance aux données d'entraînement. Elles ont besoin d'avoir des données d'entraînement qui permettent de caractériser chaque état ou défaillance. Néanmoins, il n'est pas nécessaire d'avoir un très grand nombre de données et il faut éviter le surentraînement (privilégier la généralisation). Certaines techniques permettent aussi de faire la mise à jour des classes (actualisation, apprentissage actif) en fonction de données reconnues (e.g. LAMDA).

Il est nécessaire de remarquer de plus, que leurs performances sont fortement liées à la bonne sélection des paramètres de chaque algorithme [BAB98][CAS03][FRA02][KAY95][LUR03].

Pour toutes les techniques basées sur la fouille de données, il faut différencier deux grandes phases : l'entraînement ou apprentissage afin d'établir les caractéristiques des classes et la reconnaissance de nouvelles données. La surveillance et le diagnostic correspondent à la phase de reconnaissance qui peut être faite en ligne et, en fonction de la méthode, peut posséder une procédure d'adaptation des classes.

Un prétraitement des données est fortement conseillé, d'une part pour établir les variables les plus pertinentes qui permettent de détecter les états du système, d'autre part pour éliminer des bruits et obtenir la meilleure information possible de chaque variable.

Dans le cas du diagnostic, un prétraitement des données est recommandé parce que les données brutes ne permettent pas toujours d'obtenir l'information nécessaire pour caractériser les classes. Ce problème se rencontre tout spécialement dans le cas de systèmes industriels, à cause du bruit de mesure qui est toujours présent sur les capteurs. Le prétraitement inclut des opérations classiques en automatique et en traitement du signal (filtrage, analyse fréquentielle, ...)[CHI01].

Le prétraitement, peut être appelé aussi « mise en forme des données » [GEN07]. Cette mise en forme, éventuellement accompagnée d'un premier niveau de validation, peut être implantée au niveau des capteurs et des actionneurs, aujourd'hui dotés de capacités de traitement numérique et d'une interface bidirectionnelle de

communication, et connus également sous le noms d'instruments « *smarts* » ou intelligents [RIV96].

Dans les systèmes de commande actuels, on peut accéder à des historiques des variables (enregistrement de la variable sur une durée paramétrable), ou de plusieurs variables à la fois s'il est intéressant de corrélérer leur évolution. Il peut être par exemple souhaitable de suivre une variable régulée en même temps que sa variable de commande et sa consigne. La durée de l'historique doit être choisie en fonction de la dynamique du procédé, en réaction à des changements de consigne, ou encore en réaction à certains défauts. En effet, les défauts les plus difficiles à détecter sont les défauts qui s'installent lentement, sous forme de dérive, car on peut mettre un certain temps à voir leurs effets apparaître clairement [GEN07]. Une alternative très intéressante pour prétraiter les variables en incluant l'évolution est d'analyser sa tendance pendant un période de temps déterminée et de classer une représentation qualitative de son comportement (e.g. tendance à augmenter, tendance à diminuer, etc.) Kemposky T. [KEM04a] présente deux exemples industriels où cette représentation permet d'identifier des états plutôt qu'en utilisant des variables numériques.

Les alternatives pour le prétraitement des données sont aussi très vastes. Par exemple, on peut effectuer une analyse en fréquence des signaux pour utiliser dans le diagnostic uniquement les fréquences les plus informatives. Un cas classique est d'analyser les fréquences parasites d'un signal, apportées par les vibrations de structures mécaniques ou dans les machines tournantes par exemple, pour en déduire des défauts. De même, des caractéristiques particulières, brutes ou issues du traitement des données, permettent de classer les observations pour établir les états comportementaux du système.

1.5 Diagnostic en utilisant des méthodes de classification

Les méthodes de classification sont une partie importante des méthodes basées sur la fouille des données.

Le principe de la surveillance d'un procédé dynamique à partir d'une méthode de classification, consiste à déterminer à chaque instant, la classe actuelle qui a préalablement été associée avec un état fonctionnel du système. La Figure 1.5. représente la supervision des processus utilisant des méthodes de classification.

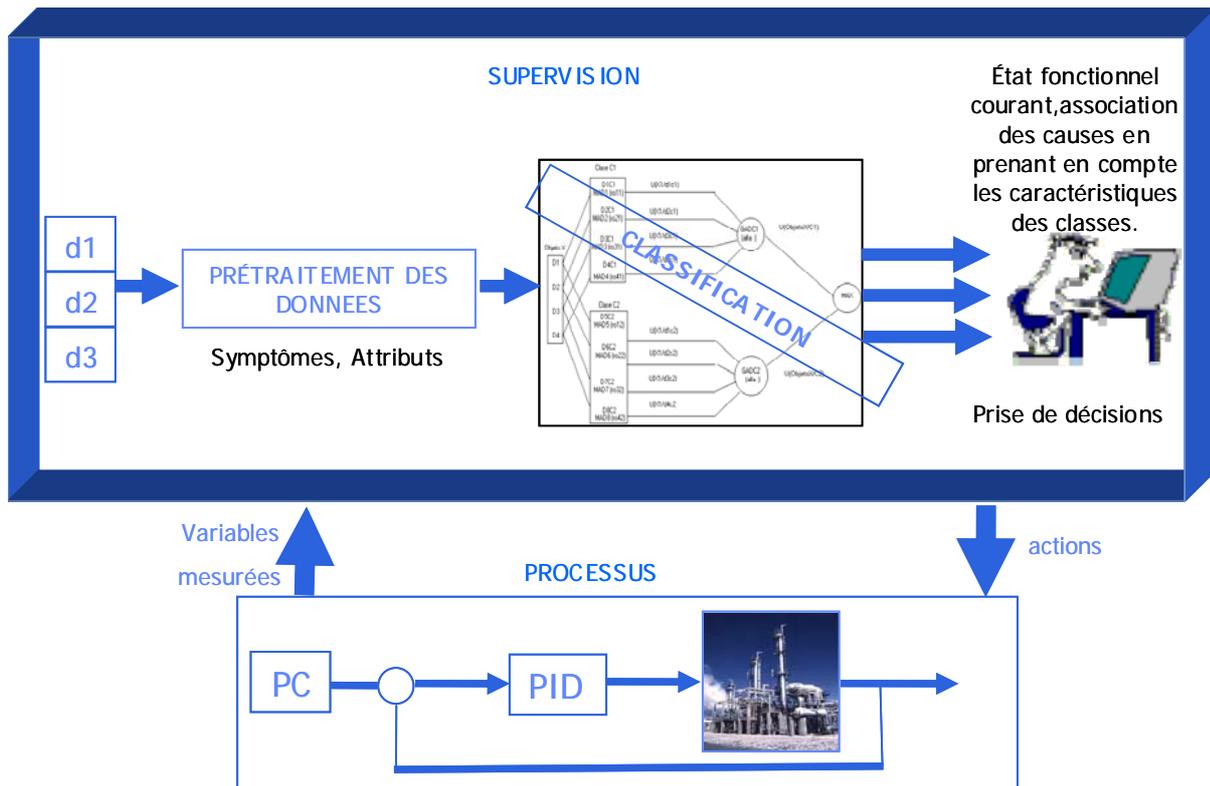


Figure 1.5. Schéma de la supervision utilisant des techniques de classification

Une vision synthétique du procédé ne peut pas être apportée par la visualisation des valeurs précises de centaines de variables sur un synoptique. La classification, qui permet de projeter ces nombreuses valeurs dans un espace de dimension réduite, est un outil très intéressant pour le diagnostic. Dans cette espace, l'opérateur peut apprendre à identifier des zones à risque ou relatives à des dysfonctionnements particuliers, il aura en conséquence des réactions beaucoup plus rapides aux problèmes.

Les méthodes de classification, spécialement les méthodes de type statistique, permettent d'établir un modèle des états du système (modèle comportemental), en incluant les états de défaillance. On peut recouvrir la connaissance de différents attributs relatifs à un comportement particulier, sans que ce comportement soit représenté par un ensemble de relations analytiques (modèle au sens classique de l'automatique). Ainsi, des caractéristiques statistiques d'un signal (e.g. moyenne et écart type) peuvent être considérées comme un « modèle » représentatif d'un comportement (normal ou non). Les modifications de ces caractéristiques permettent de détecter un fonctionnement anormal, que l'on pourra relier à certains défauts (variables en défaut) si l'on a une connaissance précise de leur influence [GEN07].

L'utilisation de techniques d'Intelligence Artificielle et de la Logique Floue pour le diagnostic de systèmes permet d'obtenir des résultats interprétables et offre une information utile pour la décision d'actions à effectuer en cas de défaillance.

Pour pouvoir réaliser un diagnostic en utilisant une méthode de classification, il y a deux phases principales: l'entraînement ou apprentissage et la reconnaissance.

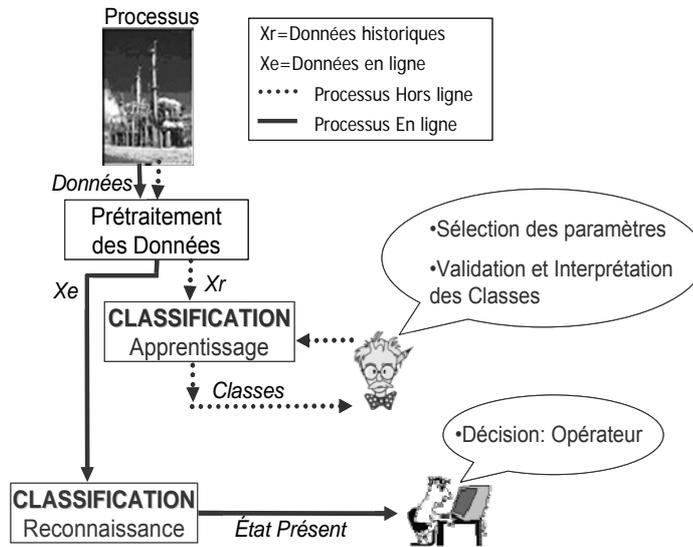


Figure 1.6. Schéma du diagnostic utilisant des techniques de classification

1.5.1 Phase d'entraînement ou apprentissage

Dans la première phase (*entraînement ou apprentissage*), l'objectif est de trouver, à partir de l'ensemble de données, les caractéristiques du comportement du système qui vont permettre de différencier les états dans lesquels le système peut se trouver (Figure 1.6). Cette phase est réalisée hors ligne et l'intervention de l'expert sur le processus est déterminante.

Dans cette phase, les objets sont regroupés et discriminés pour arriver à la construction de classes. Le résultat de l'entraînement est un classifieur ; on peut l'associer au modèle comportemental du système. Selon la méthode utilisée, le classifieur peut correspondre à des paramètres d'une équation ou aux valeurs et à la structure d'un système (par exemple un réseau de neurones). Le classifieur permet d'associer aux différentes classes, de nouvelles observations qui correspondent aux valeurs présentes des variables et permet d'associer les causes des défaillances quand la classe correspond à un état de défaillance.

En ce qui concerne l'apprentissage, il y a certaines questions qui n'ont pas encore reçu de réponse définitive [DUD01]. Par exemple, comment déterminer si l'algorithme d'apprentissage a convergé dans le temps imparti ? Quel est le nombre adéquat de données d'entraînement ? Quel est le nombre de variables qui permet d'extraire l'information nécessaire pour caractériser les états ? Quel est le nombre de classes optimal ? La réponse à ces questions dépend presque toujours de l'application. Nos travaux qui constituent des compléments aux méthodes de classification, cherchent à répondre en partie à ces questions.

Les travaux effectués lors de cette thèse se situent dans l'optimisation de l'espace de partition obtenu avec des méthodes de reconnaissance de formes de type flou. Nous cherchons à diminuer la dépendance aux paramètres propres à chaque algorithme d'apprentissage et à traiter le problème de la détermination du nombre de classes pour les algorithmes de classification où ceci est nécessaire *a priori*. Pour l'entraînement, nous proposons d'inclure une étape de validation et d'ajustement des classes afin d'optimiser la partition obtenue sans relancer l'algorithme de reconnaissance.

1.5.2 Phase de reconnaissance

Lors de la deuxième phase, la *reconnaissance* de nouvelles données est effectuée. De cette manière, l'état présent du système est identifié. Les données sont classées en fonction du classifieur obtenu lors de la phase précédente. Cette phase de reconnaissance peut être effectuée en ligne.

Avec des méthodes de classification, dans la phase d'entraînement on obtient un modèle comportemental du procédé en caractérisant les états (normaux ou non) par des paramètres extraits des variables du procédé. Dans la phase de reconnaissance quand un état est reconnu, si on parle d'un état de défaillance, les valeurs qui représentent la classe sont associées pour effectuer la localisation de la cause de la défaillance.

1.5.3 Méthodes d'apprentissage et de reconnaissance de formes utiles pour le diagnostic

Pour identifier les états des systèmes à partir de données, plusieurs algorithmes ont été proposés. Chacun permet d'obtenir une partition différente de l'espace des données. Dans ce qui suit, nous présentons une courte description de trois méthodes utilisées pour le diagnostic de systèmes qui permettent d'obtenir une partition floue.

- *Méthodes basées sur l'analyse de distance (Méthodes de coalescence) :*

Ces algorithmes ont comme paramètre d'entrée le nombre de classes K . Ils divisent l'ensemble de N objets dans K . D'abord ils proposent des centres des groupes et ensuite ils assignent chaque objet au centre le plus proche. Chaque classe est caractérisée par le centre ou prototype et par ses éléments.

C-Means est l'algorithme le plus connu parmi les méthodes basées sur la distance. Il fait une séparation de l'espace de type stricte (« dure »). La mesure de similitude qu'il emploie est la distance euclidienne, par conséquent les formes des groupes sont des hyper sphères. Une variation de *C-Means* est le *Fuzzy C-Means* (FCM) qui permet d'obtenir une matrice de partition floue ($U=[\mu_{kn}]$) qui contient les degrés d'appartenance de chaque objet x_n à chacune des k classes.

Un autre algorithme de ce type est le *Gk-Means*. Cet algorithme utilise une norme adaptative de la distance, en obtenant des classes hyper ellipsoïdales avec leurs axes

inclinés par rapport aux axes de l'espace des données [BAB98]. Dans ces méthodes, la première sélection des centres est aléatoire, et les résultats sont sensibles à l'initialisation.

- *Réseaux de Neurones* :

Les réseaux de neurones ont montré depuis de nombreuses années de grandes capacités à résoudre des problèmes de classification pour le diagnostic de systèmes [LUR03][BEZ81]. La structure générale d'un réseau de neurones utilisé pour le diagnostic consiste en trois couches. Le nombre de neurones de la première couche (d'entrée) est le nombre de variables du problème (d =nombre d'éléments par chaque individu x_n) et le nombre de neurones de la couche de sortie correspond aux K classes. Normalement la performance du réseau est directement liée à la taille de la couche cachée.

Dans la plupart des algorithmes d'apprentissage, il est nécessaire d'établir *a priori* le nombre de neurones de la couche cachée, mais les réseaux de neurones à fonctions de base radiales (RBF) trouvent automatiquement cette valeur [MARK96][JAN97]. Pour obtenir une partition floue avec les RBF, les neurones de la couche cachée ont des fonctions de base radiales (e.g. gaussiennes) et les fonctions de la couche de sortie sont de type linéaire. La partition de l'espace correspond à un ensemble de fonctions non linéaires.

- *LAMDA (Learning Algorithm for Multivariate Data Analysis)*:

LAMDA [AGU82] est une technique de classification floue basée sur l'évaluation de l'adéquation des éléments à chaque classe. Les éléments peuvent être décrits aussi bien de façon numérique que qualitative. Les degrés d'adéquation globale (GAD) d'un élément à chaque classe sont une fonction des adéquations marginales (MAD) représentant la contribution de chaque composant de l'individu x_n . Le résultat est une partition floue. La structure de l'algorithme est similaire à celle du réseau de neurones mais où les adéquations GAD et MAD sont des fonctions d'appartenance floue et leur agrégation se fait au moyen de connectifs de la Logique Floue (t-norme et sa duale).

L'avantage des techniques floues est que les résultats sont les appartenances des mesures courantes aux différents états fonctionnels identifiés sur des données historiques. Cette caractéristique permet d'avoir une connaissance plus complète de la situation réelle du système complexe et permet de prendre des décisions pour prévenir des changements vers des états non désirables (e.g. états de défaillance). Ceci est très intéressant parce que cela permet d'anticiper le basculement vers un état de défaillance. En effet, si le système est dans un état de fonctionnement normal, mais qu'il présente un degré d'appartenance élevé à un état de défaillance, ceci permet de d'anticiper et prendre des mesures pour diminuer la tendance au basculement dans l'état de défaillance. Cette caractéristique d'anticipation inhérente aux méthodes de classification floue est très intéressante dans le cadre du diagnostic. De plus, la connaissance sur l'état complet du système est une aide importante pour la décision de

l'opérateur, de cette manière l'opérateur a plus d'information (synthétique) sur les conditions du système et ses décisions seront plus acérées.

Prédire le comportement du procédé est important dans l'aide à la décision de l'opérateur humain. Cela permet d'anticiper sur la propagation de certains défauts, d'évaluer les risques et de prendre les bonnes décisions dans le temps requis. Posséder un modèle du comportement du procédé est évidemment un atout important pour une telle prédiction. Il est risqué bien sur de prédire avec un modèle de comportement normal l'évolution d'une situation [GEN07]. C'est pourquoi il est important de remarquer que le modèle comportemental doit être obtenu en incluant les défaillances qui ont été présentes antérieurement sur le procédé.

C'est dans ce contexte que nous avons procédé par ailleurs à une comparaison raisonnée des algorithmes séquentiels existants: la méthode LAMDA [AGU82][WAI00][KEM03], d'autres approches floues [BAB98] [JAN97][BEZ81] et neuronales [JAN97][MIC94][BIS95][MAR96], ou des méthodes statistiques [LIM00][HAN01]. Nous présentons une description plus détaillée des méthodes et une comparaison entre elles dans le chapitre suivant.

Kempowsky T. [KEM04a] montre qu'un des inconvénients à utiliser des méthodes de classification pour le diagnostic est qu'en général, le nombre de classes est supposé connu au départ, i.e. la connaissance est supposée exhaustive. Toute observation doit appartenir à une des classes définies. Ceci est loin d'être le cas dans le contexte du diagnostic. Par ailleurs, il reste souvent des modes de fonctionnement dont nous ignorons les caractéristiques. Dans ce cas, il est nécessaire d'utiliser des techniques qui permettent de prendre en compte la détection de nouvelles classes, d'apprendre leurs caractéristiques et de les inclure dans le système de détection ; ce qui implique un système évolutif, capable de prendre en compte tout nouveau phénomène normal ou anormal présent dans le processus surveillé.

La connaissance des états du système est obtenue à partir de données historiques et si la technique de classification est adaptative (par exemple comme LAMDA), il est possible de traiter de nouveaux états qui n'ont pas été inclus dans les données historiques initiales. Ces méthodes n'ont pas besoin d'un modèle mathématique, et par conséquent peuvent être utilisées dans des systèmes complexes, pour la plupart non linéaires et composés de plusieurs sous-processus qui interagissent entre eux. Néanmoins, ces méthodes dépendent fortement de la qualité des données et des valeurs des paramètres de l'algorithme. Pour cela, il est important de travailler sur le développement de techniques qui permettent d'améliorer la performance en optimisant automatiquement la segmentation de l'espace des données. Dans ce document nous proposons une approche qui va dans ce sens. Elle est basée sur un critère d'évaluation de la partition (i.e. de la distribution et de la forme des classes) qui permet de choisir automatiquement la meilleure partition ou de faire un regroupement de classes si nécessaire.

1.6 Conclusion

Dans ce chapitre nous avons présenté brièvement les différents concepts généraux sur lesquels se base le diagnostic des systèmes complexes. Parmi les méthodes proposées dans la littérature, nous nous sommes plus particulièrement intéressés à celles basées sur la fouille de données. Nous avons montré qu'à partir des classificateurs flous il est possible d'établir un modèle comportemental du procédé qui soit utile pour identifier les états fonctionnels incluant les états de défaillance.

En effet, l'extraction d'information pour caractériser les états d'un procédé a montré des résultats intéressants pour les systèmes complexes où il est difficile voire impossible d'estimer un modèle qui soit complet et utile pour le diagnostic. L'utilisation de méthodes de fouille de données pour le diagnostic de processus est une alternative intéressante à l'utilisation d'un modèle mathématique et ces méthodes fournissent généralement de bons résultats (en termes de qualité de performances). Le potentiel des techniques basées sur la fouille des données est énorme.

L'objectif du diagnostic est de constater l'apparition d'un défaut, d'en trouver la cause puis d'en déduire la marche à suivre afin d'assurer la sûreté de fonctionnement du procédé ; l'utilisation donc des méthodes basées sur des variables du système permet d'identifier plus facilement la cause du système parce que la définition des états est directement en relation avec les variables du système.

Parmi les différentes méthodes proposées, nous nous sommes plus particulièrement intéressés aux méthodes de coalescence qui sont basées sur l'analyse des distances (*C-Means*, ses extensions floues *Fuzzy C-Means*, *GK-means*), aux réseaux de neurones et à la méthode LAMDA qui peut être vue comme un compromis entre les deux premiers types de méthode. Dans le chapitre suivant, une description plus fine de ces méthodes est effectuée ainsi qu'une comparaison de leurs performances afin d'établir leurs atouts ou leurs faiblesses pour pouvoir ensuite proposer des améliorations.

Chapitre 2

2. COMPARAISON DE METHODES DE CLASSIFICATION

2.1 Introduction

Comme nous l'avons présenté dans le premier chapitre, le diagnostic de systèmes complexes peut être vu comme la détection de modes de fonctionnement en incluant les défaillances. L'utilisation des techniques de classification permet d'extraire des données historiques, l'information qui caractérise les états et les défaillances d'un système.

Les techniques de fouille de données permettent de résoudre les problèmes d'extraction de la connaissance. Ces techniques sont intéressantes pour le développement de lois de supervision et de commande de systèmes pour lesquels il est difficile d'obtenir un modèle précis et utile pour le diagnostic. Ainsi, la connaissance est automatiquement extraite des données, où les classes obtenues correspondent à des situations de fonctionnement normal ou de défaillance [CHI01].

Actuellement, les techniques qui permettent d'extraire l'information à partir des données sont de plus en plus appliquées dans l'industrie pour la supervision et la commande de processus. Ceci est dû à la possibilité d'obtenir une connaissance du comportement des systèmes complexes, par exemple de systèmes avec des dynamiques hautement non linéaires et des paramètres variant dans le temps [LUR03][MARI03][CAS03][GRI02][WAI00].

Dans ce chapitre l'objectif est de présenter le contexte général des méthodes de classification basées sur la fouille de données. Nous présentons une comparaison entre différentes techniques de classification et nous introduisons la méthode de classification LAMDA. Cette comparaison met en lumière les avantages et les limitations de ces

méthodes, spécialement celles de type flou et la méthode LAMDA est replacée dans le contexte général des méthodes de classification utiles pour le diagnostic des systèmes.

2.2 Contexte général des méthodes de classification basées sur la fouille de données

De manière générale, on peut diviser les méthodes de classification entre méthodes avec apprentissage supervisé et méthodes avec auto-apprentissage.

Les méthodes avec apprentissage supervisé ont besoin de connaître *a priori* les classes associées aux données d'entraînement. L'apprentissage est basé, pour la plupart des méthodes, sur la minimisation de l'erreur d'entraînement.

Pour les méthodes avec auto-apprentissage, l'étape d'entraînement est utile pour ajuster les paramètres de l'algorithme de classification. Il n'y a pas de « professeur », ni de classes établies *a priori*. L'algorithme de classification regroupe les données en fonction de caractéristiques semblables [DUD01]. Dans ce cas, l'apprentissage n'est pas basé sur une erreur d'entraînement. Dans certaines méthodes (e.g. LAMDA), la classification peut être effectuée chaque fois qu'une donnée est analysée, et les paramètres du classifieur sont mis à jour à chaque échantillonnage. Dans d'autres cas (e.g. GK-Means), la définition des classes est basée sur l'optimisation d'une fonction objectif (définie dans l'espace de données) : l'idée est de diminuer la distance entre les données d'une même classe. Ainsi de manière itérative et en utilisant tout l'ensemble des données, les classes sont adaptées pour regrouper les données qui sont les plus proches.

Dans tous les cas, un vecteur résumant l'information (individu) qui correspond aux échantillons du système, est analysé pour trouver les caractéristiques des classes. La différence entre les deux types de méthodes est la phase d'apprentissage.

Dans le contexte des méthodes basées sur la fouille de données, les méthodes qui se situent dans le "*Machine Learning*" permettent d'obtenir des règles ou des expressions de classification plus simples, de telle sorte que les résultats obtenus soient facilement interprétables. Un des objectifs de ces méthodes est que l'intervention humaine dans la phase d'entraînement soit minimale et qu'en même temps, les classes obtenues soient facilement interprétables.

C'est dans ce contexte, que ce chapitre procède à une comparaison raisonnée des algorithmes séquentiels existants, à savoir : la méthode LAMDA, des approches floues comme GK-Means, neuronales comme les réseaux de neurones à fonctions de base radial-RBF et des méthodes statistiques comme le discriminant linéaire. Une brève description des cinq méthodes employées pour la comparaison est donnée dans ce qui suit, avec tout d'abord les méthodes avec apprentissage supervisé, puis les méthodes avec auto-apprentissage.

2.3 Description des différents Méthodes

2.3.1 Méthodes avec Apprentissage Supervisé

2.3.1.1 Discriminant Linéaire – ACP

C'est la technique de classification la plus classique. Des travaux effectués sur la comparaison de méthodes de classification [LIM00][MIC94] et des travaux récents dans le cadre du diagnostic basé sur ces méthodes [MARI03][CAS03], concluent qu'il est important d'analyser le discriminant linéaire car dans plusieurs cas cette méthode a montré de très bonnes performances. Il s'agit d'un algorithme standard et qui est le plus utilisé dans différents systèmes de diagnostic industriel.

Le discriminant linéaire [FIS36] suppose une distribution normale des données avec des moyennes différentes. Ce type de distribution est justifié quand il y a une grande quantité de données. L'objectif est de trouver les hyperplans qui divisent l'espace des données. En principe, un hyperplan de dimension d (avec d égal au nombre de descripteurs) est tracé de telle manière qu'il passe par le milieu de la ligne qui unit les centres de chaque classe. L'algorithme cherche à obtenir des fonctions discriminantes paramétriques (hyperplans), à partir de la minimisation de l'erreur quadratique de classification. Par conséquent, c'est une technique qui requiert de connaître *a priori* les classes de l'ensemble des données d'entraînement.

Une fois la partition de l'espace obtenue, de nouvelles données peuvent être classées.

2.3.1.2 Les K plus proches Voisins (*K- Nearest Neighbors*)

C'est la première méthode de classification qui ne suppose pas un type de distribution des données et c'est la première à proposer une analyse de similitude entre des données en utilisant la distance entre elles [FIX51].

L'algorithme fait un calcul de distance entre tous les individus et chaque objet est classé dans le groupe où se trouvent ses K plus proches voisins. Les K plus proches voisins sont des individus de l'ensemble des données d'entraînement. Habituellement, la distance employée pour l'analyse est la distance euclidienne. Normalement le nombre des K voisins est un ou trois, mais cette valeur dépend de l'application [DUD01].

2.3.1.3 Réseaux de Neurones

Les réseaux de neurones ont montré depuis de nombreuses années de grandes capacités à résoudre des problèmes de classification pour le diagnostic de systèmes [LUR03]. Le traitement en parallèle et la distribution de l'information entre les neurones permettent d'avoir une grande robustesse et de diviser l'espace de données avec des fonctions non-linéaires.

La structure générale d'un réseau de neurones utilisée pour le diagnostic de systèmes, est généralement constituée de trois couches : une couche d'entrée, une couche cachée et une couche de sortie (Figure 2.1). Le nombre de neurones de la première couche correspond au nombre de descripteurs du problème (d =nombre de descripteurs de chaque individu x_n) et le nombre de neurones de la couche de sortie correspond aux K classes.

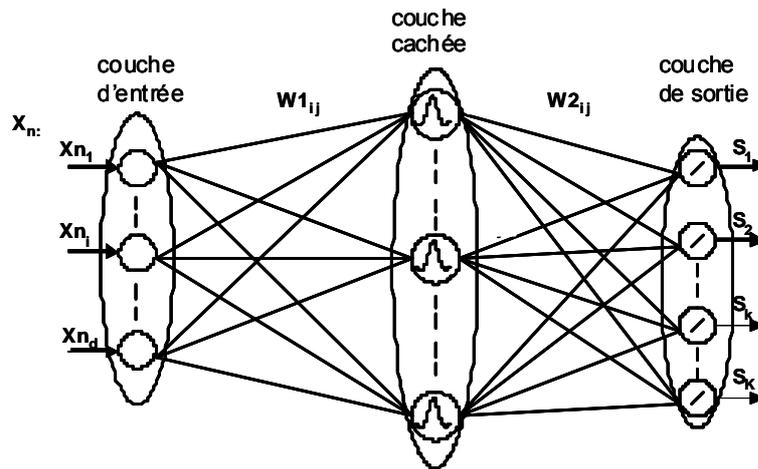


Figure 2.1. Structure du réseau de type RBF pour le diagnostic

L'équation générale d'un neurone est la suivante :

$$y_i(t) = \sum_{j=1}^N w_{ij} x_j(t) \quad (2.1)$$

$$u_i(t) = f(y_i(t))$$

$y_i(t)$ est l'état interne du neurone i et $u_i(t)$ est son état de sortie ou l'activation de sortie. L'apprentissage d'un réseau de neurones est basé sur l'ajustement des poids d'interconnexion w_{ij} entre les neurones. Pour l'entraînement, il y a différentes techniques d'apprentissage, qui se caractérisent par le besoin d'un grand nombre d'exemples d'entraînement et d'un temps de calcul souvent très long.

Normalement, les performances du réseau sont directement liées à la taille de la couche cachée. Pour la plupart des algorithmes d'apprentissage, il est nécessaire d'établir *a priori* la quantité de neurones de la couche cachée [LUR03] [BIS95]. Dans le cas du diagnostic, les algorithmes qui trouvent automatiquement cette valeur sont des plus intéressants; c'est le cas des réseaux de neurones à fonction de base radiale (RBF).

L'algorithme d'entraînement des RBF [POW87] détermine les poids de la couche cachée en utilisant la fonction de distance euclidienne. Les paramètres sont définis en utilisant les distances entre les données d'entraînement qui appartiennent à une même classe. L'algorithme d'apprentissage des RBF est plus rapide que ceux d'autres types de

réseaux [CHE91], puisque leur apprentissage peut être réalisé à l'aide de la méthode des moindres-carrés ordinaires (MCO) et les RBF en général ont beaucoup moins de connexions que les réseaux multicouches, mais ils sont aussi basés sur une minimisation de l'erreur d'entraînement, par conséquent ils ont besoin d'un grand nombre d'itérations pour améliorer leurs performances.

Pour la reconnaissance de nouvelles données, la classification est faite à partir de la structure de distribution des données d'entraînement.

Avec les RBF, pour obtenir une sortie non binaire, c'est-à-dire des valeurs entre 0 et 1, des fonctions de base radiale définies entre 0 et 1 (par exemple de type gaussienne) sont choisies pour les neurones de la couche cachée et les fonctions de la couche de sortie sont de type linéaire (voir Figure 2.1). La partition de l'espace correspond à un ensemble de fonctions non linéaires qui permettent d'obtenir des classes avec des formes très complexes.

Les applications des RNA dans le cas du diagnostic automatique montrent que grâce à la capacité de généralisation de cette technique, les résultats sont satisfaisants. Les inconvénients avec les RNA sont le temps d'apprentissage et la difficulté d'interprétation de la structure des classes, interprétation souvent nécessaire dans l'implémentation d'une technique de classification pour le diagnostic.

2.3.1.4 Arbres de Décision – CART

Les arbres de décision sont des techniques très employées dans les cas où il est important de trouver des règles afin de définir la classification. Les caractéristiques de la structure des arbres permettent une représentation interprétable des résultats.

Un arbre de décision correspond à une structure arborescente formée par des nœuds internes (unité de décision qui permet de déterminer le nœud suivant) et terminaux (lorsqu'il n'y a pas de nœud fils) reliés par des branches. Pour arriver à chaque nœud terminal, il y a une trajectoire unique qui débute avec le nœud racine. La trajectoire correspond à une règle de décision formée par une conjonction (ET) de plusieurs conditions de test.

Un arbre de décision divise l'espace des données en régions et associe à chacune un niveau ou une valeur qui caractérise les données du sous-espace.

Pour classifier en utilisant un arbre de décision, un objet se présente dans un nœud "racine" et en fonction du résultat d'une règle de décision effectuée par un nœud interne, est dirigé suivant la branche de l'arbre vers un des nœuds fils. Le processus est répété jusqu'à arriver à un nœud externe où une valeur ou un niveau est assigné à l'objet.

Dans les arbres de décision utilisés pour résoudre des problèmes de classification, chaque nœud externe a un niveau qui correspond à la classe de l'objet qui est analysé. Pour chaque objet à classer, il existe uniquement un chemin valable à travers l'arbre.

Les algorithmes les plus représentatifs des arbres de décision sont ID3 [QUI86], C4 et CART (*Classification And Regression Trees*) [LIM00] [JAN97] [LOP89].

CART est utilisé pour l'analyse de régression ou classification d'ensembles grands et complexes de données multi-variables [BRE84]. L'algorithme CART construit l'arbre en utilisant des partitions récursives. Lors de l'apprentissage, un premier arbre est créé en utilisant l'ensemble des données d'entraînement. L'algorithme cherche de manière récursive, les relations entre les variables pour obtenir une séquence optimale de divisions binaires, ce qui aboutit ainsi à un arbre très étendu. Ensuite l'arbre est réduit en tenant compte du principe de coût minimal de l'erreur.

Les arbres de décision permettent de travailler avec des données quantitatives et qualitatives. Pour les arbres de décision non flous, la classification est dite "dure" (un élément ne peut appartenir qu'à une seule classe), ce qui peut être un inconvénient surtout quand il s'agit de problèmes avec beaucoup de variables. En effet, un même élément peut être à la frontière de plusieurs classes sans que l'on ait cette information qui pourrait être très utile pour le diagnostic.

2.3.2 Méthodes avec Auto-Apprentissage

2.3.2.1 Méthodes de coalescence (*clustering*)

Ces algorithmes ont comme paramètre d'entrée le nombre de classes. Ils partagent l'ensemble de N objets en K groupes, ainsi la similarité à l'intérieur d'un même groupe est élevée mais faible entre les différentes classes. Pour ce faire, ces algorithmes itèrent en deux étapes, d'abord ils calculent les centres des groupes et deuxièmement ils assignent chaque objet au centre le plus proche. Chaque classe est caractérisée par le centre ou prototype et par ses éléments. Le prototype des classes est le point dans l'espace de dimension d (d correspond au nombre de descripteurs) où la somme des distances à tous les objets d'un même groupe est minimale.

La similarité est analysée au travers de la minimisation d'une fonction objectif qui est définie en fonction de la distance de chaque objet à l'élément prototype de chaque classe. La forme des groupes dépend du type de distance employée. Ces algorithmes fournissent une représentation géométrique des classes [BEZ05].

Parmi les différentes méthodes basées sur la distance, *K-Means* est sans nul doute l'algorithme le plus connu. Il effectue une séparation stricte de l'espace (c'est-à-dire chaque objet n'est associé qu'à une seule classe). La mesure de similarité utilisée est la distance, pour le calcul plusieurs mesures peuvent être utilisées (e.g. euclidienne, Mahalanobis, etc.), par conséquent la forme géométrique change en fonction de la mesure employée (e.g. pour le cas euclidien la géométrie des groupes est une hypersphère) [GEN07]. Une variante de *K-Means* est le *Fuzzy K-Means* (FKM) qui permet d'avoir un regroupement de type flou. FKM est un processus itératif qui permet d'obtenir comme résultat la matrice de partition floue ($U=[\mu_{kn}]$) laquelle contient les degrés d'appartenance de chaque objet x_n à chacune des K classes.

La fonction objectif sur laquelle est basée le FKM est formulée comme suit [BEZ81]:

$$J(X; \mu, V) = \sum_{k=1}^K \sum_{n=1}^N (\mu_{kn})^m D_{nk}^2 \quad (2.2)$$

La mesure de distance D_{nk}^2 est donnée par :

$$D_{nk}^2 = (x_n - v_k)^T A (x_n - v_k) \quad (2.3)$$

Le plus souvent, la matrice A est la matrice identité, et dans ce cas on retrouve la définition de la distance euclidienne. V est le vecteur des centres ou des prototypes de chaque classe (v_k). L'exposant de pondération (*fuzziness*) m ($m > 1$) conduit à l'obtention d'une partition plus ou moins floue. Ce paramètre doit être choisi par l'utilisateur ainsi que le nombre K de classes.

Un autre algorithme de ce type, reconnu pour ses très bonnes performances, est le *GK-Means* [GUS92]. Cet algorithme utilise une norme adaptative de la distance, en obtenant des classes de géométrie hyper-ellipsoïdale avec des axes qui ne sont pas nécessairement parallèles aux axes de l'espace des données. Ceci constitue un avantage face à l'algorithme FKM parce que le GK-Means permet à chaque groupe d'adapter sa norme de distance en fonction de la structure topologique locale des données. Chaque groupe a une matrice de norme induite A_k qui permet d'obtenir la norme de distance D_{nk}^2 propre à chaque classe.

Les matrices A_k sont des variables d'optimisation afin que chaque groupe adapte sa mesure de distance en fonction de la structure topologique des données mais avec une contrainte : un volume constant pour toutes les classes $|A_k| = \zeta_k$.

$$A_k = [\zeta_k \det(F_k)]^{1/N} F_k^{-1} \quad (2.4)$$

F_k est la matrice de covariance floue de la classe k :

$$F_k = \frac{\sum_{n=1}^N (\mu_{kn})^m (x_n - v_k)^T (x_n - v_k)}{\sum_{n=1}^N (\mu_{kn})^m} \quad (2.5)$$

Pour ces méthodes, le choix initial des centres (v_k) est aléatoire, ce qui fait que les résultats varient selon l'initialisation. D'autre part, les algorithmes FKM et GK-Means présentent une sensibilité élevée au paramètre m . De plus, l'interprétation des paramètres à choisir par l'utilisateur n'est pas évidente, ceci ne permet pas une utilisation aisée par un non spécialiste de l'algorithme (par exemple un expert du système de supervision ou un médecin).

2.3.3 LAMDA (*Learning Algorithm for Multivariate Data Analysis*)

LAMDA (*Learning Algorithm for Multivariate Data Analysis*) est une technique de classification floue basée sur l'évaluation de l'adéquation des individus à chaque classe. LAMDA a été proposée par Joseph Aguilar-Martin, en collaboration avec plusieurs chercheurs [AGU82]. LAMDA a été utilisé dans des domaines très divers tels que la psychologie [GALI02], les procédés biotechnologiques [WAI00] ainsi que les procédés industriels [KEM04a][HER06][ORA05].

Une des motivations, à l'origine du développement de cette méthode, est la volonté d'appréhender les difficultés inhérentes au diagnostic à base de données historiques. En effet :

- Plus il y a de descripteurs considérés pour la description d'un individu, plus il est probable de lui trouver des différences par rapport aux autres. Ceci peut amener à une classification avec un nombre excessif de classes, dont la plupart avec un seul individu.
- Certaines variables qui caractérisent les individus sont représentées par des valeurs imprécises ou approximatives, ou bien correspondent à des situations qualitatives pas forcément binaires, c'est-à-dire que leur présence ou leur absence dans un individu est connue avec un certain degré d'incertitude.

En ce qui concerne la classification avec LAMDA, les éléments peuvent être décrits aussi bien de façon numérique que qualitative. Le degré d'adéquation globale (GAD, Global Adequacy Degree) d'un individu x_n à chaque classe k correspond à l'agrégation des fonctions des adéquations marginales (MAD, Marginal Adequacy Degree) représentant la contribution de chaque descripteur x_{nd} de l'individu. Une classe non informative (NIC, Non Informative Class) est toujours fournie de telle sorte que tous les éléments dans l'espace de données aient le même degré d'adéquation à la NIC fournissant ainsi un seuil minimum pour la décision d'attribution. La figure 2.2 fournit de manière schématique la structure de LAMDA pour l'analyse du degré d'appartenance de la donnée x_n à chaque classe.

Le résultat, analogue à la matrice de partition floue ($U=[\mu_{kn}]$) du FKM, est un vecteur des appartenances μ_{kn} de chaque individu n à chaque classe k . Ces valeurs représentent une partition floue de l'espace des données. La structure pour l'analyse des individus est similaire à celle du réseau de neurones mais les adéquations GAD et MAD sont des fonctions d'appartenance floue et leur agrégation se fait en utilisant des connectifs issus de la logique floue (t-norme et sa duale) [PIE91].

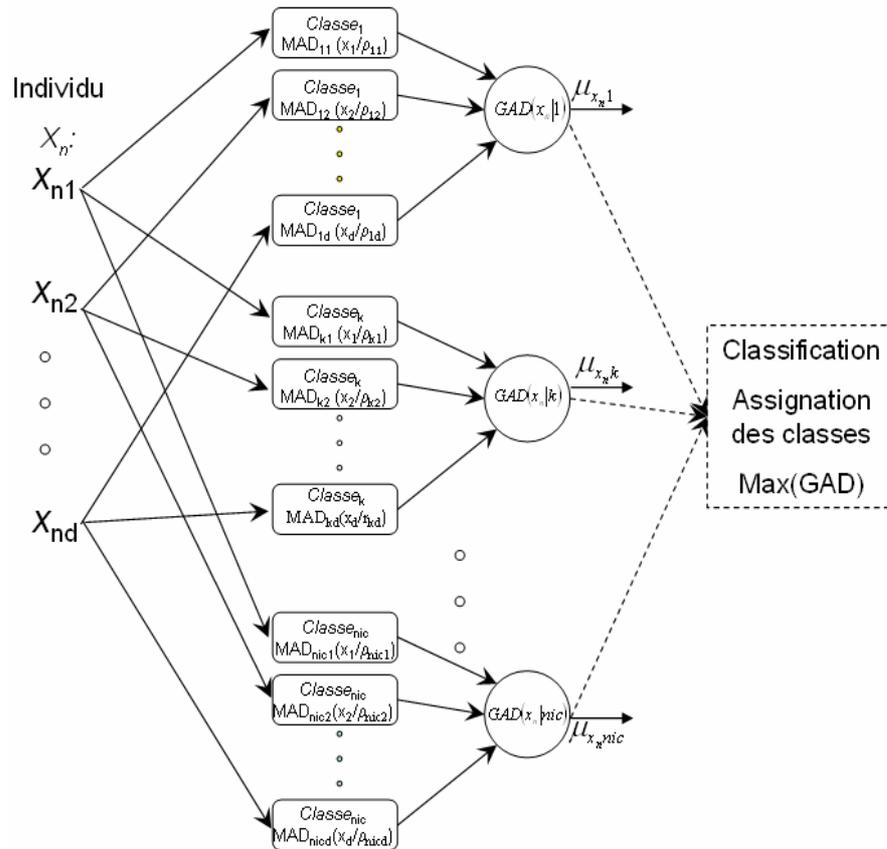


Figure 2.2. Schéma de la méthode de classification LAMDA

L’algorithme d’apprentissage calcule les paramètres des fonctions MAD. Le type d’apprentissage peut être supervisé ou non (auto apprentissage).

Dans le cas d’un apprentissage supervisé, les paramètres des fonctions MAD sont calculés à partir de données d’entraînement. Si une reconnaissance passive a été choisie, toutes les fois qu’un individu est assigné à la NIC, c’est-à-dire que le degré d’appartenance à la NIC est le plus élevé (par rapport à ceux des autres classes), l’individu est considéré comme non reconnu.

Si l’auto-apprentissage est choisi, une nouvelle classe est créée et initialisée avec cet individu et les paramètres de la NIC. D’autre part, si l’individu est assigné à une classe existante, les paramètres de cette classe sont mis à jour avec les valeurs de l’individu. Avec cet algorithme (d’auto-apprentissage) il est possible de commencer une classification sans information *a priori*, n’ayant seulement que la classe NIC quand le premier individu est traité. Mais il est aussi possible d’initialiser le processus de classification en présentant à l’algorithme des données historiques classées par un expert (professeur) et d’effectuer une actualisation des classes. D’ailleurs, il est également possible de maintenir l’apprentissage après une première phase d’entraînement supervisé. La figure 2.3 présente un schéma du principe d’apprentissage de l’algorithme.

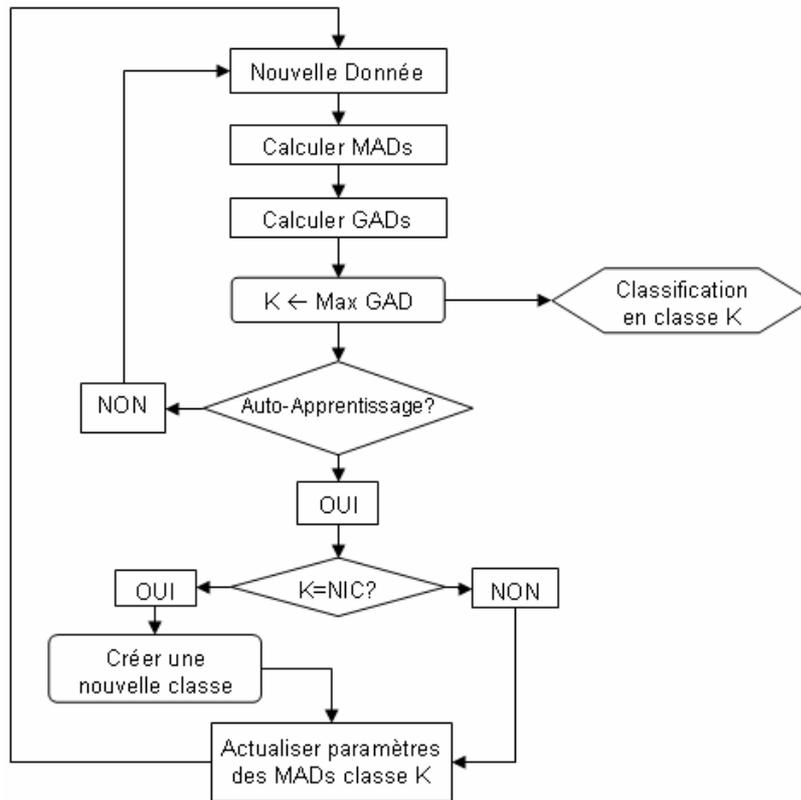


Figure 2.3. Algorithme d'apprentissage de LAMDA

2.3.3.1 Degré d'Appartenance Marginale - Fonctions MADs

Dans une direction j qui correspond à un descripteur $x_j = x_{nj}$, de l'individu x_n , l'adéquation de l'individu à la classe k est une fonction paramétrique qui correspond à $MAD(x_j|k) = MAD(x_j|w_j)$. La classe k est caractérisée par le vecteur de paramètres w qui est défini dans l'espace des descripteurs R^d

Pour maintenir la concordance avec la logique floue, les conditions suivantes dans les équations du MAD ont été imposées :

$$0 \leq MAD(x_j|k) \leq 1 \quad (2.6)$$

$$\exists \bar{x}_j; MAD(\bar{x}_j|k) \geq MAD(x_j|k) \quad (2.7)$$

où \bar{x}_j est le prototype de la classe k .

- Descripteurs Quantitatifs : Dans ce cas la valeur originale y_j de chaque descripteur est normalisée par : $x_j = \frac{(y_j - y_{j\min})}{(y_{j\max} - y_{j\min})}$ où $y_{j\min}$ et $y_{j\max}$ correspondent respectivement aux valeurs minimum et maximum pour le descripteur x_j . De cette manière LAMDA ne traite que des entrées normalisées ($0 \leq x_j \leq 1$) sans perte de généralité.

Deux types de fonction d'adéquation marginale peuvent être considérés :

1) Binaire floue, où le résultat attendu est une frontière de décision, mais il n'y a aucun élément central ou prototype. C'est le cas de la fonction Binomiale floue qui sera présentée dans ce qui suit.

2) Approche Prototypique, où une fonction de similitude établit la relation de n'importe quel élément à un prototype comme dans le cas des fonctions Gaussiennes et des fonctions Binomiales centrées floues décrites ci-dessous.

- Fonction Binomiale floue : avec $x_n = \begin{cases} 0 \\ 1 \end{cases}$ et une classe k caractérisée par $\rho = \text{prob}[x=1]$ et le $MAD[x|\rho] = \begin{cases} 1-\rho \\ \rho \end{cases}$. La fonction binomiale floue est défini comme une fonction monotone qui satisfait ces conditions de frontière et qui peut être considérée comme une mesure floue qui interpole la probabilité binomiale. En tenant compte de ces conditions, l'équation (2.8) a été proposée :

$$MAD_{kj}[x_j|\rho_{kj}] = \rho_{kj}^{x_j} (1-\rho_{kj})^{(1-x_j)} \quad (2.8)$$

La valeur maximale d'adéquation pour cette fonction peut être soit en 0 soit en 1 si $\rho \neq 0.5$. Si $\rho=0.5$, aucune valeur ne peut être considérée comme prototype. Il en ressort que la valeur $\rho_{\text{nic}}=0.5$ a été choisie comme celle qui donne le seuil minimum pour la décision d'attribution.

- Fonction Gaussienne : Pour respecter un prototype $\bar{x} = \rho$, il est nécessaire d'introduire un paramètre σ qui représente la dispersion en fonction de ce prototype, de telle sorte que : $\forall x \neq \rho; MAD[\rho|\rho, \sigma] \geq MAD[x|\rho, \sigma]$ et si $\sigma_1 \geq \sigma_2 \quad \forall x \neq \rho$, les degrés d'adéquation sont dans l'ordre : $MAD[x|\rho, \sigma_2] \geq MAD[x|\rho, \sigma_1]$. La fonction typique qui satisfait ces conditions est la fonction Gaussienne :

$$e^{-\frac{1}{2} \left(\frac{x_j - \rho_{kj}}{\sigma_{kj}} \right)^2} \quad (2.9)$$

La valeur minimale d'adéquation, quand $x_n = \begin{cases} 0 \\ 1 \end{cases}$ est obtenue avec $\rho = \frac{1}{2}$ et $\sigma = \frac{1}{4}$.

Choisir une fonction de distribution gaussienne $F(x)$, avec un écart type de 0.25 et une moyenne 0.5, permet d'avoir une fonction très étendue. Compte tenu que pour ce type de fonction, les valeurs de $F(x) \quad \forall x = \rho \pm 3\sigma$ sont pratiquement négligeables, on peut considérer cette fonction comme une approximation de la fonction de distribution

uniforme pour les x compris entre 0 et 1. Pour cette raison, ces valeurs ont été choisies comme paramètres de la classe NIC pour la fonction MAD Gaussienne.

- Fonction Binomiale Centrée : De manière similaire aux fonctions présentées précédemment, s'il y a un prototype $\bar{x} = \rho$, et un paramètre φ qui représente la dispersion, une fonction qui inclut les conditions considérées antérieurement correspond à l'équation (2.10). Cette fonction est basée sur la fonction Binomiale floue (eq. (2.8)), et elle inclut l'analyse de la distance entre le prototype et l'individu.

$$MAD_{kj} [x_j | \rho_{kj}, \varphi_{kj}] = \varphi_{kj}^{1-|x_j-\rho_{kj}|} (1-\varphi_{kj})^{|x_j-\rho_{kj}|} \quad (2.10)$$

La valeur maximale d'adéquation pour cette fonction correspond au cas où x_j est égal au prototype. Le seuil minimum d'appartenance s'obtient par $\varphi_{nic}=0.5$. En effet, si $\varphi=0.5$, aucune valeur de x_j ne peut être considérée comme prototype car avec cette valeur le résultat du MAD est toujours 0.5.

- Descripteurs Qualitatifs

Chaque descripteur de type qualitatif x_j peut prendre différentes valeurs possibles appelées modalités ($1, \dots, i, \dots, m$). Pour le traitement des descripteurs qualitatifs, chaque classe k suivant la direction j est représentée par la probabilité fréquentielle de chaque modalité du descripteur (soit $\Psi_{kj} = (\pi_{kj1}, \dots, \pi_{kji}, \dots, \pi_{kjm})$ le vecteur de fréquences), c'est-à-dire par la fréquence d'apparition de la modalité sur le nombre total d'individus. Le calcul du MAD s'effectue en considérant la présence de la modalité i dans le descripteur x_j et la fréquence de la modalité correspondante :

$$MAD_{kj} [x_j | \Psi_{kj}] = \pi_{kji} \quad \text{où } i \text{ est la modalité que possède } x_j. \quad (2.11)$$

Le type de fonction d'adéquation marginale est un paramètre à choisir par l'utilisateur.

2.3.3.2 Paramètres d'apprentissage

L'évaluation des paramètres des fonctions d'adéquation marginale est un problème statistique. Il est possible de déterminer les paramètres pour chaque classe de deux manières ; directe avec apprentissage supervisé ou de manière itérative par auto-apprentissage.

- **Apprentissage Supervisé :**

Pour l'approche directe, l'ensemble des individus d'entraînement $\{x(1), \dots, x(t), \dots, x(T_k)\}$ est considéré. La classe de chaque individu d'entraînement est assignée *a priori* par l'expert du système. T_k correspond au nombre d'individus qui appartiennent à la classe k .

Les paramètres ρ pour chaque classe sont déterminés en calculant la valeur moyenne des individus qui appartiennent à chaque classe :

$$\rho_{k_j}(T_k) = \frac{1}{T_k} \sum_{t=1}^{t=T_k} x_j(t) \quad (2.12)$$

Le paramètre σ de la fonction gaussienne peut être vu comme l'écart type d'une distribution Gaussienne, alors l'équation (2.13) donne l'estimation de la dispersion

$$\sigma_{k_j}^2(T_k) = \frac{1}{T_k - 1} \sum_{t=1}^{t=T_k} (x_j(t) - \rho_{k_j}(T_k))^2 \quad (2.13)$$

Pour la fonction binomiale centrée, le paramètre φ_{k_j} peut être déterminé en utilisant la valeur moyenne des distances entre ρ_{k_j} et chaque individu qui appartient à la classe k :

$$\varphi_{k_j}(T_k) = \frac{1}{T_k - 1} \sum_{t=1}^{t=T_k} |x_j(t) - \rho_{k_j}(T_k)| \quad (2.14)$$

Dans le cas des descripteurs qualitatifs, la probabilité fréquentielle π_{kji} pour chaque modalité i dans la classe k suivant la direction j , est calculée en fonction des modalités qui sont présentes dans les individus d'entraînement qui appartiennent à cette classe.

$$\Pi_{kji}(T_k) = \frac{1}{T_k} \sum_{t=1}^{t=T_k} \omega_j(t) \quad (2.15)$$

où $\omega_j(t) = 1$ si la modalité i est présente dans $x_{j,t}$ et $\omega_j(t) = 0$ dans le cas contraire.

- Auto-Apprentissage :

Les paramètres donnés pour l'apprentissage supervisé peuvent être calculés au fur et à mesure que chaque donnée est traitée. De cette manière, l'algorithme LAMDA crée une classe quand un objet est classé dans la NIC ou sinon l'algorithme effectue la mise à jour des classes. (voir Figure 2.3.)

L'estimation itérative du paramètre ρ_{k_j} , correspond à l'équation (2.16) :

$$\rho_{k_j}(T_k + 1) = \rho_{k_j}(T_k) + \frac{1}{T_k + 1} (x_j(T_k + 1) - \rho_{k_j}(T_k)) \quad (2.16)$$

Dans le cas des fonctions centrées, les paramètres de dispersion σ_{k_j} et φ_{k_j} ne peuvent pas être calculés itérativement parce qu'ils ont besoin de la valeur finale de ρ_{k_j} . Pour retrouver la récursivité, la moyenne arithmétique peut être remplacée par sa valeur à l'instant T , ainsi pour le calcul itératif de $\sigma_{k_j}(T_k + 1)$ et $\varphi_{k_j}(T_k + 1)$ la valeur ρ_{k_j} est remplacée dans les équations (2.13) et (2.14) par la valeur instantanée $\rho_{k_j}(T_k + 1)$.

Pour les descripteurs qualitatifs, le calcul itératif de la probabilité fréquentielle est effectué en utilisant l'équation (2.17) :

$$\Pi_{kji}(T_k + 1) = \Pi_{kji}(T_k) + \frac{1}{T_k + 1} (\omega_j(T_k + 1) - \Pi_{kji}(T_k)) \quad (2.17)$$

2.3.3.3 Degré d'Appartenance Globale – Fonction d'Agrégation Floue GAD

Un individu à classer est un vecteur $x_n = [x_{n1} \dots x_{nj} \dots x_{nd}]$, et les degrés d'adéquation marginale à la classe k correspondent à un vecteur de MADs $MAD(x_n|k) = [MAD(x_{n1}|k) \dots MAD(x_{nj}|k) \dots MAD(x_{nd}|k)]$.

En utilisant la fonction d'agrégation γ , l'adéquation globale inclut tous les degrés de l'adéquation marginale pour la classe k . Ainsi, la valeur GAD représente le degré d'appartenance de l'individu x_n à la classe k .

$$GAD(x_n|k) = \gamma(MAD(x_{n1}|k), \dots, MAD(x_{nj}|k), \dots, MAD(x_{nd}|k)) \quad (2.18)$$

Les connectifs de la logique floue sont des versions floues des connectifs de la logique binaire, en particulier, d'intersection (ET) et d'union (OU).

L'opérateur ET peut être associé aux fonctions appelées T-normes, ainsi γ peut inclure les t-normes les plus utilisées : le **produit** appelé fonction *probabiliste* et le **minimum** appelé fonction *possibiliste*.

Dans la théorie des fonctions t-normes, la fonction duale d'une t-norme $\gamma(a,b)$ est appelée une t-conorme et est donnée par : $\beta(a,b) = 1 - \gamma(1-a, 1-b)$. Cette fonction duale correspond à l'opérateur flou OU.

- La fonction duale de $\gamma(a,b) = a.b$ est $\beta(a,b) = a + b - a.b$,
- La fonction duale de $\gamma(a,b) = \min[a,b]$ est $\beta(a,b) = \max[a,b]$

En considérant seulement les fonctions d'agrégation basées sur l'opérateur logique ET, le résultat correspond à l'intersection des ensembles flous, donc les valeurs résultantes seront faibles. Au contraire, en considérant l'opérateur OR qui correspond à l'union des ensembles flous, les résultats seront plus élevés.

Avec comme objectif de disposer d'un algorithme qui fournisse une classification ni trop stricte ni trop permissive, il est nécessaire d'inclure une fonction d'agrégation qui interpole entre la t-norme et sa duale. Pour ce faire, les connectifs mixtes ont été proposés [PIE91]. Ces connectifs correspondent à une interpolation linéaire entre une t-norme et sa t-conorme: $\alpha.\gamma(a,b) + (1-\alpha).\beta(a,b)$, où $0 \leq \alpha \leq 1$. Ce type de connectif mixte est utilisé pour calculer le GAD à partir des MAD's.

$$GAD(x_n|k) = \alpha\gamma(MAD(x_1|k), \dots, MAD(x_d|k)) + (1-\alpha)\beta(MAD(x_1|k), \dots, MAD(x_d|k)) \quad (2.19)$$

Le paramètre α est appelé *exigence*. $\alpha = 1$ correspond à l'exigence maximale, dans ce cas uniquement l'opérateur ET est appliqué, et le résultat est l'intersection de ensembles flous qui est plus restrictive que l'union.

La figure 2.4 montre un exemple de la partition de l'espace des données avec deux descripteurs. La représentation a été effectuée pour les différents types de MAD et en utilisant les deux types de fonction d'agrégation présentées (probabiliste et possibiliste). Pour chaque cas, les résultats de la reconnaissance des données en quatre classes et la NIC sont présentés. Chaque point du plan a été classé dans la classe avec le degré d'appartenance le plus élevé. La valeur d'exigence a été choisie égale à $\alpha = 1$.

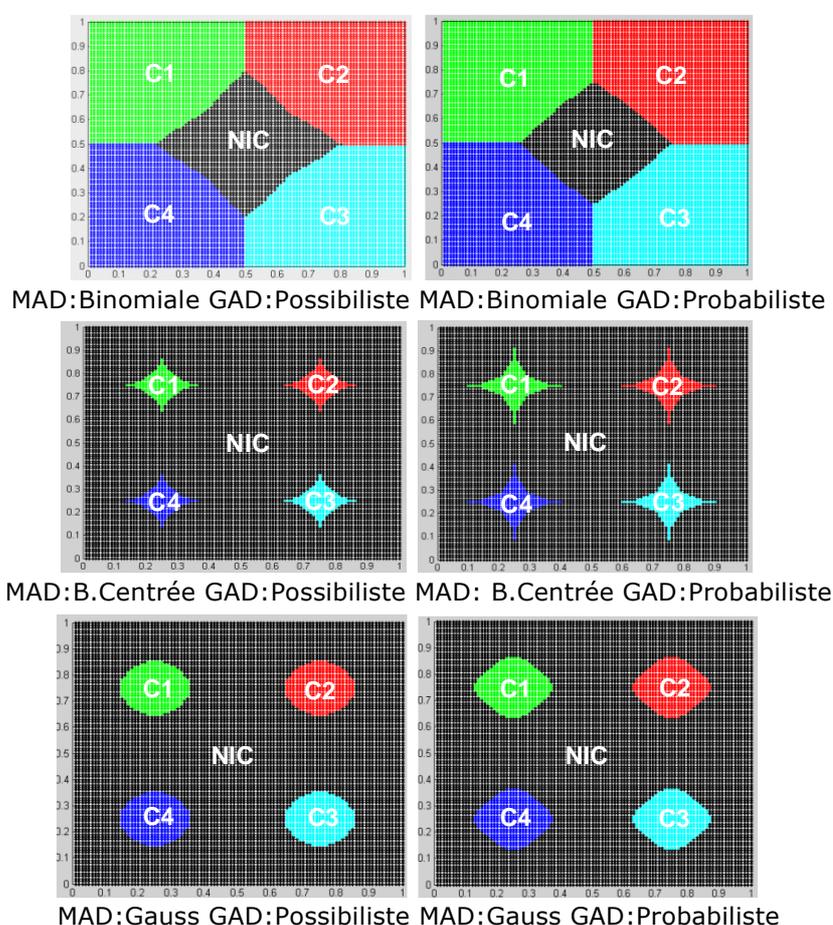


Figure 2.4. Partition de l'espace des données selon le type de fonction MAD et GAD

LAMDA est une méthodologie flexible, qui permet de s'adapter aux différentes distributions statistiques des ensembles de données. En fonction du type d'adéquation marginale choisi, il est possible d'obtenir une partition différente de l'espace des données. En fonction du paramètre d'exigence α , le nombre d'individus non reconnus en reconnaissance passive et le nombre de classes en auto-apprentissage sont croissantes. Si sa valeur est trop grande, le nombre d'individus assignés à la NIC est

élevé dans le cas d'un apprentissage supervisé, et dans le cas de l'auto-apprentissage le résultat est un nombre élevé de classes créées.

2.3.3.4 Similitude entre LAMDA et le Réseau de Neurones de type Perceptron

Si la fonction binomiale, l'exigence $\alpha=1$ et le connectif mixte probabiliste sont choisis, alors le GAD pour chaque classe correspond à :

$$GAD_k = \prod_{j=1}^d \rho_{kj}^{x_j} (1-\rho_{kj})^{1-x_j} \quad (2.20)$$

Si la fonction logarithme est appliquée à l'équation (2.20), le $\log(GAD)$ est donné par :

$$\log(GAD_k) = \sum_{j=1}^d x_j (\log(\rho_{kj}) - \log(1-\rho_{kj})) + \sum_{j=1}^d \log(1-\rho_{kj}) \quad (2.21)$$

Si on pose :

$$W_j = \log\left(\frac{\rho_{kj}}{1-\rho_{kj}}\right) \text{ et } W_0 = \sum_{j=1}^d \ln(1-\rho_{kj}) \quad (2.22)$$

alors :

$$\log(GAD_k) = \sum_{j=1}^d x_j W_j + W_0 \quad (2.23)$$

L'équation (2.23) est l'équation d'un neurone dans un réseau de type perceptron [BIS95] à deux couches. La couche d'entrée avec le nombre de neurones égal au nombre de descripteurs d et la couche de sortie avec k neurones. W_j et W_0 correspondent alors aux poids de chaque neurone qui représente la classe k . Pour une valeur $\theta_1 > \theta_2$, le $\log(\theta_1) > \log(\theta_2)$ donc l'individu est classé dans la classe qui correspond au neurone avec la valeur de sortie la plus élevée, comme ceci est effectué lors de la classification avec LAMDA.

Les valeurs des poids du perceptron sont obtenues sans itération. En fonction du type d'apprentissage (supervisé ou non), l'équation (2.12) ou l'équation (2.16) est utilisée pour déterminer la valeur ρ_{kj} .

De la même manière, on peut remarquer la similitude entre la structure de l'algorithme LAMDA et le réseau de neurones de type RBF si une fonction linéaire est choisie pour les neurones de la couche de sortie. Toutefois LAMDA possède plus de flexibilité que les réseaux neuronaux puisqu'il peut travailler avec des données mixtes et permet un auto-apprentissage. LAMDA peut combiner la reconnaissance de données avec une mise à jour constante des classes. En outre, le type d'apprentissage de LAMDA ne requiert pas d'itérations et les résultats sont facilement interprétables.

Afin d'avoir une vision générale des méthodes qui ont été présentées, les caractéristiques de chacune sont résumées dans le Tableau 2.1.

	LAMDA	D. Linéaire	CART	RBF	K-Voisins	GK Means
Type d'Apprentissage	supervisé et auto apprentissage	supervisé	supervisé	supervisé	supervisé	auto apprentissage
Nature de Données	numérique et symbolique	numérique	numérique et symbolique	numérique	numérique	numérique
Besoin du Nombre de Classes	non	oui	oui	oui	oui	oui
Paramètres à choisir	type mad, exigence α	---	critère de segmentation et d'optimisation	critère finalisation, amplitude	nombre des voisins.	m, c, e
Type de Classification	flou	strict	strict	strict	strict	flou
Mis à Jour des Classes	oui	non	non	non	non	Non

Tab. 2.1. – Principales caractéristiques des méthodes présentées.

2.4 Comparaison des méthodes

Pour évaluer la performance de LAMDA par rapport à des algorithmes de classification bien connus, nous avons fait une comparaison sur des cas d'étude (« benchmarks ») publics. Nous présentons tout d'abord 4 cas d'étude testés avec 5 techniques de classification. Pour traiter les variables qualitatives avec les méthodes qui ne le permettent pas, nous avons proposé une représentation des variables qui ne modifie pas les caractéristiques statistiques des données. Une discussion sur les méthodes de transformation de variable qualitative en variable numérique est donnée dans la deuxième partie de cette section et illustrée par un exemple.

2.4.1 Exemples

Chacune des méthodes présentées a été testée sur 4 cas d'étude reconnus dans le domaine de la fouille de données ; ils sont d'utilisation libre et disponibles dans [MER96]. Les bases de données utilisées permettent d'explorer les différents domaines d'application des méthodes de classification basées sur la fouille de données. Un des cas d'études correspond au diagnostic médical, un autre à une application dans la banque et l'autre est une application sur un système industriel. Deux des cas d'étude ont des variables mixtes (quantitatives et qualitatives).

Le Tableau 2.2. montre les caractéristiques des bases de données : "Fleur IRIS", test introduit par Fisher en 1936, le cas de données de type crédit bancaire "Australian Credits", le cas de données de type diagnostic médical "Heart Disease", le cas de

données de type détection d'états de processus industriel "Waste Water Treatment Plant".

Nom Base de Données	Individus	Descripteurs	Classes	Var. Symboliques
<i>Iris</i> , Introduit par Fisher	150	4	3	0
<i>Australian Credit</i> , Cas de données bancaires	690	14	2	8
<i>Heart Disease</i> , Cas de diagnostic médical	920	13	5	6
<i>Waste Water Treatment Plant</i> , Cas de diagnostic système industriel	527	38	13	0

Tab. 2.2. – Les 4 *benchmarks* utilisés pour la comparaison des méthodes.

Pour les méthodes qui ne permettent de traiter que des variables de type quantitatif (voir Tableau 2.1.), et dans le cas de la base de données *Heart Disease*, la représentation proposée dans la section suivante a été employée. Avec cette représentation, les caractéristiques statistiques des données ne sont pas modifiées, ce qui permet de meilleures performances par rapport à d'autres types de représentations numériques [ISA04]. Néanmoins, le nombre de descripteurs est augmenté, ce qui constitue un problème avec la présence d'un grand nombre de modalités pour un ou plusieurs descripteurs. C'est le cas de la base de données *Australian Credits*. Pour la comparaison, la représentation numérique proposée dans [MER96] a été utilisée, cette approche représente les modalités avec une échelle numérique de la variable.

2.4.1.1 Critère d'évaluation

Pour choisir les paramètres de chaque méthode (voir Tableau 2.1.), il est nécessaire de faire un ajustement de ces valeurs. En fonction de la méthode, cet ajustement est plus ou moins difficile. Les bonnes performances d'une méthode de classification dépendent fortement du choix de ces paramètres. Dans tous les cas, il est nécessaire de choisir un ensemble de données d'entraînement et d'effectuer l'ajustement paramétrique en utilisant soit l'erreur d'entraînement soit la validation des résultats par l'expert du processus. L'objectif est de minimiser l'erreur dans l'ensemble des données d'entraînement.

La méthode "10 fold cross validation" [LIM00] a été utilisée pour l'analyse de l'erreur. Cette méthode a été recommandée par différents travaux sur la comparaison de techniques de classification [MARI03] [LIM00][MIC94]. Cette méthode réalise une analyse de l'erreur de telle façon que le choix de l'ensemble de données d'apprentissage et de test n'ajoute pas de caractéristiques statistiques n'appartenant pas aux données.

Avec la méthode "10 fold cross validation", l'ensemble des données est divisé de manière aléatoire (en conservant l'ordre chronologique dans les cas où le temps est important, comme c'est le cas pour la supervision en temps réel de processus

industriels) en 10 sous-ensembles. Chaque sous-ensemble a approximativement le même nombre de données.

Pour ajuster les paramètres de chaque méthode de classification, 9 sous-ensembles ont été choisis comme données d'entraînement (90% du total). Pour chaque méthode, les paramètres sont ajustés en cherchant à diminuer l'erreur d'apprentissage sur les données d'entraînement. Avec le sous-ensemble restant (10% du total) l'erreur de reconnaissance est estimée.

Après avoir trouvé les meilleurs paramètres de l'algorithme de classification, un autre sous-ensemble est choisi comme données d'entraînement (90% du total). Un nouvel entraînement est réalisé sans changer les paramètres de la méthode. La reconnaissance est faite avec les données restantes (10% du total). Les erreurs d'entraînement et de reconnaissance sont calculées à nouveau. Ceci est répété jusqu'à ce que chacun des sous-ensembles ait été utilisé comme données de reconnaissance.

Les erreurs (voir Figure 2.5) de l'entraînement et de la reconnaissance correspondent à la moyenne des 10 erreurs obtenues avec les sous-ensembles de données d'entraînement et de test respectivement. De même, le nombre d'itérations de l'apprentissage est la moyenne sur les 10 sous-ensembles.

Le choix des paramètres peut être effectué en utilisant le premier sous-ensemble, puisqu'en divisant de manière aléatoire les données, ses caractéristiques statistiques ne sont pas altérées, et elles sont approximativement les mêmes pour les autres sous-ensembles.

Avec toutes les méthodes, une classification de type strict a été faite. Pour les méthodes de classification qui ne donnent pas une classification stricte ou dure (LAMDA, GK-Means, RBF avec des fonctions linéaires dans la couche de sortie), l'individu appartient à la classe dont le degré d'appartenance est le plus élevé.

2.4.1.2 Résultats

La figure 2.5 montre les résultats obtenus avec les 6 algorithmes, en termes de pourcentage de bonne classification. Dans le cas de LAMDA, on a utilisé les 3 fonctions d'adéquation marginale avec apprentissage supervisé et les résultats correspondent aux trois premières colonnes de la figure 2.5. La figure 2.6 montre le nombre moyen d'itérations pour les méthodes qui ont besoin d'itérer.

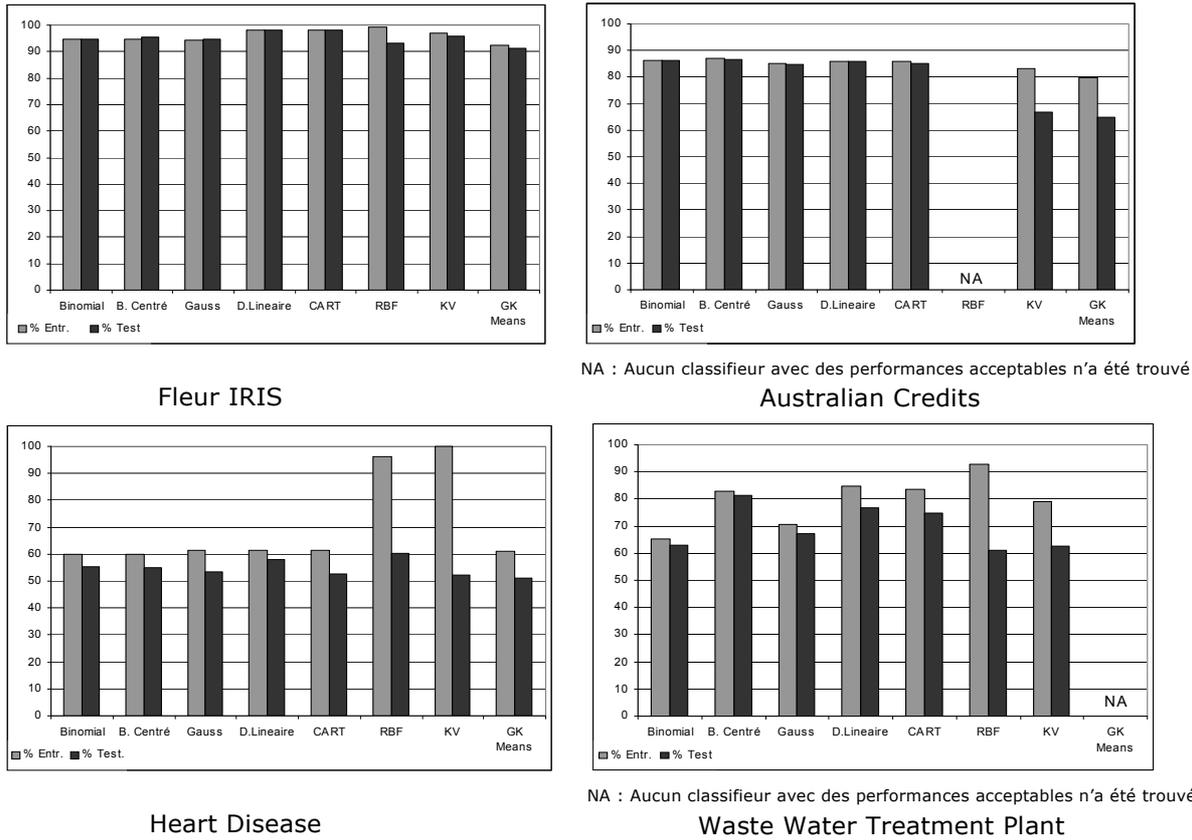


Figure 2.5. Résultats obtenus avec les différentes méthodes

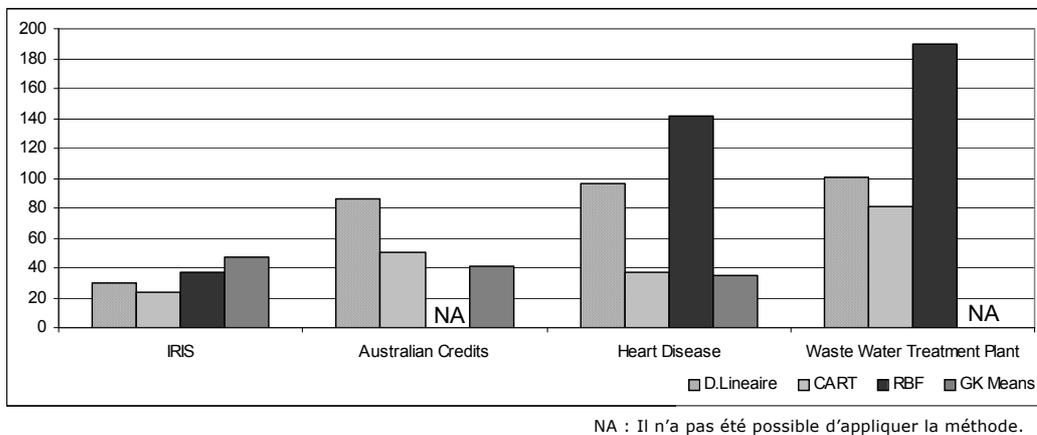


Figure 2.6. Nombre Moyen d'Itérations

Ces résultats montrent qu'il est tout-à-fait intéressant d'utiliser des méthodes basées sur la fouille des données pour résoudre des problèmes de classification. Toutefois, nous remarquons que pour obtenir de bonnes performances, il est nécessaire de faire des tests pour trouver les paramètres optimaux pour chaque cas.

Il est important de noter que pour le cas de la base de données de type risque de crédit, il n'a pas été possible d'obtenir un réseau neuronal qui ait des performances acceptables avec les données de test. De la même manière, avec la méthode GK-Means et dans le cas de *Waste Water Treatment*, les performances obtenues sur les données de test ne sont pas acceptables (situation représentée par NA sur les figures).

Dans le cas des RBF, si l'objectif est seulement d'obtenir une identification des données employées dans l'entraînement, le pourcentage de réussite est toujours de 100%. Ceci est dû au fait que l'algorithme d'apprentissage augmente le nombre de neurones de la couche cachée jusqu'à obtenir une erreur d'entraînement de 0%. Cette valeur est toujours obtenue quand il y a un neurone par individu d'entraînement. Un réseau de ce type ne permet pas une généralisation. C'est-à-dire qu'il n'a pas de bonnes performances pour la classification de nouvelles données. En conséquence, il est plus intéressant d'augmenter l'erreur d'entraînement afin d'obtenir une meilleure performance avec les données de test, et ainsi d'éviter le sur-entraînement.

La méthode GK-Means, en général présente des résultats semblables à ceux de LAMDA avec la fonction MAD de type gaussienne. Ceci est dû au fait que les classes obtenues dans les deux cas ont une forme ellipsoïdale. GK-Means montre de très bonnes performances pour le diagnostic, mais le choix des paramètres influence considérablement les résultats et il est difficile de trouver une relation permettant de traduire cette influence. C'est pour cela, que dans le cas de cette méthode, l'ajustement des paramètres est très compliqué. D'autre part, étant donné que GK-Means est une méthode d'optimisation, elle dépend fortement de l'initialisation (en principe aléatoire) des centres des classes. Bien qu'étant une méthode non supervisée, il est nécessaire de donner *a priori* le nombre de classes.

La plupart des algorithmes de classification ont besoin d'un grand nombre d'itérations, ce qui augmente le temps d'apprentissage. (voir Figure 2.6). Les RBF en général ont le plus grand nombre d'itérations, ce qui est normal pour un réseau de neurones. En revanche, les méthodes LAMDA et *K-Nearest Neighbors* n'ont pas besoin d'itération. C'est un avantage pour l'apprentissage, parce que ceci permet d'effectuer l'apprentissage avec différents paramètres et de trouver ceux qui permettent d'obtenir les meilleures performances.

LAMDA montre des performances semblables à celles des autres méthodes. Comme avantage, elle permet une interprétation facile des résultats, et l'algorithme d'apprentissage est très rapide car il n'est pas basé sur la minimisation de l'erreur et donc des résultats sont obtenus dès la première itération. LAMDA permet d'extraire la connaissance en utilisant des données d'entraînement, mais permet aussi d'ajouter de nouvelles classes non prévues *a priori*. Cette caractéristique permet de diminuer la dépendance aux données d'apprentissage qui est commune à toutes ces méthodes. L'interprétation aisée des paramètres à choisir permet d'avoir une méthode d'apprentissage plus facile à utiliser dans un système complet de diagnostic, qui va peut-être être mis en œuvre par l'expert du processus, non spécialiste des méthodes de classification.

2.4.2 Étude particulière des données qualitatives sur un exemple.

Nous présentons dans ce qui suit un exemple de classification de données incluant des descripteurs qualitatifs. Trois méthodes de classification ont été testées, deux méthodes sont de type flou (LAMDA et FKM) et l'autre (*K-Means*) effectue une partition

stricte. Les méthodes FKM et *K-Means* sont des méthodes basées sur le calcul de distances (Voir Section 2.3.2.1). Au contraire de LAMDA, les algorithmes FKM et *K-Means* ne permettent pas de travailler directement avec des descripteurs de type qualitatif. Cet exemple nous permet de présenter une représentation ou transformation des variables qualitatives en variables quantitatives qui ne modifie pas les caractéristiques statistiques des données. Cette représentation a été reprise dans l'exemple précédent.

En ce qui concerne la transformation, dans le cas des descripteurs dont les modalités ont un ordre naturel (e.g. haut, bas, etc.) une échelle numérique a été employée. Dans le cas des descripteurs où il n'y a aucune notion d'ordre (e.g. couleur, forme, etc.), la représentation numérique proposée transforme chaque modalité des variables qualitatives en un nouveau descripteur. La présence de la modalité dans l'individu à classer est représentée avec un 1, et 0 est utilisé dans le cas contraire.

D'ailleurs, cet exemple nous permet de mettre en évidence les caractéristiques que possède LAMDA pour classer des données qualitatives. En raison de la capacité de LAMDA de pouvoir classer des données de type mixte (qualitatif/quantitatif), deux approches ont été considérées pour cet algorithme : soit le type qualitatif de tous les descripteurs est conservé, soit ils sont transformés en descripteurs quantitatifs. Une comparaison des résultats obtenus suivant les deux approches est présentée dans ce qui suit.

Pour illustrer le traitement de données qualitatives de manière facilement interprétable, un petit exemple a été choisi. Il a été précédemment utilisé pour analyser la génération de règles dans des arbres de décision (cours dispensé au département de *Computer Science* de l'université de Regina au Canada [HAM03]).

L'exemple consiste à classer 8 individus en deux groupes (classes), ceux qui risquent le coup de soleil et ceux qui ne le risque pas. Chaque individu est caractérisé par quatre attributs de type qualitatif, deux représentent les possibles facteurs du coup de soleil et les deux autres sont ajoutés pour établir la capacité des algorithmes à ne pas tenir compte d'information superflue pour la classification. Les données sont présentées dans le Tableau 2.3. Les descripteurs avec leurs modalités respectives sont :

Cheveux: Blonds, Bruns et Roux
Taille: Grande, Moyenne et Petite
Poids: Elevé, Moyen et Faible
Application Crème Solaire: Oui, Non

Nom	Cheveux	Taille	Poids	Crème Solaire	Résultat
Sarah	Blonds	Moyenne	Faible	Non	Coup de Soleil (1)
Dana	Blonds	Grande	Moyen	Oui	Non (2)
Alex	Bruns	Petite	Moyen	Oui	Non (2)
Annie	Blonds	Petite	Moyen	Non	Coup de Soleil (1)
Emily	Roux	Moyenne	Elevé	Non	Coup de Soleil (1)
Pete	Bruns	Grande	Elevé	Non	Non (2)
John	Bruns	Moyenne	Elevé	Non	Non (2)
Katie	Blonds	Petite	Faible	Oui	Non (2)

Tab. 2.3. – Description des Individus.

Les données transformées en descripteurs quantitatifs sont présentées dans le Tableau 2.4. La représentation des variables qualitatives en variables quantitatives telle qu'elle a été décrite précédemment, a été employée. Ainsi pour la taille et les attributs de poids, l'échelle de $\{0, \frac{1}{2}, 1\}$ a été utilisée. Pour l'attribut « application de crème solaire », une association binaire a été faite et dans le cas de la couleur des cheveux, le descripteur a été transformé en trois descripteurs (Blonds, Bruns et Roux) ; la présence de la modalité dans l'individu est représentée par un 1 dans le descripteur correspondant et 0 pour les deux autres.

Individu	Blonds	Bruns	Roux	Taille	Poids	Crème Solaire	Classe
1	1	0	0	0.5	0	0	1
2	1	0	0	1	0.5	1	2
3	0	1	0	0	0.5	1	2
4	1	0	0	0	0.5	0	1
5	0	0	1	0.5	1	0	1
6	0	1	0	1	1	0	2
7	0	1	0	0.5	1	0	2
8	1	0	0	0	0	1	2

Tab. 2.4. – Représentation quantitative des Individus.

2.4.2.1 Résultats

Résultats obtenus avec l'algorithme FCM

Le Tableau 2.5. présente les résultats en termes de pourcentage d'éléments correctement classés pour différents degrés de floue (*fuzziness*) m (voir Section 2.3.2.1). Compte tenu que cet algorithme effectue une classification de type flou, l'individu appartient à la classe dont le degré d'appartenance est le plus élevé. La référence utilisée pour estimer l'erreur de classement est la classification présentée dans le Tableau 2.4.

Degré de Floue (m)	% de Succès
100	87.5
70	75
50	62.5
25	87.5
2	62.5

Tab. 2.5. – Pourcentages de succès obtenus avec FCM

Il est important de remarquer que cet algorithme, même avec les mêmes paramètres d'initialisation, n'assure pas l'obtention de classifications identiques pour différents tests.

Résultats obtenus avec l'algorithme K-Means

Cet algorithme demande le choix du type de distance employé pour établir la fonction de minimisation, ainsi que le critère d'arrêt.

Nous avons testé deux approches pour arrêter l'algorithme. Les deux portent sur l'arrêt pour un nombre maximal d'itérations : dans la première approche, à chaque itération, de nouveaux centres des classes sont aléatoirement initialisés. La deuxième approche n'utilise qu'une seule initialisation des centres de classe. Cette dernière approche a fourni les meilleurs résultats, résultats donnés dans le Tableau 2.6.

Type de Distance	% de Succès
Distance Euclidienne Carrée	62.5
Somme des Différences Absolues	75
1 – Cosinus de l'angle entre Individus	75
1 – Corrélation des Données	75

Tab. 2.6. – Pourcentage de succès obtenus avec K-Means (Deuxième approche).

Il est à noter que 62,5% de succès c'est le minimum que nous puissions obtenir car c'est ce qui se passerait si tous les éléments étaient affectés à une seule classe, la classe 2.

Comme nous l'avons expliqué dans l'exemple précédent (voir Section 2.4.1), les algorithmes de coalescence (y compris les K-means et FKM bien entendu) montrent d'importantes variabilités des résultats, liés à l'initialisation aléatoire des centres (v_k) qui est nécessaire pour démarrer l'algorithme. Il est possible de diminuer cette variabilité en augmentant le nombre d'itérations.

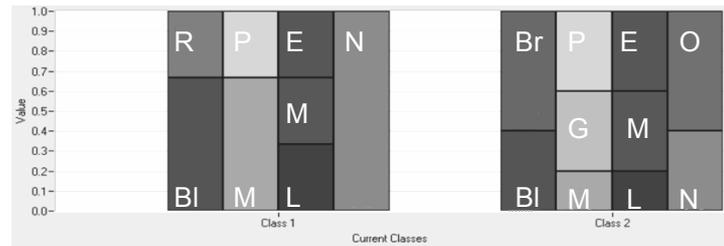
Résultats obtenus avec l'algorithme LAMDA

Les résultats de la classification des 8 individus avec des attributs qualitatifs (Tableau 2.3.) sont donnés dans le Tableau 2.7. Ils ont été obtenus en utilisant la fonction MAD binomiale (Section 2.3.3.1) et différentes valeurs de l'exigence (α).

Exigence	Succès %
$0.1 < \alpha < 0.6$	75
0.8	87.5
1	100

Tab. 2.7. – Pourcentage de Succès avec LAMDA (Cas Qualitatif)

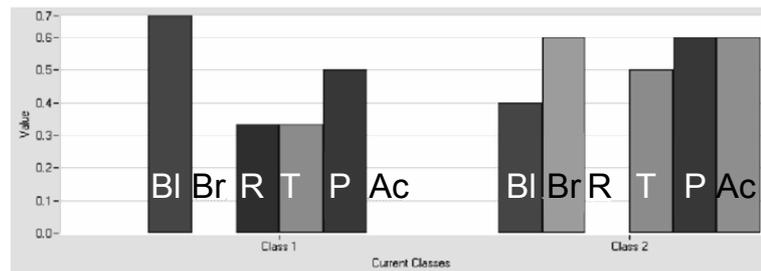
La Figure 2.7 illustre les profils des classes, qui correspondent aux valeurs de fréquence des modalités pour chaque descripteur. Cette caractéristique de LAMDA facilite l'interprétation des classes tout spécialement quand les variables sont de type qualitatif. A partir des profils des classes, on peut avoir une idée des facteurs les plus influents pour chaque classe. Dans le cas des caractéristiques peu importantes pour l'estimation de la classe, la distribution des modalités est semblable pour les deux classes, néanmoins pour avoir une caractérisation adéquate des classes, il est important d'avoir un nombre élevé d'exemples pour trouver tous les attributs dans toutes les classes. Pour cet exemple, il manque en effet la représentation d'un individu avec des coups de soleil et de grande taille.



R:Roux, BI:Blonds, Br:Bruns, P:Petite, M:Moyenne,G:Grande,E:Elevé,M:Moyen,L:Léger,O:Oui,N:Non

Figure 2.7. Profil des Classes obtenue avec LAMDA (Cas Qualitatif)

En utilisant la représentation quantitative, toutes les modalités sont représentées par des descripteurs quantitatifs binaires séparés. Les pourcentages de succès obtenus avec la fonction binomiale et pour différents indices d'exigence (voir Tableau 2.7.) sont de 100%. Il est à noter que la dimension de l'espace (donnée par le nombre de descripteurs) a été augmentée, elle était de 4 dans la première approche et de 6 dans celle-ci. La figure 2.8 donne le profil des classes ; chaque barre correspond à la valeur moyenne pour chaque descripteur des deux classes.



R:Roux, BI:Blonds, Br:Bruns, T:Taille, P:Poids,Ac: Application Crème Solaire

Figure 2.8. Profil des Classes obtenu avec LAMDA (Cas Quantitatif)

LAMDA présente des avantages pour le traitement de données qualitatives et ses performances sont semblables à celles obtenues avec la représentation quantitative proposée.

Une caractéristique très intéressante de LAMDA est l'interprétation qui peut être facilement effectuée à partir des profils des classes. Par exemple, les profils obtenus indiquent quel est le descripteur qui peut le plus influencer le coup de soleil. Dans cet exemple, la modalité « brun » pour des cheveux et la modalité « OUI » pour l'application de lotion n'apparaissent pas dans la classe 1 (figure 2.8), donc les individus qui n'ont pas employé de crème et qui sont blonds ou roux, auront beaucoup plus de chance d'attraper des coups de soleil. Dans le cas qualitatif, on trouve le même comportement, les modalités NON et Brun ne sont pas dans la classe 1 (coup de soleil). Il est clair que le facteur qui affecte la plupart du temps le coup de soleil est l'attribut relatif à l'application de lotion.

De manière similaire au cas qualitatif pur, la transformation quantitative proposée permet d'obtenir des résultats de classification avec 100% de succès. La transformation proposée peut être donc considérée comme une alternative viable pour les algorithmes qui ne permettent pas de travailler avec des données qualitatives. Néanmoins, en convertissant les modalités en descripteurs quantitatifs, la dimension de l'espace est augmentée, ce qui peut représenter un problème pour un nombre élevé de modalités, car cela peut conduire à augmenter fortement le temps de calcul de l'algorithme.

2.5 Conclusion

Les méthodes qui ont été comparées sont considérées adéquates pour faire le diagnostic automatique des systèmes quand il est considéré comme la différenciation d'états en prenant en compte les états de défaillance.

En pratique, les différences de performances entre les méthodes sont probablement insignifiantes. Parfois obtenir un petit accroissement en performance demande beaucoup de temps, et en pratique n'est pas justifié. Pour la plupart des applications où le temps est un facteur important, il est peut-être plus intéressant d'utiliser un des algorithmes les plus rapides et les plus simples.

Les méthodes floues de reconnaissance de formes ont montré des performances satisfaisantes par rapport celles obtenues avec des méthodes de partition stricte. De plus, il est très intéressant dans le cas du diagnostic de posséder les appartenances d'un individu à plusieurs classes (représentant les états du système). Cette caractéristique permet de mieux connaître la situation réelle du système et prévoir des changements vers des états de défaillance en fonction du degré d'appartenance de l'individu actuel aux classes de défauts.

La méthode de classification floue LAMDA a été resituée dans le contexte des méthodes de reconnaissance de formes. Cet algorithme est très flexible, il permet de travailler avec des données de type qualitatif, quantitatif et mixte. La méthode peut être considérée comme faisant partie des méthodes de "Machine Learning", car l'interprétation des résultats et du choix des paramètres est très simple et directe. L'apprentissage est basé sur des principes possibilistes et pourtant la méthode peut être aussi considérée comme statistique tout en ne supposant pas une distribution des données *a priori*.

LAMDA montre des performances très semblables à celles des autres méthodes reconnues de classification basées sur la fouille de données. L'avantage que LAMDA présente est qu'il ne requiert pas d'itération pour l'entraînement et qu'il permet l'ajout de classes non prévues dans l'étape d'entraînement ainsi que l'adaptation des classes. Le choix des paramètres est intuitif, ce qui peut être considéré comme très utile pour un système de diagnostic développé par un expert du processus non spécialiste des méthodes de classification.

D'ailleurs, le traitement effectué par LAMDA pour les données qualitatives ou mixtes, permet d'avoir des profils de classes utiles pour l'interprétation des classes obtenues en relation directe avec les variables qui décrivent ces données. Le profil des classes permet d'estimer la contribution de chaque mesure (descripteur) à partir de sa fréquence dans le cas qualitatif ou de sa valeur moyenne dans le cas quantitatif.

Il est évident qu'avec les techniques de reconnaissance de formes, il existe une forte dépendance aux données d'entraînement et aux paramètres de chaque algorithme. D'ailleurs, en général le choix du nombre de classes est critique dans les cas où on n'a pas une idée *a priori* de la distribution de l'espace des données. Toutefois, avec des paramètres adéquats, bien que non optimaux, il est possible d'avoir une classification initiale avec plusieurs classes qui peut être améliorée ensuite.

En conséquence, nous considérons particulièrement intéressant de proposer une technique applicable aux méthodes de classification du type flou en général, qui permette d'améliorer automatiquement la partition initiale obtenue. Notre objectif dans cette thèse est de proposer une méthode pour obtenir la meilleure partition possible en termes de compacité et de séparation des classes. Les mesures que nous considérons nécessaires pour estimer la qualité d'une partition sont présentées dans le chapitre suivant et la méthode proposée pour l'amélioration est présentée dans le chapitre 3. Pour pouvoir implémenter cette amélioration dans les différents algorithmes de classification floue, les mesures proposées sont basées uniquement sur le résultat commun délivré par les méthodes de classification floue, à savoir la matrice de partition obtenue par une classification initiale. Cette proposition permet d'obtenir un meilleur entraînement quand le nombre des classes est inconnu.

Chapitre 3

3. CRITERES D'EVALUATION DES PARTITIONS DE L'ESPACE DE DONNEES

3.1 Introduction

Dans le premier chapitre nous avons présenté la théorie du diagnostic basée sur la fouille de données et la possibilité d'établir un classificateur qui soit utile pour identifier les états fonctionnels d'un procédé. Dans le chapitre deux, une comparaison des méthodes de classification basées sur la fouille de données, met en évidence, que la performance de ces méthodes est hautement dépendante des paramètres choisis par l'utilisateur. D'autre part, en utilisant une technique non supervisée, en particulier s'il n'y a pas une connaissance préalable sur la distribution ni sur le nombre des classes, il est nécessaire de proposer une évaluation de la partition obtenue qui conduira à l'amélioration automatique et permettra d'obtenir une partition finale qui représente le plus correctement possible les états du processus. C'est pourquoi l'objectif principal de cette thèse est de proposer une méthode qui permette d'améliorer automatiquement les partitions, et qui constitue une aide à l'expert pour pouvoir identifier les états de fonctionnement et établir ainsi le modèle comportemental du procédé.

Nous nous sommes intéressés aux méthodes de classification floues. Les méthodes de classification floue, d'une part ont montré des performances très satisfaisantes et d'autre part elles fournissent aussi les appartenances d'un individu à toutes les classes qui représentent les états du système, et elles permettent ainsi de mieux connaître la situation réelle du système et de prévoir de possibles changements vers des états de défaillance

Avant même de vouloir améliorer une partition, il faut d'abord évaluer sa qualité. Ce chapitre est dédié à la description des critères (caractéristiques) qui permettent d'avoir une partition de l'espace de données interprétable (avec des classes bien

définies), utile pour le diagnostic et pour la prise de décisions dans le cadre de la surveillance. Nous proposons des mesures de distance et de dispersion des classes qui nous permettent de définir un nouvel indice, indépendant de la nature des données, pour évaluer la qualité des partitions. L'indice est calculé uniquement en fonction des degrés d'appartenance qui sont obtenus par toutes les méthodes de classification floues, par conséquent cet indice est indépendant de la méthode de classification choisie.

L'indice de qualité que nous proposons, constitue la base de l'algorithme d'amélioration des partitions que nous présenterons dans la dernière section.

Les applications que nous présenterons dans le dernier chapitre utilisent la méthode LAMDA comme méthode de classification floue, mais elle pourrait être remplacée par toute autre méthode sans avoir à modifier l'algorithme d'amélioration des partitions.

3.2 Critères pour évaluer une partition de l'espace des données

Généralement, la question de la validité de partitions est complexe et implique beaucoup d'interrogations qui dépendent de l'application. Dans le cadre de la vision artificielle, généralement les mesures de validité sont basées sur la séparation et la cohésion des classes [JAI88] [BEZ81]. Dans la plupart de travaux, les critères qui sont considérés importants pour évaluer une partition, sont : le nombre de classes, la séparation et la dispersion des classes [XIE91][FRA02] [GAT89] [KAY95][KAY02] [WAN05].

Dans le cadre du diagnostic, nous proposons de chercher une partition avec des classes compactes et séparées, mais pour éviter la solution triviale qui propose une classe par élément, il faut tenir compte du nombre de classes.

En recherchant des classes floues compactes, l'objectif est d'avoir des groupes avec de hauts degrés d'appartenance pour les éléments qui sont semblables aux prototypes des classes et des degrés d'appartenance faibles dans les autres cas. Cette caractéristique est utile dans le diagnostic parce qu'elle permet d'avoir une plus grande certitude sur la décision de classification, donc sur l'identification de l'état fonctionnel du système, sans toutefois perdre la connaissance des appartenances aux autres états possibles du système.

La recherche de classes séparées facilite l'identification des états en général, et améliore, en particulier, la détection des comportements anormaux du procédé. En général, les états normaux d'un procédé suivent une certaine tendance et les groupes de données qui représentent ces états sont proches entre eux. Un comportement hors tendance peut représenter une défaillance du système.

Si le nombre de classes dans l'évaluation d'une partition n'était pas pris en compte, la solution optimale correspondrait à la partition avec un seul individu par classe, ou des classes avec des individus absolument identiques. De plus, le fait de conserver un nombre élevé de classes génère un modèle comportemental inutilement complexe. En revanche, si on se contentait d'un nombre très faible de classe, le modèle serait incomplet ou peu précis [KAY95].

En conséquence, nous voulons souligner qu'un indice de qualité de l'ensemble des classes doit tenir compte des 3 caractéristiques ; privilégier l'une plus que l'autre ne permet pas d'avoir une meilleure partition. Les indices proposés dans la littérature prennent en compte ces aspects, mais, soit ils sont basés sur les caractéristiques géométriques des classes en restreignant leur usage à quelques algorithmes de classification (Section 3.3), soit ils dépendent des valeurs des données ou de valeurs autres que les degrés d'appartenance.

Dans la suite, nous présentons des mesures pour évaluer chacune des caractéristiques des partitions. Ces mesures sont ensuite utilisées dans l'indice que nous proposons pour évaluer la qualité des partitions. Dans les sections suivantes, d'abord nous discutons les approches les plus fréquemment présentées dans la littérature et nous exposons les raisons pour lesquelles nous ne les utilisons pas telles quelles, ensuite nous présentons la nouvelle approche que nous proposons.

3.3 Mesures pour évaluer les critères de qualité des partitions.

Le problème d'évaluation de la qualité des classes et de la détermination du nombre de classes pour un ensemble de données se rencontre de façon naturelle dans les méthodes de coalescence (Chapitre 2). Ces méthodes sont de type non supervisé et montrent de très bonnes performances, mais il est toujours nécessaire d'établir *a priori* le nombre de classes et de plus elles sont très sensibles aux nombreux paramètres que doit choisir l'utilisateur, spécialement lorsque le nombre de classes choisi n'est pas correct [BEZ98] [KAY02].

Plusieurs mesures pour évaluer les caractéristiques des classes complètent les algorithmes de groupement basés sur des distances; elles utilisent pour ceci, les caractéristiques géométriques des classes [KRI92] [KRI00]. Ces caractéristiques sont basées sur les concepts d'hypervolume et de densité des classes. Les valeurs sont calculées à partir de la matrice de covariance floue de chaque groupe (F_k , Chapitre 2). On peut citer: *The Fuzzy Hyper Volume*, *The Total Fuzzy Hyper Volume* et *The Average Partition Density*.

Dans le but de proposer un indice de qualité de partitions applicable à toutes les méthodes de classification floue quelles qu'elles soient, on doit abandonner ces mesures, parce qu'elles sont basées sur les caractéristiques de la matrice de covariance

floue des classes (F , Chapitre 2), qui est un résultat particulier de la méthode de coalescence choisie, et en outre dépend des valeurs des données.

Nous interprétons chaque classe comme un ensemble flou défini par les degrés d'appartenance de chaque donnée à chaque groupe (voir Figure 3.1). La totalité des ensembles flous, correspond à la totalité des classes qui ont été trouvées avec l'algorithme de classification.

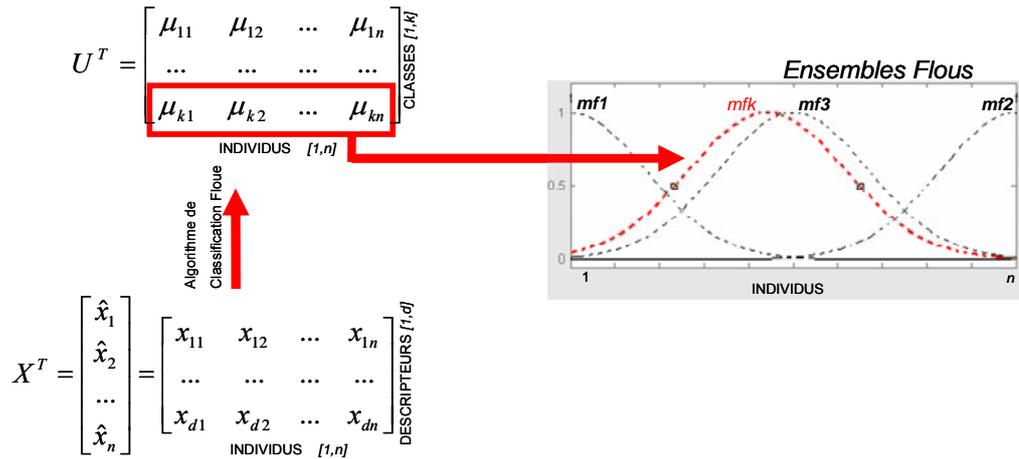


Figure 3.1. Représentation des classes comme ensembles flous

Pour le calcul de dispersion des classes floues, des mesures basées sur les degrés d'appartenance et sur les moyennes des données ont été proposées [XIE91] [FRA02]. Un des indices de qualité des partitions floues le plus reconnu a été proposé par Xie L. et Xuanli B. [XIE91]. Cet indice inclut l'analyse de dispersion en utilisant "the compactness fuzzy c-partition value Π ":

$$\Pi = \frac{\sigma}{N} \quad \text{où} \quad \sigma = \sum_{k=1}^K \sum_{n=1}^N (\mu_{kn} \|x_n - v_k\|)^2 \quad (3.1)$$

La valeur Π mesure la compacité de la partition ; une valeur basse correspond à une partition avec des classes compactes. μ_{kn} est le degré d'appartenance de x_n à k et $\|\bullet\|$ la norme euclidienne. $\mu_{ij} \|x_j - v_i\|$ peut donc être interprétée comme la "déviation floue" ou « l'écart flou » (*the fuzzy deviation*) de l'individu x_n par rapport au centre v_k de la classe k .

Un autre indice de qualité qui dénote des performances similaires à celui de Xie L. et Xuanli B., est l'indice proposé par Franco et al. [FRA02]. La mesure de dispersion, dans ce cas, est basée sur les matrices de dispersion de Fisher R. [FIS36], mais en incluant le degré d'appartenance pour pouvoir traiter des cas issus de la logique floue.

En utilisant les concepts de Fisher R. pour évaluer la compacité et la séparation des classes non floues, Franco et al. [FRA02], définit la mesure associée à la dispersion de la classe k et permet de mesurer la position des centres des classes (eq. 3.2.).

$$sbe_k = \sum_{n=1}^N \mu_{kn} \cdot (m_{ke} - m) \cdot (m_{ke} - m)^T \quad (3.2)$$

où N est le nombre de données d'entraînement, m_{ke} est le prototype de la classe k et m le centre de l'ensemble des données (eq. 3.3.).

$$m_{ek} = \frac{\sum_{n=1}^N \mu_{kn} \cdot x_n}{\sum_{n=1}^N \mu_{kn}} \quad \text{et} \quad m = \frac{1}{N} \cdot \sum_{n=1}^N x_n \quad (3.3)$$

μ_{kn} est le degré d'appartenance de la donnée n (vecteur x_n) à la classe k . La somme des dispersions de toutes les classes permet de mesurer la dispersion de la partition.

Clairement, ces mesures utilisent les valeurs des descripteurs des données, ce qui représente un coût de calcul élevé lorsque les données ont de nombreux descripteurs pour l'entraînement, ce qui est très fréquent. Nous cherchons à être indépendants des valeurs des données, non seulement à cause du coût de calcul, mais aussi pour pouvoir employer des techniques qui permettent de classer directement des données de type qualitatif (e.g. LAMDA, Chapitre 2) sans passer par une représentation quantitative.

Concernant la mesure de distance entre classes, cette mesure est normalement estimée entre les centres des classes ($\|v_i - v_j\|, \forall i, j : [1, K]$). Comme la plupart des travaux de classification ont été faits pour le cas de données de type quantitatif, la distance est mesurée en utilisant des mesures classiques comme la distance euclidienne ou celle de Mahalanobis.

Pour notre approche, nous avons besoin de travailler avec une mesure de distance entre classes qui soit uniquement dépendante des degrés d'appartenance. C'est possible si on représente chaque classe comme un ensemble flou discret (voir Figure 3.1). Dans cette représentation, les K colonnes de la matrice de partition floue U qui a été obtenue comme résultat de la partition de l'ensemble des données X en K classes, sont interprétées comme K ensembles flous discrets, définis dans l'univers de discours $[1, N]$, où N est le nombre d'individus classés. De cette manière, nous pouvons établir une mesure de distance entre classes comme une mesure de distance entre ensembles flous.

Cette représentation nous permet de travailler avec la théorie des ensembles flous pour établir les caractéristiques des partitions, sans tenir compte des données ni avoir besoin de connaître la structure géométrique des classes. Ces caractéristiques sont

particulièrement utiles lorsque les données ont des composantes qualitatives pour lesquelles il est difficile d'associer une représentation.

Nous présentons dans ce qui suit, les mesures de dispersion et de distance qui seront employées pour la construction d'un indice de qualité des partitions.

3.3.1 Mesure de distance entre classes

Pour évaluer la distance entre deux classes floues (A et B), nous utilisons la représentation proposée sur la Figure 3.1. Nous pouvons définir pour ceci, une mesure qui permet d'établir la distance entre des ensembles flous définis sur le même univers de discours [ISA06].

Nous voulons inclure dans la mesure de distance, la possibilité d'analyser la dissimilitude entre classes. Pour ce faire, nous proposons de remplacer les mesures classiques de distance entre deux ensembles flous présentées dans le Tableau 3.1 [MART02], par la mesure de distance ultra-métrique de l'équation 3.4.

Nom	
Hamming	$d(A, B) = \mu_A - \mu_B $
Euclidienne	$d(A, B) = (\sum (\mu_A - \mu_B)^2)^{1/2}$
Minkowski	$d(A, B) = (\sum (\mu_A - \mu_B)^w)^{1/w}$ avec $w \in [1, \infty]$

Tab. 3.1 – Mesures classiques de distance entre ensembles flous.

Les mesures classiques permettent d'estimer la distance entre classes mais ne prennent pas en compte la similitude entre elles.

THÉORÈME : La relation (3.4.) définit une mesure de distance entre deux sous ensembles flous A et B .

$$d^*(A, B) = 1 - \frac{M[A \cap B]}{M[A \cup B]} = 1 - \frac{\sum_{n=1}^N (\mu_{A_n} \cap \mu_{B_n})}{\sum_{n=1}^N (\mu_{A_n} \cup \mu_{B_n})} \quad \text{où} \quad (3.4)$$

$$\mu_{A_n} \cap \mu_{B_n} = \min(\mu_{A_n}, \mu_{B_n})$$

$$\mu_{A_n} \cup \mu_{B_n} = \max(\mu_{A_n}, \mu_{B_n})$$

$d^*(A, B)$ est une ultra-métrique dans l'ensemble $\mathcal{P}(X)$ des parties floues de l'ensemble des données X .

Démonstration : Il s'agit de prouver que les propriétés des ultra-métriques sont satisfaites :

1. Propriété de Symétrie :

$$\forall_{A,B} \in \mathcal{P}(X) : d^*(A,B) = d^*(B,A) \quad (3.5)$$

En représentant $d^*(A,B)$ comme $1 - \frac{M[Tnorme(A,B)]}{M[Snorme(A,B)]}$ et en tenant compte que les $Tnormes$ et les $Snormes$ remplissent la propriété de commutativité, $Tnorme(A,B) = Tnorme(B,A)$ et $Snorme(A,B) = Snorme(B,A)$:

$$1 - \frac{M[A \cap B]}{M[A \cup B]} = 1 - \frac{M[B \cap A]}{M[B \cup A]} \quad (3.6)$$

2. Propriété de Séparation :

$$\forall_{A,B} \in \mathcal{P}(X) : d^*(A,B) = 0 \Leftrightarrow A = B \quad (3.7)$$

En tenant compte que :

$$1 - \frac{M[A \cap B]}{M[A \cup B]} = 0 \Leftrightarrow M[A \cap B] = M[A \cup B] \quad (3.8)$$

et en prenant en compte les axiomes des $Tnormes$ et des $Snormes$, $Snorme(A,B) = Tnorme(A,B) \Leftrightarrow A = B$, $d^*(A,B)$ remplit par conséquent la propriété de Séparation.

3. Propriété d'Inégalité Triangulaire:

$$\forall_{A,B,C} \in \mathcal{P}(X) : d^*(A,C) \leq \max(d^*(A,B), d^*(B,C)) \quad (3.9)$$

En remplaçant la distance $d^*(A,B)$ par la définition donnée en (3.4), l'inégalité triangulaire devient:

$$1 - \frac{M[A \cap C]}{M[A \cup C]} \leq \max\left(1 - \frac{M[A \cap B]}{M[A \cup B]}, 1 - \frac{M[B \cap C]}{M[B \cup C]}\right) \quad \text{ou} \quad (3.10)$$

$$\frac{M[A \cap C]}{M[A \cup C]} \geq \min\left(\frac{M[A \cap B]}{M[A \cup B]}, \frac{M[B \cap C]}{M[B \cup C]}\right)$$

On doit considérer les deux hypothèses complémentaires uniques H1 et H2 :

Hypothèse H1 :

$$\frac{M[A \cap B]}{M[A \cup B]} \leq \frac{M[B \cap C]}{M[B \cup C]} \quad (3.11)$$

D'après la condition d'inégalité triangulaire et l'équation (3.9) on a :

$$\frac{M[A \cap C]}{M[A \cup C]} \geq \frac{M[A \cap B]}{M[A \cup B]} \quad (3.12)$$

En tenant compte que:

$$M[Tnorme(X,Y)] + M[Snorme(X,Y)] = M[X] + M[Y] \quad \text{où } X, Y \text{ et } Z \text{ sont ensembles flous} \quad (3.13)$$

et en utilisant la propriété (3.13), l'équation (3.12) se transforme en :

$$\frac{M[A]+M[C]}{M[Snorme(A,C)]} \geq \frac{M[A]+M[B]}{M[Snorme(A,B)]} \quad (3.14)$$

D'après la propriété (3.13) et l'équation 3.14 on a :

$$\begin{aligned} M[Snorme(A,C)] &\geq M[A]+M[C] \quad \text{et} \\ (M[A]+M[C]) M[Snorme(A,B)] &\geq (M[A]+M[B]) M[Snorme(A,C)] \end{aligned} \quad (3.15)$$

Alors,

$$(M[A]+M[C]) M[Snorme(A,B)] \geq (M[A]+M[B]) M[Snorme(A,C)] \geq (M[A]+M[B])(M[A]+M[C])$$

Donc :

$$(M[A]+M[C]) M[Snorme(A,B)] \geq (M[A]+M[B])(M[A]+M[C]) \quad (3.16)$$

En satisfaisant la condition :

$$M[Snorme(A,B)] \geq (M[A]+M[B]) \quad (3.17)$$

Hypothèse H2 :

$$\frac{M[A \cap B]}{M[A \cup B]} \geq \frac{M[B \cap C]}{M[B \cup C]} \quad (3.18)$$

En utilisant la propriété (3.13), l'équation (3.18) se transforme en :

$$\frac{M[A]+M[C]}{M[Snorme(A,C)]} \geq \frac{M[B]+M[C]}{M[Snorme(B,C)]} \quad (3.19)$$

Et de manière similaire au cas H1 :

$$\begin{aligned} M[Snorme(A,C)] &\geq M[A]+M[C] \quad \text{et} \\ (M[A]+M[C]) M[Snorme(B,C)] &\geq (M[B]+M[C]) M[Snorme(A,C)] \end{aligned} \quad (3.20)$$

Alors,

$$(M[A]+M[C]) M[Snorme(B,C)] \geq (M[B]+M[C]) M[Snorme(A,C)] \geq (M[B]+M[C])(M[A]+M[C])$$

La condition est aussi satisfaite :

$$M[Snorme(B,C)] \geq (M[B]+M[C]) \quad (3.21)$$

Par conséquent, la mesure proposée est une mesure ultra-métrique.

De plus, il est facile de constater que si A et B sont des ensembles disjoints la valeur de $d^*(A,B)$ est 1 et si $A=B$ la valeur de $d^*(A,B)$ est 0. On a donc une propriété utile de cette distance: il s'agit d'une distance normalisée dans l'intervalle unitaire, puisqu'il est facile de constater que $0 \leq d^*(A,B) \leq 1$.

Dans le chapitre suivant, nous présentons un exemple académique qui montre que cette mesure de distance (eq. 3.4) a le même comportement que la distance euclidienne entre classes, mais elle a l'avantage d'être applicable pour tout classificateur flou.

Le complément de la mesure de distance nous permet d'établir la similitude entre classes. La mesure de similitude $G(A,B)$ entre classes (A, B) que nous utilisons est donnée par l'équation (3.24). Cette mesure de distance est dérivée de la similitude

entre ensembles flous. On peut mesurer la similitude en utilisant la mesure d'information de chaque ensemble.

L'entropie floue, comme extension de la proposition de Shannon C.E.[SHA48] a été appliquée dans différents domaines pour estimer l'information d'un ensemble. Les formules pour calculer l'entropie varient en fonction de l'application pour laquelle elles ont été proposées (e.g. [ZAD68], [LUC72], [PAL92], [KAU75]). L'évaluation de l'information d'un ensemble flou, utilise la notion d'entropie floue (« *fuzziness* »). L'entropie floue est une mesure utile pour déterminer la similitude entre des ensembles flous [LEE01] [GOU05]. Les classes qui apportent la même quantité d'information peuvent être considérées comme similaires. Toutefois, pour obtenir une formule de similitude utile pour le diagnostic, il faut tenir compte de la distance relative entre les groupes. Pour cette raison, ce n'est pas suffisant de calculer l'entropie de chaque ensemble individuellement.

Quelques approches pour le calcul de l'entropie floue sont basées sur la comparaison entre un ensemble et son complément. Yager R. [YAG79] a proposé une mesure en tenant compte que l'intersection de l'ensemble flou A et de son complément A^c n'est pas nulle. L'entropie floue pour l'ensemble A est alors :

$$H_A = \frac{1 - \varphi(A, A^c)}{N} \quad \text{où} \quad \varphi(A, A^c) \text{ est une distance entre } A \text{ et } A^c \quad (3.22)$$

Kosko [KOS86] a employé la géométrie des ensembles flous et les distances entre eux pour proposer une autre formulation de l'entropie. Puis il a effectué une extension de la formule initiale pour la représenter en fonction de la cardinalité des ensembles (H_A):

$$H_A = \frac{M(A \cap A^c)}{M(A \cup A^c)} \quad \text{où} \quad M(A) = \sum_{n=1}^N \mu_{A_n} \quad (3.23)$$

A partir de la formule de Kosko et de Yager, en remplaçant A^c par B , l'expression obtenue est alors une relation entre les deux ensembles (A et B) qui correspond au complément de la mesure de distance $d^*(A, B)$ (eq.3.4). Cette mesure permet d'établir la similitude entre les classes en prenant en compte la quantité d'information des deux ensembles (A et B) et de leurs positions respectives (eq. 3.24).

$$G(A, B) = \frac{\sum_{n=1}^N (\mu_{A_n} \cap \mu_{B_n})}{\sum_{n=1}^N (\mu_{A_n} \cup \mu_{B_n})} \quad (3.24)$$

Cette mesure de similitude, associée à la mesure de dispersion que nous proposons par la suite, sera utilisée pour l'amélioration automatique des partitions en utilisant l'algorithme proposé dans la Section 3.5.

3.3.2 Mesure de dispersion des classes

Nous cherchons des partitions avec des classes les plus compactes possibles. Pour ce faire, nous évaluons la dispersion de la partition à partir des dispersions individuelles de chaque classe, de façon analogue à la méthode proposée dans plusieurs travaux antérieurs [FRA02][XIE91]. Cependant dans notre approche, nous proposons d'employer, pour la mesure de la dispersion, une mesure de quantité d'information qui est basée uniquement sur les degrés d'appartenance. Dans ce cas, nous considérons chaque classe comme un ensemble flou en utilisant la représentation présentée sur la Figure 3.1. La mesure d'information d'un sous-ensemble flou que nous cherchons doit avoir une valeur faible pour des ensembles qui sont très dispersés et une valeur élevée pour des ensembles bien définis (avec de forts degrés d'appartenance pour les données similaires au prototype de la classe et des degrés d'appartenance faibles pour les données très dissemblables au prototype).

Nous proposons de mesurer la quantité d'information de chaque classe en utilisant une nouvelle mesure d'entropie pour éviter le problème des ensembles qui ont tous des degrés d'appartenance élevés. En effet, pour ceux-ci, une mesure classique d'entropie donnerait une forte valeur de quantité d'information alors qu'en réalité, ce ne sont pas des ensembles bien conditionnés pour prendre une décision [ISA07].

Généralement, la quantité d'information est associée à l'entropie des données. Dans le cas des ensembles flous, la formule d'entropie non probabiliste proposée par De Luca et Termini [LUC72] a été largement utilisée.

Si pour l'estimation de la quantité d'information, on utilise une mesure classique d'entropie floue [LUC72][SHA48], l'ensemble peut être considéré informatif (bien défini) quand la quantité d'information est importante. Toutefois, un ensemble avec des valeurs des degrés d'appartenance faibles peut donner une valeur élevée de l'information. C'est pourquoi, nous proposons d'utiliser un indice d'information [ISA07] qui a été inspiré de la mesure d'information floue (entropie floue de Luca et Termini), mais qui permet de tenir compte du rapport entre les faibles degrés d'appartenance et la dispersion de l'ensemble.

Les fonctions d'entropie non probabilistes représentent le degré de flou d'un ensemble flou A discret vis-à-vis des éléments μ_{An} qui le composent. D'après l'approche proposée par De Luca et Termini, les fonctions de type entropie floue doivent répondre aux axiomes suivants [LUC72][TRI79][TRI78]:

$$\begin{aligned}
 P1 : H(A) = 0 &\Leftrightarrow \mu_{An} \in \{0,1\} \\
 P2 : \max H(A) &\Leftrightarrow \forall n \mu_{An} = 1/2 \\
 P3 : H(A) \leq H(B) &\Leftrightarrow A \geq_S B
 \end{aligned} \tag{3.25}$$

La relation d'ordre \geq_s est un opérateur de comparaison appelé "*sharpeness*". Un ensemble flou A est considéré comme plus "aigu" (*sharp*) que l'ensemble flou B si :

$$\begin{aligned} \forall \mu_{A_n}, \mu_{B_n} \in \mathcal{P}(X) \quad \text{Si } \mu_{B_n} \leq 0.5 \text{ alors } \mu_{A_n} \leq \mu_{B_n} \\ \text{Si } \mu_{B_n} \geq 0.5 \text{ alors } \mu_{A_n} \geq \mu_{B_n} \end{aligned} \quad (3.26)$$

De nombreuses études [ALS01] ont montré la validité de l'entropie proposée par De Luca et Termini comme mesure de l'information floue :

$$H(A) = \frac{-1}{N} \sum_{i=1}^N \mu_{A_i} \log \mu_{A_i} + (1 - \mu_{A_i}) \log(1 - \mu_{A_i}) \quad (3.27)$$

Une forme plus générale de l'entropie de l'ensemble discret flou A (eq. 3.27) est donnée par 3.28; elle permet un choix plus étendu des fonctions représentant l'information individuelle de chaque élément et de sa négation:

$$H_g(A) = \frac{-1}{N} \sum_{i=1}^N \mu_{A_i} g(\mu_{A_i}) + (1 - \mu_{A_i}) g(1 - \mu_{A_i}) \quad (3.28)$$

Cependant, dans le cadre d'une décision à prendre, cette entropie ne tient pas compte de la valeur du choix d'un élément de l'ensemble. Nous considérons que la dispersion d'un ensemble flou discret est un concept équivalent à l'analyse de la validité du choix de l'élément qui présente l'appartenance maximum à cet ensemble. C'est pourquoi, nous devons considérer l'ensemble flou formé par les différences d'appartenance avec le maximum μ_{MA} , il correspond à :

$$\delta_n = \mu_{MA} - \mu_{A_n} \quad \forall n \neq M \quad \text{et} \quad \mu_{MA} = \max[\mu_{A_n}] \quad (3.29)$$

La mesure d'information proposée est basée sur les mêmes principes que l'entropie classique, et correspond à [ISA07]:

$$I_D(A) = \Delta \cdot \sum_{n=1}^N \delta_n \cdot g(\delta_n) \quad (3.30)$$

La constante normalisatrice correspond à l'absence de choix:

$$\Delta = \frac{1}{N \cdot \mu_{MA} \cdot g(\mu_{MA})} \quad (3.31)$$

En choisissant $g(\xi) = e^{(\xi)}$, on définit la mesure d'information pour l'ensemble A :

$$I_D(A) = \frac{1}{N \cdot \mu_{MA} \cdot e^{\mu_{MA}}} \cdot \sum_{n=1}^N \delta_n \cdot e^{(\delta_n)} \quad (3.32)$$

Et donc l'entropie est

$$H_D(A) = 1 - I_D(A) \quad (3.33)$$

Cette mesure d'entropie permet d'établir la similitude entre un ensemble flou donné et l'ensemble flou le plus informatif qui serait le singleton correspondant à la valeur d'appartenance maximum qui est l'ensemble flou le plus compact. La dispersion d'une partition s'évalue à partir du complément de la mesure de l'information des classes H_{DA} .

L'équation 3.34 présente la mesure de dispersion que nous utilisons pour définir l'indice de qualité des partitions présenté dans la Section 3.4.1.3 .

$$Dis = \sum_{k=1}^K 1 - \frac{\sum_{n=1}^N \delta_{kn} \cdot \exp(\delta_{kn})}{N \cdot \mu_{Mk} \cdot \exp(\mu_{Mk})} \quad \text{où } \delta_{kn} = \mu_{Mk} - \mu_{kn} \quad \text{et} \quad (3.34)$$

$$\mu_{Mk} = \max[\mu_{kn}] \quad \forall n, n \in [1, N],$$

$K =$ nombre de classes

En utilisant la mesure d'information $I_D(A)$ pour évaluer la dispersion, nous incluons de manière indirecte, l'analyse de la qualité de chaque ensemble. De cette manière, nous recherchons des ensembles avec une information élevée ($I_D(A)$ élevé correspond à la dispersion minimale) qui sont utiles pour prendre des décisions. Une valeur de dispersion élevée correspond à une valeur basse de la quantité d'information pour l'ensemble des classes de la partition, et pourtant les classes seraient mieux définies (plus semblables au singleton) si la valeur de dispersion était basse.

La mesure de la quantité d'information a plusieurs applications, autres que celle de l'évaluation de la dispersion des partitions floues. Nous avons proposé une méthode de validation des transitions entre états de fonctionnement d'un processus dynamique. Cette application est basée sur la mesure d'information obtenue à chaque instant du fonctionnement du procédé [ISA07]. Dans l'étape de reconnaissance des états à partir des classes obtenues par un algorithme de classification floue, nous proposons de valider l'apparition d'un changement d'état. Ceci permet d'éviter des fausses transitions ou des transitions vers des états mal conditionnés, et d'éliminer des éventuelles fausses alarmes. Normalement, un individu est classé dans l'état pour lequel il a le plus grand degré d'appartenance, mais on considère une décision "mal conditionnée" quand les degrés d'appartenance à toutes les classes sont semblables ou proches d'un niveau d'incertitude. Dans ce cas, la décision n'est pas sûre et on ne doit pas valider la transition. Le critère essentiel associé à une décision a été considéré comme l'évaluation de l'information qui a été utilisée pour la prendre. C'est dans cette optique que nous avons cherché à mesurer cette information. Au contraire des mesures sur les partitions (e.g. dispersion), le problème qui nous concerne dans ce cas n'est plus l'analyse d'appartenance de plusieurs individus à une classe, mais le choix entre plusieurs classes (états) auxquelles un individu peut appartenir. Pour ceci, nous définissons un nouvel ensemble flou où l'univers de discours est défini par les classes K et les degrés d'appartenance correspondent aux valeurs d'appartenance de chaque individu à chaque classe. Nous pouvons chercher à analyser l'information instantanée sur laquelle une attribution à une classe (état) est faite; alors la transition ne sera

validée que dans les cas où cette information est considérée comme suffisante. Par conséquent, puisque l'analyse est basée sur l'ensemble flou défini par les degrés d'appartenance, nous mesurons la quantité d'information de cet ensemble en utilisant l'équation 3.32 et nous la comparons avec un niveau d'incertitude introduit *a priori* ; la décision est alors validée seulement si ce niveau d'incertitude est dépassé.

Cette méthode a été utilisée pour le diagnostic à partir de données historiques d'un procédé réel de production d'eau potable qui est en service dans la ville de Tuxtla au Mexique. Avec cette approche, la maintenance du système a été mieux programmée (en fonction des besoins réels) en évitant des états non désirables (quantité de boue élevée et trop grande fréquence des rétrolavages des filtres)[ISA07].

3.4 Indices d'évaluation de la qualité d'une partition

De manière générale il y a trois types d'approches pour établir la qualité des partitions [HAL02]:

1. Approches basées sur des critères externes: le résultat du groupement est évalué en tenant compte d'une structure spécifique qui est imposée aux données et qui représente l'idée que l'utilisateur a sur la structure de l'ensemble des données qui ont été classées. (e.g. comparaison avec une partition de référence).
2. Approches basées sur des critères internes : la partition de l'espace des données est évaluée en termes de mesures qui sont inhérentes aux données. Dans ce cas, il n'y a pas de comparaison entre partitions (e.g. matrice de proximité).
3. Approches basées sur des critères relatifs : l'idée est de comparer plusieurs partitions obtenues avec différents paramètres de l'algorithme de classification pour établir la meilleure en fonction d'un critère déterminé. Les mesures de qualité des partitions basées sur des critères proposés précédemment, sont utilisées dans ces approches.

Les approches 1 et 2 sont en général basées sur des critères statistiques. Les indices en relation avec ces approches permettent de mesurer le degré de similitude entre la partition obtenue et un schéma spécifié *a priori*. En revanche, l'objectif de la troisième approche est d'évaluer la meilleure partition que l'algorithme de classification peut trouver avec un bon choix des paramètres.

Dans cette section, nous décrirons trois indices proposés dans la littérature :

Le premier indice (Indice de Dissemblance, [LOP77]) permet d'estimer la similitude des partitions dans les cas où il y a une partition de référence (approche de type critère externe). Nous l'avons utilisé pour évaluer la performance de notre méthode, en

comparant la partition obtenue automatiquement et la partition proposée par l'expert du procédé dans le cas des applications industrielles présentées dans le Chapitre 4.

Les deuxième et troisième indices (approches avec critères externes) sont basés sur des critères de dispersion et de distance. Les mesures qu'ils utilisent ont été présentées dans la Section 3.3. Nous nous sommes inspirés de la structure de ces deux derniers indices pour proposer le nouvel indice de qualité.

L'indice de qualité des partitions que nous proposons est calculé uniquement à partir des seuls degrés d'appartenance en utilisant les mesures de dispersion et de distance que nous avons proposées dans la Section 3.3. L'indice est présenté dans la Section 3.4.3, cet indice est utilisé dans la méthode d'amélioration des partitions que nous décrivons dans la Section 3.5.

3.4.1.1 Indice de dissemblance entre partitions

L'indice proposé par Lopez de Mantaras [LOP77] permet de mesurer la dissimilitude entre deux partitions, ceci est très utile si on a une partition de référence. L'indice est calculé à partir de l'estimation des probabilités d'avoir des individus classés dans des classes différentes: une classe correspond à la partition obtenue automatiquement et l'autre à la partition de référence qui implique une connaissance *a priori* de la distribution des données.

La matrice de contingence est établie pour deux partitions non floues, C (dont les classes sont notées $c_1, c_2, \dots, c_i, \dots, c_p$) et F (dont les classes sont notées $f_1, f_2, \dots, f_j, \dots, f_r$) et les individus classés correspondent à l'ensemble des données $X = (x_1, x_2, \dots, x_N)$. Les probabilités correspondantes à chacune des classes et la probabilité de l'intersection entre une classe de la partition C et une classe de la partition F sont notées :

$$P_i = P(c_i) \quad P_j = P(f_j) \quad P_{ij} = P(c_i \cap f_j) \quad (3.35)$$

où $c_i \cap f_j$ dans ce cas, correspond à l'ensemble formé par les éléments qui appartiennent en même temps à la classe c_i et à la classe f_j .

Les probabilités satisfont les relations suivantes :

$$P_i = \sum_{j=1}^r P_{ij} \quad P_j = \sum_{i=1}^p P_{ij} \quad \sum_{i=1}^p \sum_{j=1}^r P_{ij} = 1 \quad (3.36)$$

Pour établir la similitude entre partitions, il faut déterminer la probabilité d'avoir des éléments qui appartiennent à la classe c_i et à la classe f_j , c'est à dire il faut estimer la probabilité de:

$$P(c_i \cap f_j) = \frac{M(c_i \cap f_j)}{N} \quad (3.37)$$

$M()$ correspond à la cardinalité et N le nombre total d'individus classés ($M(X)$).

La matrice de contingence, est une matrice formée par les valeurs des fréquences de $(c_i \cap f_j)$ pour $i=[1,p]$ et $j=[1,r]$ de la façon suivante :

		Partition F					
		fi	f2	...	fj	...	fr
Partition C	c1	P11	P12	...	P1j	...	P1r
	c2	P21	P22	...	P2j	...	P2r
	c3	P31	P32	...	P3j	...	P3r

	ci	Pi1	Pi2	...	Pij	...	Pir

	cp	Pp1	Pp2	...	Ppj	...	Ppr

Tab. 3.2. – Matrice de contingence

Cette matrice satisfait les propriétés suivantes [LOP77]:

Propriétés:

- 1) Si deux partitions sont égales la matrice formée est carrée et possède un seul terme différent de zéro par ligne et par colonne ; avec une permutation d'indices elle deviendra donc diagonale. Cette matrice sera appelée matrice « quasi-diagonale ».
- 2) Si deux partitions n'ont pas le même nombre de classes, mais qu'en ajoutant des lignes ou des colonnes, on peut obtenir une matrice « quasi-diagonale », cette matrice sera appelée « quasi-diagonalisable ».

La définition suivante nous permet d'établir la comparaison entre les partitions C et F à partir de la matrice de contingence [LOP77]:

Définition: Une partition F comprenant moins de classes qu'une autre partition C est *compatible* avec C, si et seulement si, C est une partition plus fine que F et alors la matrice de contingence ayant ses lignes caractérisées par C et ses colonnes par F est « quasi-diagonalisable ». On dit aussi que les partitions sont égales « modulo zero ».

Proposition: Si une matrice est « quasi-diagonalisable » la partition F qui a moins de classes, est compatible avec la partition C.

La *démonstration* est immédiate en vertu des propriétés énoncées. On en déduit que:

-Deux partitions sont *égales* modulo une permutation d'indices si la matrice qu'elles forment est « quasi-diagonale ».

-Deux partitions sont *compatibles* si la matrice de contingence est « quasi-diagonalisable », dans ce cas la partition qui a le plus grand nombre de classes est la plus fine.

Lopez de Mantaras [LOP77], à partir des probabilités de $(c_i \cap f_j)$, a proposé une mesure de la différence entre les deux partitions. Cette mesure est nulle si et

seulement si, la matrice de contingence est « quasi-diagonale » ou bien « quasi-diagonalisable », c'est-à-dire quand les partitions sont soit égales, soit compatibles ou égales modulo zéro. Il a proposé une première mesure de dissemblance (Id) estimée à partir de l'information conditionnelle entre les partitions F et C . L'information est calculée à partir de la formule d'entropie proposée par Shanon [SHA48].

$$Id(F,C) = \bar{H}(F/C) = - \sum_{i=1}^p \sum_{j=1}^r P_{ij} \log_2 \left(\frac{P_{ij}}{P_i} \right) \quad (3.38)$$

Il a appelé cette mesure : l'indice de dissemblance. Cet indice est positif dans tous les cas. Une normalisation de l'indice de dissemblance ($I'd(F,C)$) $0 \leq I'd(F,C) \leq 1$ est donnée par:

$$I'd(F,C) = \frac{\bar{H}(F/C)}{\bar{H}(C,F)} = \frac{\bar{H}(F/C)}{\bar{H}(F/C) + \bar{H}(C,F)} = \frac{- \sum_{i=1}^p \sum_{j=1}^r P_{ij} \log_2 \left(\frac{P_{ij}}{P_i} \right)}{- \sum_{i=1}^p \sum_{j=1}^r P_{ij} + \sum_{i=1}^p \sum_{j=1}^r P_{ij} \log_2 (P_{ij})} \quad (3.39)$$

Si la partition F est compatible ou égale à C , $I'd=0$. Au fur et à mesure que F et C sont plus dissemblables, la valeur de $I'd$ est plus élevée.

L'indice de dissemblance normalisé ($I'd$) a les 3 propriétés suivantes:

- 1) Positivité: $I'd(F,C) \geq 0$.
- 2) Annulation: $I'd(F,C)=0 \Leftrightarrow F=C$ ou $C \subset F$
- 3) Inégalité triangulaire: Si F , C et G sont trois partitions de l'ensemble des données X , on a : $I'd(F,C) + I'd(C,G) \geq I'd(F,G)$

Cet indice de dissemblance est une pseudo-distance dissymétrique à cause de l'absence de symétrie: $I'd(F,C) \neq I'd(C,F)$

L'application de cet indice à des cas concrets sera présentée au Chapitre 4. Nous avons utilisé cet indice et la matrice de contingence pour comparer les résultats obtenus en utilisant la méthode d'amélioration des partitions que nous proposons et la comparaison avec la partition proposée par un expert du procédé dans le cas de 3 applications industrielles.

3.4.1.2 Indices de qualité des partitions floues

Différents indices basés sur l'analyse de la dispersion et la distance entre classes ont été proposés dans la littérature pour déterminer la qualité d'une partition. Quelques indices de validation standards sont présentés dans [MAU02] [FRA02][ALE04]: "The partition F coefficient and partition H entropy [BEZ98]" . Les deux indices permettent de déterminer le nombre optimal de classes pour une partition non floue. "The minimum and maximum relative fuzziness index" [GOR92] évalue le degré de séparation entre classes floues. Néanmoins sa performance diminue quand il y a des classes qui se chevauchent (*overlapped*).

De manière générale, les matrices de dispersion de la méthode discriminante linéaire de Fisher "*Fisher's linear discriminate (FLD)*" [BIS95] sont utilisées pour mesurer la séparation et la compacité entre classes d'une partition stricte [MAU02].

Franco [FRA02] a proposé "*The Extended Fisher Linear Discriminant (EFLD)*", qui est l'extension floue du FDL (*Fisher Linear Discriminant*). L'indice FDL permet de mesurer la qualité des partitions floues et non floues, l'indice se caractérise par un temps de calcul plus faible que celui d'autres indices. Toutefois s'il y a beaucoup de classes, la performance diminue surtout si les classes se chevauchent (*overlapped*). De manière générale, dans ces méthodes, les valeurs des données classées sont nécessaires pour estimer les indices ce qui limite leur application aux données quantitatives.

Un indice de validation très reconnu pour les cas des partitions floues [ALE04][MAU02][BIS95] est "*The compactness and separation validity function S coefficient*" proposé par Xie [XIE91]. Cet indice permet d'évaluer le nombre optimal de classes et de mesurer la dispersion et la distance entre classes. Nous avons présenté la mesure de dispersion de cet indice (Π , eq.3.1) dans la Section 3.3. L'indice est donné par l'équation (3.40)

$$S = \Pi \frac{1}{\min_{i,k} \|V_i - V_k\|^2} \quad (3.40)$$

La mesure de dispersion (Π , eq.3.1), est basée sur les valeurs des données, les degrés d'appartenance, et la mesure de distance minimale entre classes $\min_{i,k} \|V_i - V_k\|^2$ qui correspond à une mesure de distance euclidienne entre les centres des classes (V_i, V_k).

Une partition avec des classes séparées et compactes a une valeur faible de l'indice S. Cet indice a une performance acceptable pour établir la qualité d'une partition, mais seulement pour le cas spécifique de l'algorithme de classification *Fuzzy C-Means* (Chapitre 2). Il a été montré [TAN05] que l'indice ne permet pas d'évaluer une partition si le nombre de classes est proche du nombre d'individus et qu'il est instable ou imprédictible si la valeur du paramètre de flou m est très élevée. Néanmoins avoir un nombre de classes similaire au nombre d'individus n'est pas commun dans le cadre du diagnostic de systèmes.

Un autre indice utile pour estimer la qualité des partitions floues est le *Inter Class Contrast (ICC)* [FRA02] (eq.3.41). L'*Icc* est basé sur l'EFLD index, c'est-à-dire qu'il est aussi basé sur l'extension floue des matrices de dispersion de Fisher mais l'*Icc* inclue le nombre de classes dans le calcul. Quand le nombre de classes augmente l'*Icc* montre de meilleures performances que l'EFLD. L'*Icc* présente des qualités similaires à celui proposé par Xie [XIE91], mais requiert beaucoup moins de temps de calcul.

L'*Icc* est estimé à partir des matrices de séparation floue U et des données X en utilisant la mesure de dispersion sbe présentée précédemment (Section 3.3, eq.,3.2) et la distance euclidienne entre les centres des classes. La valeur de ICC est obtenue par :

$$Icc = \frac{sbe}{N} \cdot D_{\min} \cdot \sqrt{K} \quad (3.41)$$

où N est le nombre de données classées, K le nombre des classes, D_{\min} la distance minimale entre groupes, sbe est la valeur associée à la dispersion de la partition. La valeur la plus élevée de l'ICC est obtenue pour la partition avec les classes les plus compactes et les plus séparées.

Cet indice nous servira de base de référence pour comparer avec les résultats obtenus avec l'indice que nous proposons dans la section suivante.

3.4.1.3 Indice de qualité des partitions floues basé uniquement sur les degrés d'appartenance floue.

En utilisant la même structure que l'indice Icc , nous proposons un indice de qualité des partitions (*Clusters Validity Index*, CV) qui a l'avantage d'être calculé uniquement à partir des degrés d'appartenance flous. L'indice CV est donc applicable aux partitions floues obtenues par les différentes méthodes de classification, en incluant les méthodes qui permettent de classer des données qualitatives sans représentation numérique.

Pour estimer l'indice, nous utilisons les mesures de dispersion de la partition (Dis , eq.3.34.) et de distance ($d^*(A,B)$, eq.3.4) présentées dans les sections précédentes. L'analyse de la qualité de la partition de l'espace est alors faite uniquement à partir de la matrice de partition floue U (degrés d'appartenance), il suffit donc qu'une première phase d'apprentissage fournisse une partition floue :

$$U^T = \begin{bmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{1n} \\ \dots & \dots & \dots & \dots \\ \mu_{k1} & \mu_{k2} & \dots & \mu_{kn} \end{bmatrix} \quad (3.42)$$

De la même façon que pour l'indice Icc , nous incluons le nombre de classes dans l'indice de *Clusters Validity* (CV). L'équation 3.43 donne la relation permettant le calcul de l'indice CV :

$$CV = \frac{Dis}{N} \cdot D_{\min}^* \cdot \sqrt{K} \quad (3.43)$$

La valeur de Dis est calculée en utilisant l'indice de dispersion que nous avons proposé (eq.3.34), et la valeur de D_{\min}^* correspond à la distance minimale entre classes en utilisant la formule de distance proposée ($d^*(A,B)$, eq.3.4). La partition avec les classes les plus compactes et les plus séparées aura la valeur de l'indice CV la plus élevée. La valeur minimale de l'indice est de zéro.

Cet indice est employé dans la méthode d'évaluation et d'optimisation des partitions que nous présentons dans le chapitre suivant.

3.5 Méthode d'optimisation des partitions

Nous avons étudié plusieurs méthodes pour optimiser la séparation de l'espace des données. La méthode d'ajustement des classes pour les méthodes d'apprentissage basées sur la distance, proposée par Kaymak et Babuska [KAY95], est une modification de l'algorithme *Compatible Cluster Merging* (CCM) proposé par Krishnapuram et Freg [KRI92]. Lurette [LUR03] propose l'union de groupes qui ont été créés par un réseau neuronal adaptatif. Les prototypes des classes similaires ont un degré d'appartenance plus grand qu'un seuil choisi *a priori*. Lee et al. [LEE01], ont proposé une méthode itérative où la qualité de la partition est déterminée en fonction de l'entropie.

La plupart des techniques existantes pour trouver automatiquement le nombre des classes sont basées sur le regroupement des classes. Comme la plupart des travaux sont des compléments aux algorithmes de coalescence, les mesures sont basées sur l'union de classes définies géométriquement. S'il n'y a pas de regroupement des classes, les approches consistent à faire plusieurs classifications avec différents paramètres de l'algorithme de classification et de comparer les partitions pour choisir la meilleure.

Une alternative similaire à l'approche que nous proposons, mais qui n'est pas applicable aux méthodes de classification floue est présenté dans [WAN05]. Cette approche recherche le nombre optimal de classes non floues obtenues avec une méthode de classification non supervisée hiérarchique. Pour trouver le nombre optimal, les auteurs utilisent la mesure de dispersion basée sur les matrices de dispersion de Fisher. Pour l'amélioration de la partition, le regroupement des classes similaires a été adopté. Cette méthode n'utilise pas la partition floue et n'est pas basée sur la classification floue; la méthode de classification employée est de type hiérarchique.

Kaymak U. et Setnes [KAY02] montrent qu'une approche très utile pour trouver le nombre de classes est de partir d'une partition avec une grande quantité de groupes et de regrouper itérativement les groupes les plus semblables jusqu'à arriver au nombre optimal de classes. De plus, au moment du regroupement on peut supprimer de façon supervisée, les classes non désirables. La méthode d'optimisation des partitions floues qui est souvent proposée est basée sur ce principe. Les mesures pour estimer la qualité des partitions sont dépendantes de la géométrie des classes donc applicables uniquement aux algorithmes de coalescence. L'algorithme consiste à réunir des classes jusqu'à arriver à un seuil qui doit être estimé de manière indépendante de l'algorithme.

La méthode d'optimisation que nous proposons est un algorithme basé sur le regroupement des classes similaires et l'analyse (chaque fois que deux classes sont réunies) de la qualité de la partition obtenue. Les critères sur lesquels est évaluée la partition sont la compacité des classes et la distance entre elles: nous cherchons une partition avec des classes compactes et séparées. L'objectif est de pouvoir améliorer une partition initiale, sans avoir besoin des valeurs des données, nous utilisons donc l'indice de qualité que nous avons proposé dans la section précédente.

Pour arrêter l'algorithme d'optimisation, nous sommes partis du principe que le nombre K de classes est obligatoirement fini, le nombre de partitions obtenues sera égal ou inférieur à K . Donc le maximum de l'indice de qualité sera toujours atteint en un nombre fini d'itérations. Dans le cas où le nombre initial K de classes est très grand, on peut proposer des heuristiques qui consistent à proposer des critères d'arrêt de la recherche avant d'avoir effectué toutes les réunions de classes jusqu'à en obtenir 2. Par exemple, une stratégie est de continuer les itérations jusqu'à ce que l'indice de qualité de partitions diminue, et de récupérer la matrice de partition U de l'itération précédente, qui a été celle qui a permis d'obtenir l'indice CV le plus élevé. Ces critères sont basés sur la décroissance de l'indice sur une ou plusieurs itérations. Cependant, ils présentent le risque d'arrêter l'algorithme sur un maximum local, comme on le verra sur l'exemple académique présenté dans le chapitre suivant (Figure 4.8).

Dans la méthodologie que nous proposons, nous traitons la partition floue U comme un ensemble d'ensembles flous. Cette représentation est présentée sur la Figure 3.1. A chaque itération, nous déterminons la compatibilité entre classes en utilisant le concept de similitude floue entre deux classes A et B ($G(A,B)$, (eq.3.24)). Itérativement, nous regroupons les deux classes les plus similaires. En employant une S-Norme (par exemple le max) nous obtenons la nouvelle matrice de partition floue $U(T+1)$. Cette matrice représente la nouvelle partition. La figure 3.2 représente l'actualisation de la matrice de partition floue:

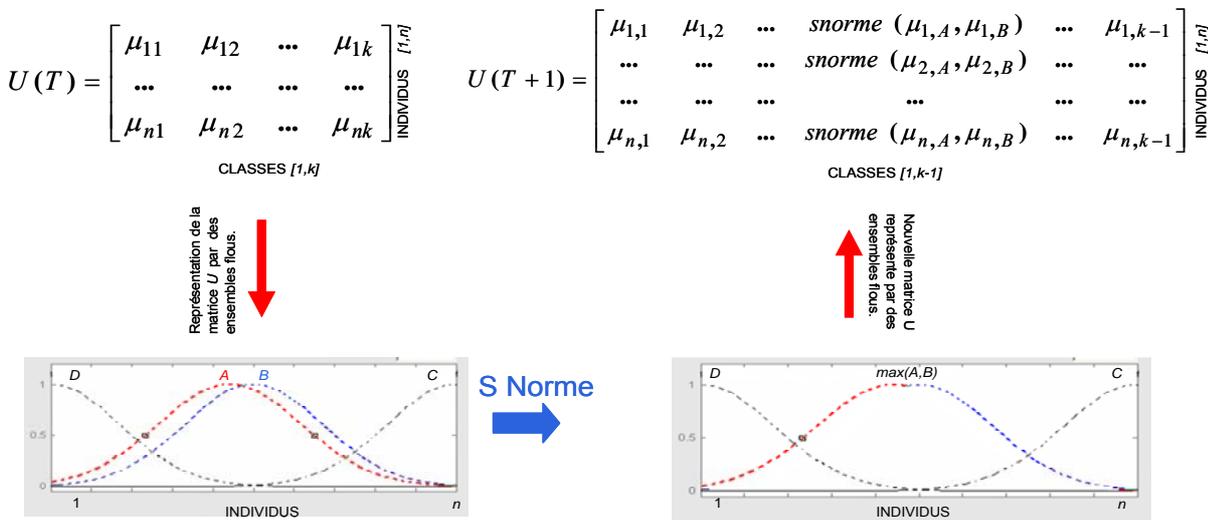


Figure 3.2. Représentation de l'actualisation de la matrice de partition floue

L'algorithme continue jusqu'à ce que la partition n'ait plus que deux classes. Nous évaluons la qualité de chaque matrice de partition floue obtenue en utilisant l'indice CV. La meilleure partition correspond à celle qui a la valeur de CV la plus élevée.

Ce processus cherche à conserver la meilleure performance de manière similaire à la technique d'apprentissage du perceptron de poche (the *pocket perceptron learning algorithm*) [BIS95] pour l'entraînement des réseaux de neurones.

3.6 Conclusion

L'évaluation et l'optimisation automatique des partitions sont des points clés permettant de diminuer la dépendance des algorithmes de classification vis-à-vis des paramètres et de trouver automatiquement le nombre de classes d'un ensemble de données. La validation de partitions est une aide pour l'évaluation des techniques de classification et pour la sélection de la classification qui représente le mieux la structure de l'ensemble des données. Plusieurs travaux ont été développés sur ce sujet, la plupart sont des compléments aux algorithmes de coalescence (fortement dépendants de paramètres à choisir par l'utilisateur). Les approches pour optimiser les partitions sont donc basées sur la matrice de covariance obtenue avec ces méthodes de classification. D'un autre côté, les indices pour évaluer les partitions sont basés en général sur des mesures de dispersion et de distance entre classes qui incluent dans les calculs les valeurs des données classées. C'est pour cela que nous avons proposé une méthode d'optimisation des partitions qui, au contraire de ces approches, ne dépend ni de la méthode de classification, ni des valeurs des données. Nous avons conservé comme critères pour évaluer la qualité des partitions, la dispersion des classes, la distance et le nombre de classes.

Au cours de ce chapitre, nous avons présenté des critères pour évaluer la qualité d'une partition floue de telle sorte à améliorer l'identification des états fonctionnels d'un procédé. Les mesures et indices que nous employons pour l'optimisation des partitions dépendent des degrés d'appartenance. L'indice de qualité des partitions, proposé tient compte des critères amplement utilisés dans les indices usuels de qualité de partitions comme la dispersion et la distance des classes.

Pour estimer la dispersion de la partition, nous proposons une mesure d'information qui inclue des critères de qualité de l'ensemble flou constitué par les degrés d'appartenance des individus aux différentes classes. Pour le calcul de la distance, nous proposons une mesure de distance qui tient compte de l'entropie des classes et de leur position.

La méthode d'optimisation des partitions floues est utile dans les cas où il n'y a pas une représentation géométrique des classes. La méthodologie proposée est considérée comme un complément aux méthodes de reconnaissance de formes qui sont utilisées pour l'identification de défaillances de systèmes complexes. Cette méthode n'a pas de restriction par rapport au type des données, elle est applicable aussi dans le cas de systèmes de diagnostic basés sur des méthodes de classification qui permettent de travailler directement avec des données qualitatives comme c'est le cas de la méthode LAMDA.

Chapitre 4

4. APPLICATIONS DE LA METHODE D'ÉVALUATION ET D'OPTIMISATION DES PARTITIONS FLOUES

4.1 Introduction

Dans le chapitre précédent nous avons développé une méthode de validation et d'optimisation des partitions floues. Cette méthode donne aux experts du procédé une orientation pour établir les états fonctionnels lors de l'implantation d'une technique de surveillance basée sur des méthodes de classification floue.

Ce chapitre est consacré à la présentation et à la discussion des résultats obtenus lors de l'application de la méthode proposée sur un cas d'étude avec des données générées artificiellement, cas communément employé pour l'évaluation des indices de qualité des partitions, et sur 3 cas tirés du domaine industriel.

Les applications industrielles sont du domaine du Génie Chimique. Le premier est un procédé complet de production de propylène glycol incluant plusieurs opérations unitaires : mélange, réaction chimique, distillation. Les deux derniers processus concernent deux réactions chimiques (réaction d'estérification et réaction du thiosulfate de sodium) mises en œuvre sur un nouveau type de réacteur chimique développé par le Laboratoire de Génie Chimique de Toulouse, en partenariat avec ALFA-LAVAL, qui utilise la technologie "*Open Plate Reactor*", il s'agit d'un nouveau concept de réacteurs avec échange thermique intégré [DEV03].

Dans chacun des cas, nous obtenons une partition initiale en utilisant la méthode de classification LAMDA (Chapitre 2) avec apprentissage non supervisé (sans donner *a priori* le nombre de classes). Cette partition est améliorée en utilisant la méthode d'optimisation proposée. La partition obtenue est comparée avec la classification qui avait été proposée par un expert du procédé. En ce qui concerne l'application de la

méthode LAMDA, nous avons travaillé avec le logiciel SALSA [KEM04b], et la méthode d'amélioration des partitions a été implantée sur *Matlab*[®].

Pour bien illustrer le comportement de la méthode de validation et d'optimisation des partitions floues proposée, nous avons choisi un exemple simple constitué par des données artificielles. Dans un exemple à deux dimensions, nous mettons en évidence les étapes du fonctionnement de la méthode, ensuite nous donnons un exemple analogue à trois dimensions qui nous a paru plus démonstratif. Ce type d'exemple a été employé dans d'autres travaux faisant référence à l'évaluation de la qualité de partitions. [FRA02][WAN05][KAY02].

Des applications de cette méthode à des procédés chimiques dans des contextes proches de l'utilisation industrielle sont ensuite décrites dans les sections suivantes. Nous présentons l'application de la méthode d'évaluation et d'optimisation de classes que nous avons développée avec comme objectif l'obtention du nombre optimal de classes, et de cette manière fournir une aide aux experts pour trouver le nombre et les classes optimales en fonction du critère mathématique de séparation et compacité des classes. Les classes optimales que nous donnons peuvent être associées plus directement aux états du processus. Nous comparons la partition optimale proposée avec la partition que l'expert du procédé avait établie pour chaque cas.

Le procédé de production de propylène glycol, la réaction d'estérification et la réaction de thiosulfate de sodium qui sont présentés dans les sections suivantes, ont été étudiés dernièrement dans une thèse développée au sein du groupe DISCO [ORA05] et dans un projet industriel [KEM06]. L'objectif principal de ces travaux a été la conception d'une méthodologie permettant de placer les capteurs les plus pertinents sur un processus en vue de son diagnostic à partir de données historiques, de manière à pouvoir détecter en ligne les états fonctionnels associés à des défaillances lors du suivi du procédé. Cette méthodologie est basée sur l'association de méthodes de mesure de la quantité d'information (entropie de Shannon) délivrée par des signaux issus d'un système et de la classification de données. A partir des données d'évolution temporelle des capteurs (constituant l'ensemble de tous les possibles) suite à des scénarii de défaillance, il est possible d'identifier les capteurs les plus pertinents et d'obtenir un modèle de comportement du processus à un niveau d'abstraction tel qu'il soit utilisable pour le diagnostic [KEM06].

La méthode proposée dans [ORA05][ORA07a] a quatre étapes essentielles :

1. Etablir les états fonctionnels (normal et de défaillances) du procédé en utilisant les données d'apprentissage avec tous les capteurs (descripteurs) disponibles. Pour identifier les états, une méthode de classification non supervisée est utilisée (LAMDA dans ces travaux, mais tout autre méthode de classification peut être employée). Tout d'abord, les différentes classes possibles sont identifiées et ensuite, avec l'aide de l'expert du procédé, une association entre classes et états fonctionnels est réalisée. Les données d'apprentissage, comme dans tous les cas d'extraction de la connaissance par fouille de données, doivent être les

plus riches possibles et doivent être obtenues de façon à couvrir la plupart des états fonctionnels et des défaillances du processus à surveiller.

2. Décider quels sont les capteurs les plus pertinents pour la détection des différents états. Pour ceci, une analyse basée sur l'entropie de l'information délivrée par chaque descripteur suivant l'état de fonctionnement est réalisée. De cette manière, les descripteurs donnant l'information la plus riche en termes d'entropie sont identifiés et comme chaque descripteur est associé à un capteur potentiel, il en résulte une liste des capteurs les plus pertinents pour constituer le modèle de comportement du procédé en vue de son diagnostic.
3. Actualiser le modèle de comportement en utilisant les données d'apprentissage uniquement avec les descripteurs les plus influents. De la même manière que dans la première étape, les états fonctionnels sont établis mais uniquement à partir des descripteurs (capteurs) restreints à ceux identifiés à l'étape précédente.
4. Cette étape comporte la reconnaissance en ligne des nouvelles données en appliquant le modèle de comportement identifié dans l'étape précédente.

Avec ces quatre étapes, un modèle dédié à la surveillance et la détection d'état fonctionnel est établi en n'utilisant que les capteurs les plus pertinents. La Figure 4.1. donne le schéma général de la méthode.

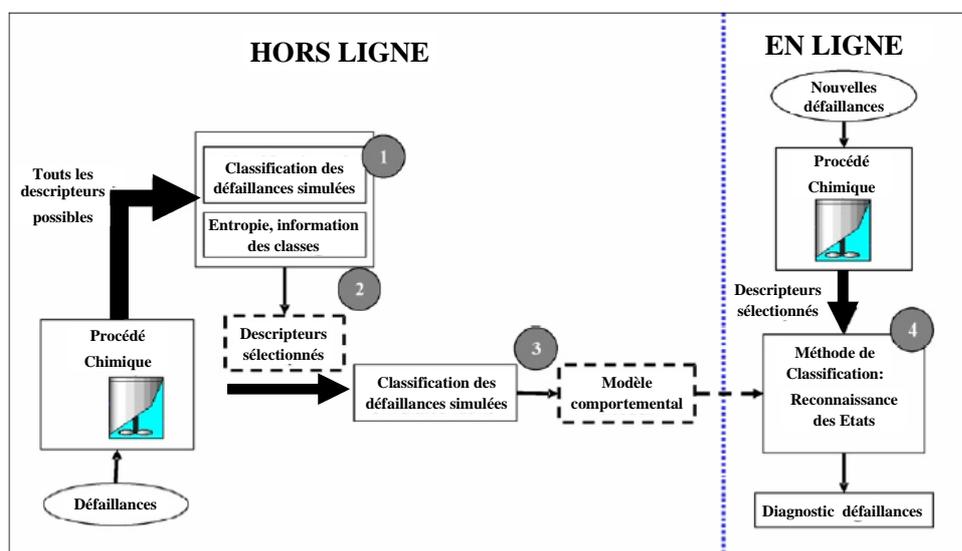


Figure 4.1. Schéma général de la méthode proposée par Orantes A. [ORA07a]

L'association des classes à des états (partie des étapes 1 et 3, Figure 4.1.) est une étape clé de cette méthodologie. Il faut remarquer que cette phase de la procédure, préalable à la phase de sélection des capteurs pertinents proprement dite, est très importante car elle doit permettre d'obtenir des classes bien définies et concises de

telle sorte que les valeurs obtenues à partir du profil de classes soient les plus précises possibles, et ainsi, permettre d'avoir une meilleure sélection des capteurs. [ORA05]. Néanmoins, dans les travaux précédents, la qualité de cette classification était laissée à l'appréciation de l'expert qui ne disposait pas d'une aide significative pour bien identifier les classes correspondant aux états fonctionnels. Dans les trois cas étudiés, le nombre de classes obtenues est très élevé ce qui amène une certaine difficulté pour l'association classes/états. Dans ces deux premiers travaux, l'assignation de classes aux états fonctionnels a été effectuée à partir d'une analyse très exhaustive des classes de la part de l'expert du procédé, sans aucun indice de la qualité des partitions et sans aucun automatisme.

Dans les sections 4.3, 4.4 et 4.5, nous discutons les résultats de l'application de la méthode d'évaluation et d'optimisation proposée aux trois processus chimiques.

4.2 Cas illustratifs avec données artificielles

4.2.1 Cas d'étude dans l'espace 2D

Deux ensembles de données ont été obtenus de manière artificielle : à partir des données de coordonnées (2,2.5), (4,2.5), (3,7), (2,5), (4,5) un ensemble de 1000 nombres aléatoires (500 par dimension) avec une distribution gaussienne ont été générés pour chacune des 5 classes, avec deux écarts type différents. Les caractéristiques des données pour chaque ensemble sont décrites dans le Tableau 4.1.

Classe	Centres	Ecart Type Ensemble 1	Ecart Type Ensemble 2	Nombre de Données
1	(2,2.5)	0.5	0.7	500
2	(4,2.5)	0.5	0.7	500
3	(3,7)	0.5	0.7	500
4	(2,5)	0.5	0.7	500
5	(4,5)	0.5	0.7	500

Tab. 4.1. – Description des classes générées aléatoirement.

Les données du premier ensemble qui ont un écart type de 0.5 (Figure 4.2.) ont une distribution facilement identifiable par un algorithme d'apprentissage. En revanche la distribution des classes pour le deuxième ensemble (Figure 4.3.) est plus difficile à identifier.

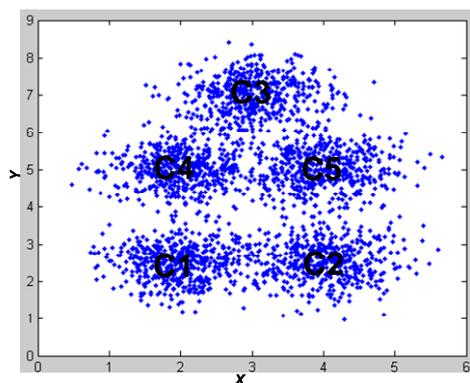


Figure 4.2. Ensemble de données 1

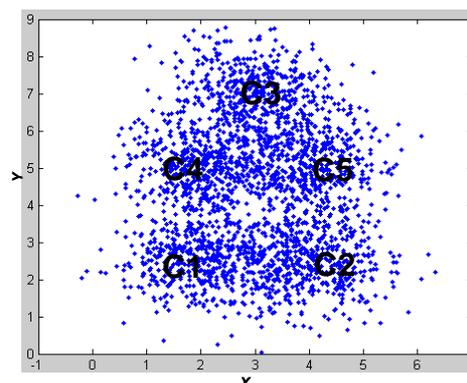


Figure 4.3. Ensemble de données 2

En utilisant l'algorithme de classification floue LAMDA avec apprentissage non supervisé (sans fixer *a priori* le nombre de classes), pour chaque cas, nous avons trouvé une première partition avec un nombre de classes plus grand que le nombre réel (5 classes). Pour le premier ensemble, LAMDA a identifié 13 classes et pour le deuxième ensemble 8 classes. Les Figures 4.4. et 4.5. présentent la partition de l'espace de données obtenue avec LAMDA respectivement pour les deux cas. Pour effectuer la représentation graphique, chaque point de l'ensemble de données a été classé dans la classe avec le degré d'appartenance le plus élevé. Chaque classe est représentée par une couleur différente et le centre est représenté par une étoile noire.

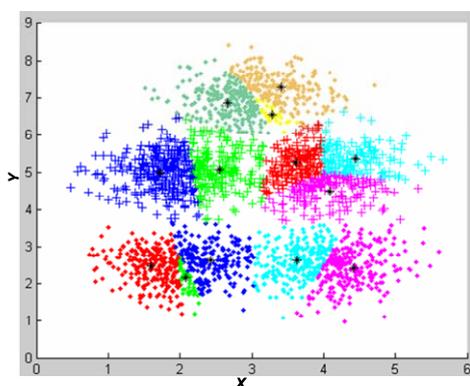


Figure 4.4. Classification Initiale Ensemble de Données 1

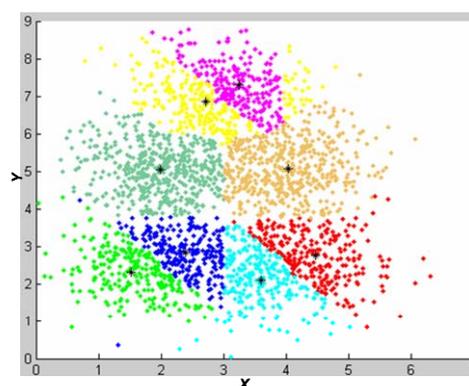


Figure 4.5. Classification Initiale Ensemble de Données 2

Nous pouvons remarquer que LAMDA a effectué une partition adéquate mais non optimale. D'une part, LAMDA a trouvé dans les deux cas, des classes qui correspondent à une partition plus granulaire que la partition optimale. D'autre part, LAMDA a effectué une partition de l'espace suivant des lignes droites. Ceci est dû au fait que la fonction MAD employée est la fonction binomiale, ce type de fonction n'est pas la plus adéquate puisque la distribution statistique des données est de type gaussien, ce qui introduit des frontières ellipsoïdales ou leurs intersections.

Pour chaque cas et en utilisant uniquement la matrice de partition obtenue comme résultat de la classification (valeurs du GAD pour chaque donnée), la méthode d'optimisation de partitions proposée (Chapitre 3) a permis d'améliorer chaque partition. La Figure 4.6. présente l'évolution des partitions du deuxième cas (données avec écart type de 0.7) au fur et à mesure des itérations. Dans le premier cas, on obtient une séquence analogue.

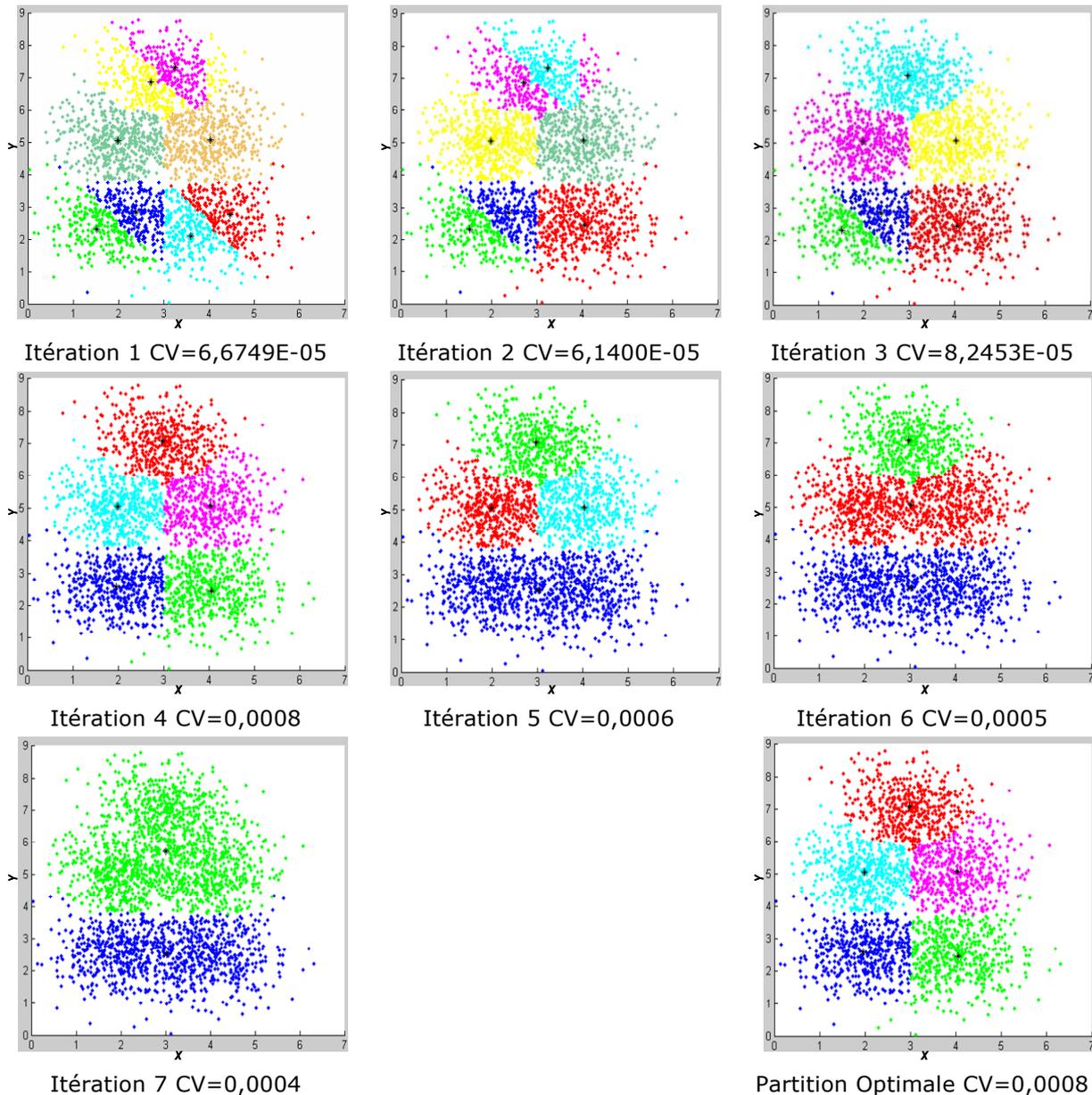


Figure 4.6. Evolution des partitions ensemble de données 2, Algorithmme d'évaluation et d'optimisation de partitions floues

A chaque itération, d'une part l'algorithme estime la qualité de la partition en termes de séparation et de compacité des classes (Indice CV, Chapitre 3), d'autre part l'algorithme effectue l'union des deux classes les plus semblables jusqu'à obtenir une

partition avec uniquement deux classes (Mesure de similitude, Chapitre 3). La partition optimale est la partition avec la valeur CV la plus élevée. La Figure 4.7. et la Figure 4.8. correspondent à la partition optimale trouvée pour chaque ensemble de données.

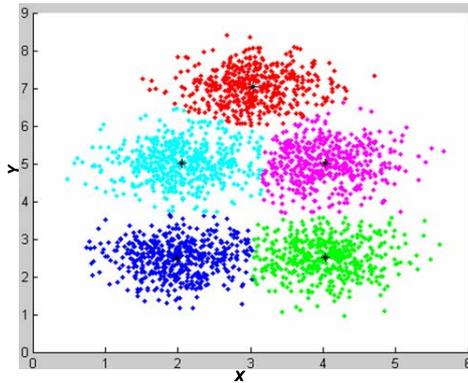


Figure 4.7. Classification Optimale Ensemble de Données 1

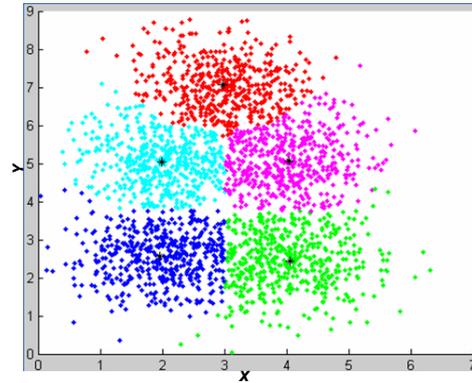
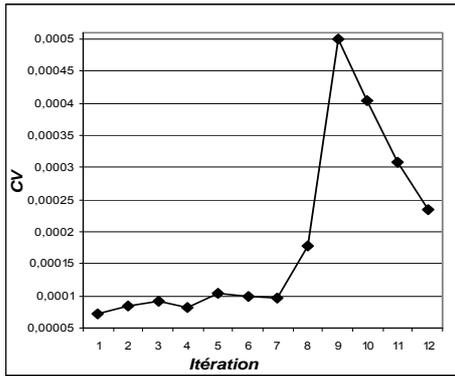


Figure 4.8. Classification Optimale Ensemble de Données 2

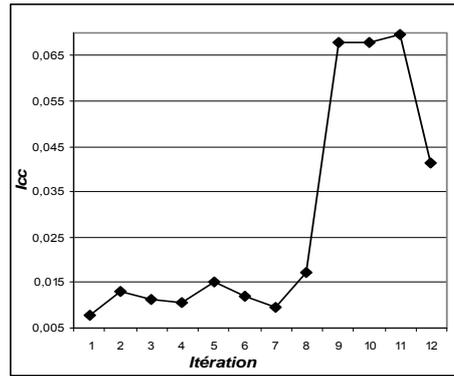
Dans les deux cas, la méthode d'optimisation a ajusté la partition initiale et a établi automatiquement la meilleure partition. Dans le premier cas, la partition adéquate était facilement identifiable par une méthode d'apprentissage mais dans le deuxième cas, il était beaucoup plus difficile de bien regrouper les classes. L'application de la méthode proposée a permis d'obtenir automatiquement les classes souhaitées.

Cette méthode fournit un critère mathématique qui permet d'établir la classe optimale obtenue à partir de la partition de l'espace de données initiale, ainsi que la matrice de partition floue qui correspond à la partition optimale. Avec cette matrice d'appartenance on peut caractériser les classes résultantes, par exemple en faisant une analyse statistique des données ou en utilisant la méthode de classification avec un apprentissage supervisé à partir des classes obtenues.

Pour évaluer la performance de la mesure de qualité proposée, à chaque itération nous avons comparé notre approche et la mesure de qualité des partitions Icc (Chapitre 3, [FRA02]). Cette comparaison est possible dans cette application parce que les données de l'exemple sont uniquement de type quantitatif; l'indice Icc est applicable uniquement aux données quantitatives puisque, pour ce calcul, il est nécessaire d'inclure les valeurs des données (descripteurs). Les Figures 4.9. et 4.10. présentent l'évolution des valeurs d'Icc et des valeurs de CV au fur et à mesure des itérations pour chaque ensemble de données : écart type = 0.5 et écart type = 0.7. La valeur la plus élevée de l'indice Icc correspond à la meilleure partition par rapport à cet indice.

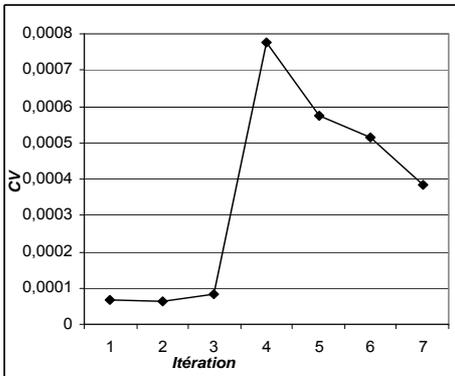


Indice de qualité des partitions CV

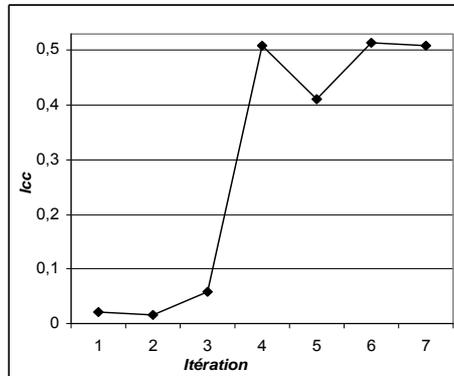


Indice de qualité des partitions Icc

Figure 4.9. Comparaison des indices CV et Icc, ensemble de données 1



Indice de qualité des partitions CV



Indice de qualité des partitions Icc

Figure 4.10. Comparaison des indices CV et Icc, ensemble de données 2

La dernière itération correspond à une partition ne possédant plus que deux classes. Dans les deux cas étudiés, l'indice Icc a une valeur élevée pour la partition attendue (5 classes) mais qui n'est pas la plus élevée, la valeur maximale d'Icc correspondant à la partition avec 3 classes. En regardant l'espace de données, l'obtention de cette partition peut paraître logique. Cependant en tenant compte que les données ont été créées à partir de 5 classes proches mais différentes, le résultat souhaité est bien une partition avec 5 classes. Dans le cas où, à partir de l'union des classes, on arrive à avoir des classes très dispersées, l'indice CV permet une meilleure relation entre la dispersion et la séparation des classes, en conséquence il permet d'obtenir la partition attendue.

Pour mettre en évidence la différence entre les deux indices, les Figures 4.11. et 4.12. donnent la valeur de dispersion de la partition à chaque itération pour les deux ensembles. La valeur *sbe* correspond à la mesure de dispersion basée sur une extension floue des matrices de dispersion de Fisher, *sbe* est la mesure proposée par Franco pour estimer l'indice Icc (Chapitre 3, [FRA02]). Pour l'indice CV proposé dans

notre travail (Chapitre 3), la mesure de dispersion (Dis) utilise uniquement les degrés d'appartenance.

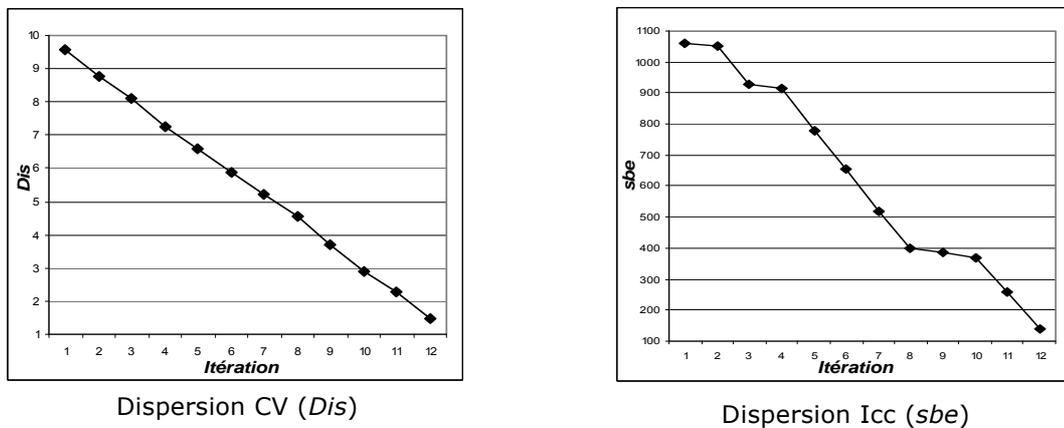


Figure 4.11. Comparaison de la dispersion indices CV et Icc, ensemble de données 1

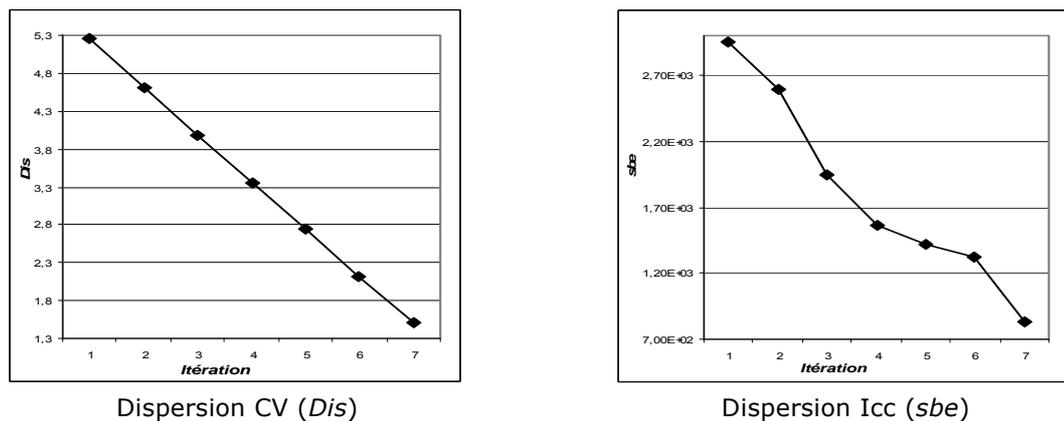


Figure 4.12. Comparaison de la dispersion indices CV et Icc, ensemble de données 2

Les Figures 4.13. et 4.14. correspondent à la comparaison de la distance minimale entre classes à chaque itération pour chaque ensemble de données. A droite, nous présentons la distance euclidienne (employée pour calculer l'Icc) et à gauche la distance estimée à partir de la similitude des classes (employée pour calculer le CV) (Chapitre 3). On peut remarquer que les deux mesures sont adéquates pour estimer la distance entre ces ensembles. La deuxième a le très net avantage d'être estimée uniquement à partir des degrés d'appartenance et son temps de calcul ne dépend pas de la dimension des données. De plus, ce calcul peut être effectué aussi lorsqu'il y a des données de type qualitatif.

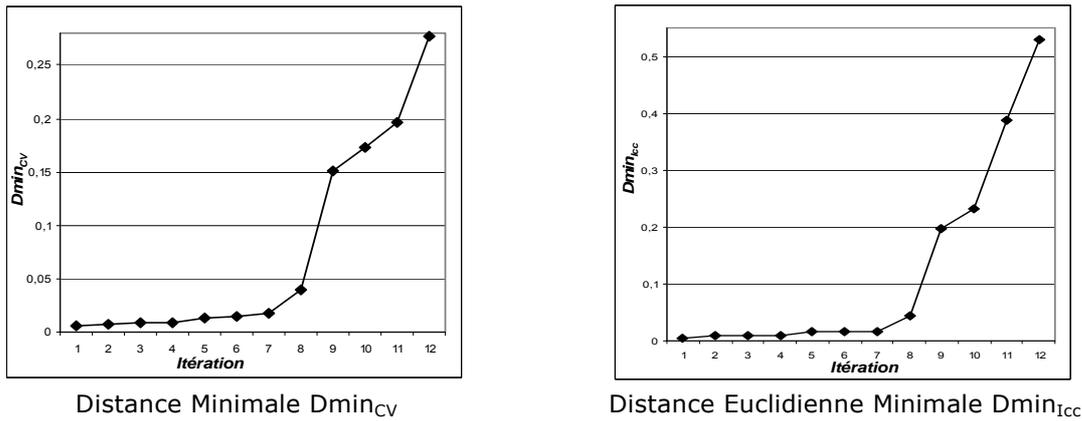


Figure 4.13. Comparaison distance minimale pour indices CV et Icc, ensemble de données 1

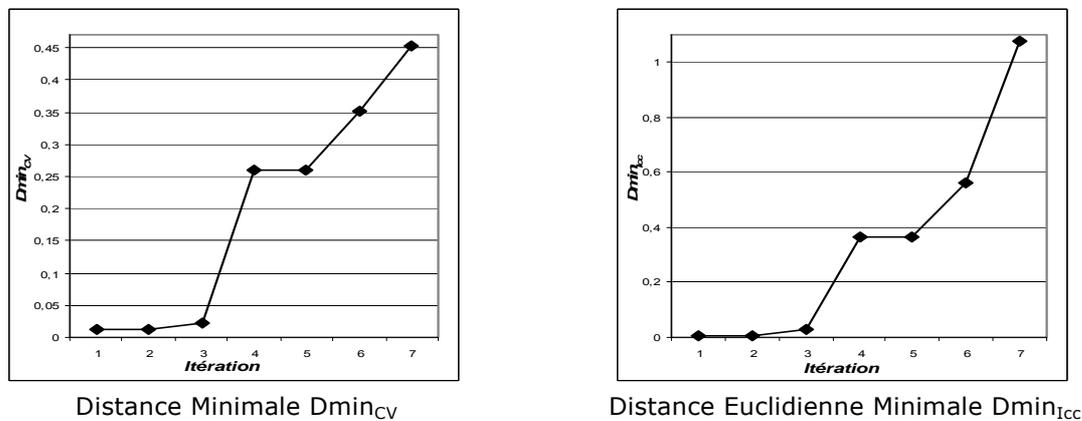


Figure 4.14. Comparaison distance minimale pour indices CV et Icc, ensemble de données 2

La mesure de la distance basée sur les degrés d'appartenance présente les mêmes tendances que la distance euclidienne, mais elle établit directement une relation de similitude entre ensembles flous et ne dépend donc pas des données. Cette caractéristique est très intéressante dans le cas de la supervision de processus industriels où souvent le nombre de descripteurs (mesures ou résultats de traitement de mesures) à traiter est important ; les performances de la méthode d'amélioration proposée ne varient donc pas en fonction du nombre de descripteurs.

On peut, en cours de supervision, faire une actualisation des classes en un temps très court, en procédant à un nouvel entraînement de l'algorithme de classification. Lors d'une phase de reconnaissance, si le processus présente un nouvel état ou un état non reconnu, il est possible d'envisager une validation et un réajustement des classes. Dans le cas de la supervision des procédés industriels, il est primordial de disposer d'un algorithme présentant un temps de calcul faible, dépendant uniquement du nombre des classes de la partition initiale, et qui fournit des partitions avec un nombre faible de classes permettant une interprétation directe pour établir les états fonctionnels du procédé nécessaires pour le système de surveillance.

4.2.2 Cas d'étude dans l'espace 3D

Dans ce qui suit, nous présentons un troisième cas qui correspond à l'extension du dernier cas d'étude à l'espace 3D. Le Tableau 4.2 présente la description des classes.

Classe	Centres	Ecart Type Ensemble 3	Nombre de Données
1	(5,5,5)	0.7	500
2	(2.5,2.5,2.5)	0.7	500
3	(7.5,7.5,7.5)	0.7	500
4	(2.5,2.5,7.5)	0.7	500
5	(7.5,7.5,2.5)	0.7	500

Tab. 4.2. - Description des classes générées aléatoirement 3D.

L'évolution des partitions au fur et à mesure des itérations est présentée sur la Figure 4.15.

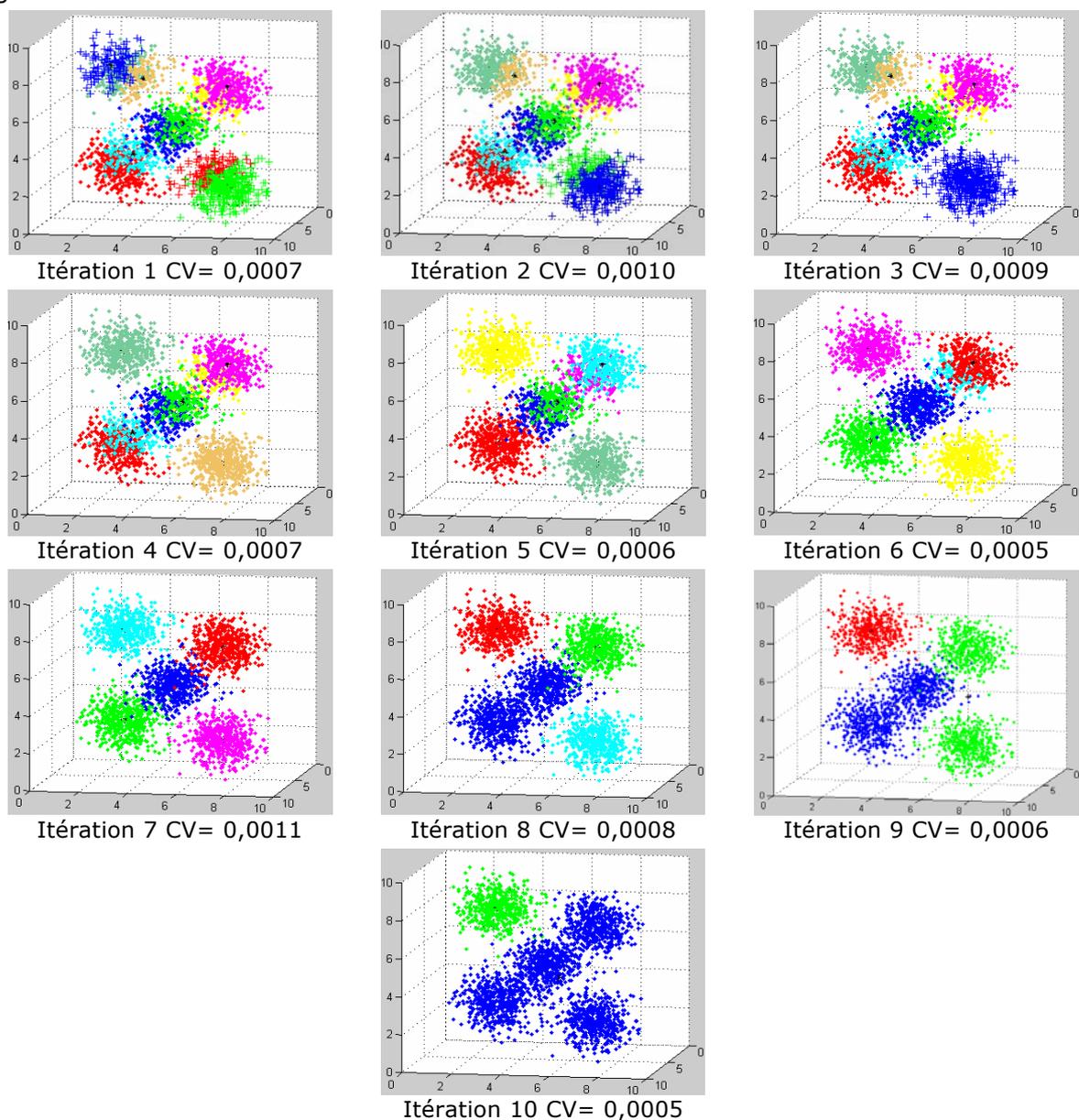
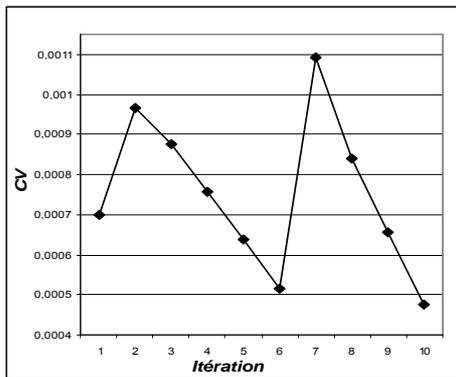
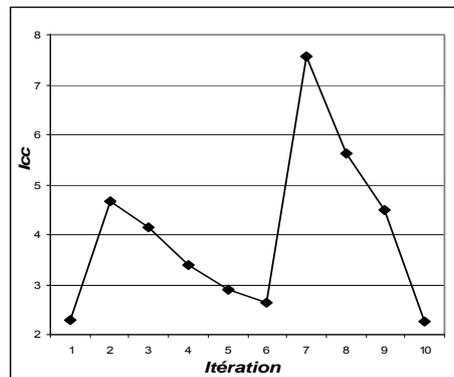


Figure 4.15. Evolution des partitions ensemble de données 3, Algorithmes d'évaluation et d'optimisation de partitions floues

La méthode d'amélioration des partitions proposée a permis, à partir d'une classification initiale de 11 classes, d'obtenir (Itération 7) une partition finale ne comportant plus que les 5 classes souhaitées. Nous présentons sur la Figure 4.16., les évolutions des indices CV et Icc, qui dans ce cas fournissent le même résultat, c'est-à-dire que les valeurs les plus élevées pour Icc et pour CV correspondent toutes les deux à la partition attendue comportant les 5 classes.



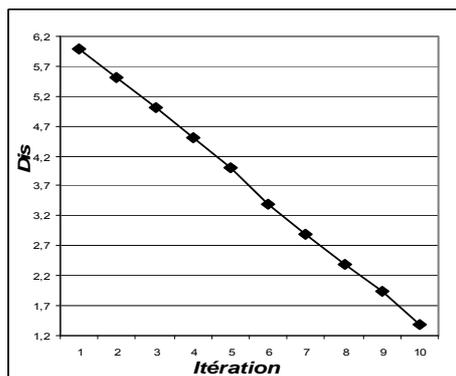
Indice de qualité des partitions CV



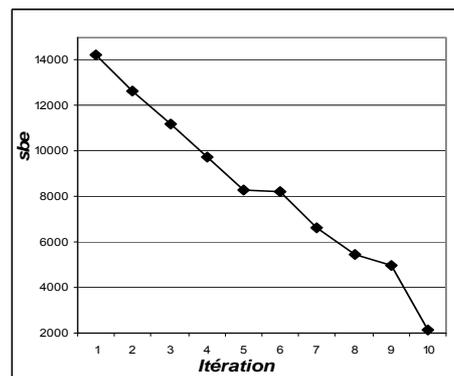
Indice de qualité des partitions Icc

Figure 4.16. Comparaison des indices CV et Icc, ensemble de données 3

La comparaison des dispersions et des distances minimales à chaque itération est présentée sur les Figures 4.17. et 4.18. respectivement.



Dispersion CV (*Dis*)



Dispersion Icc (*sbe*)

Figure 4.17. Comparaison de la dispersion indices CV et Icc, ensemble de données 3

Dans ce cas, la tendance des dispersions est semblable pour les deux indices.

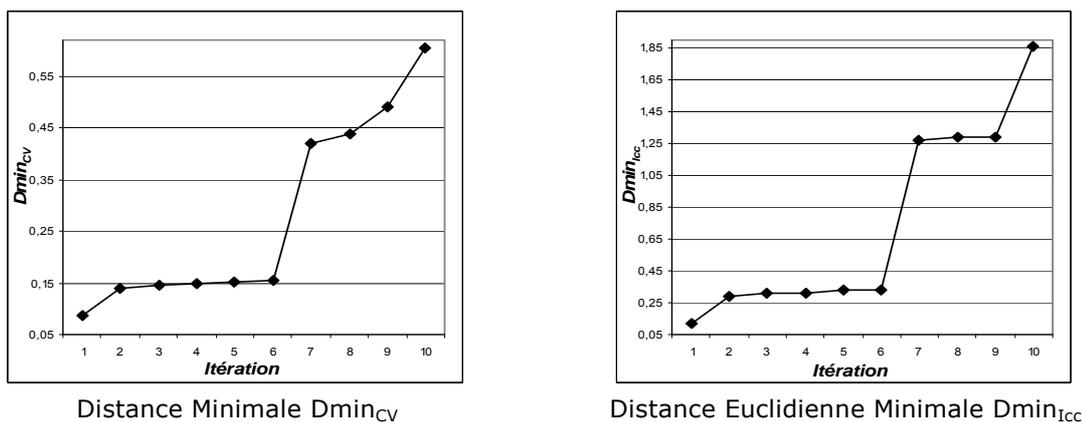


Figure 4.18. Comparaison de la distance minimale indices CV et Icc, ensemble de données 3

La distance minimale à chaque itération est aussi très semblable dans les deux cas. On a donc dans ce cas un comportement très similaire pour les deux indices.

4.3 Processus de production de propylène glycol.

4.3.1 Description du procédé.

Le propylène glycol est un liquide clair, sans couleur, légèrement sirupeux à la température ambiante. Il peut exister dans l'air sous forme gazeuse, bien que le propylène glycol doive être chauffé ou vivement secoué pour produire un gaz. Le propylène glycol est pratiquement inodore et insipide. Ce composé est employé comme antigel dans les solutions de dégivrage pour les voitures, les avions, les bateaux ; utilisé pour la fabrication de composés polyesters et comme dissolvant dans les industries de peinture et de plastiques.

The Food and Drug Administration (FDA) a classé le propylène glycol comme additif sûr pour un usage alimentaire. Il est employé pour absorber l'excès d'eau et pour maintenir l'humidité dans certains médicaments, produits de beauté ou produits alimentaires. C'est un dissolvant pour certains colorants alimentaires et arômes. Le propylène glycol est également employé pour créer la fumée ou le brouillard artificiel utilisé dans la lutte contre l'incendie et dans des productions théâtrales. Le propylène glycol affecte la chimie du corps humain en augmentant la quantité d'acide dans le corps, ayant pour résultat des problèmes métaboliques. Cependant, de grandes quantités de propylène glycol sont nécessaires pour causer cet effet (Agency for Toxic Substances and Disease Registry, [ATSDR]) [ORA05].

L'objectif du processus est de produire du propylène glycol par réaction de l'oxyde de propylène et de l'eau. Le fonctionnement du procédé a été simulé sous HYSYS [ORA07a](simulateur dynamique de procédés commercial, développé par AEA Technology Engineering Software Products, 2002 [HYSIS]) et son schéma est donné sur la Figure 4.19. Le procédé est composé d'un mélangeur MIXER1 qui permet de

mélanger l'oxyde de propylène et l'eau. Ce mélange rentre ensuite dans le réacteur chimique où la production de propylène glycol s'effectue suivant la réaction :



Les produits et réactifs liquides non consommés présents dans le courant (REACTOR PRODS) sont envoyés dans la colonne de distillation pour y être séparés. Cette colonne a pour but de récupérer au maximum (en pied de colonne) le propylène glycol. Elle comporte 10 étages théoriques et est aussi munie de deux boucles de régulation par PID qui régulent la pression de la colonne en agissant sur le refroidissement du condenseur (CondDuty) et la température du plateau 9 (avant-dernier plateau) par action sur la quantité d'énergie fournie au bouilleur (RebDuty). L'emplacement de ce capteur de température a été choisi après une analyse du fonctionnement de la colonne en régime permanent qui a montré que la sensibilité de la température à une variation de la quantité de chaleur fournie au bouilleur était la plus importante à cet endroit.

L'alimentation de cette colonne est donc composée de 3 constituants : l'eau, le propylène glycol, l'oxyde de propylène. Le courant de sortie en tête de colonne (RECYPRODS) est composé en grande partie d'eau et d'oxyde. Celui de fond (GLYCOL), même s'il est composé en grande partie de propylène glycol, contient encore de l'eau et de l'oxyde. La qualité du produit obtenu est donnée par la fraction de propylène glycol dans ce courant, l'objectif étant de récupérer le maximum de propylène en sortie.

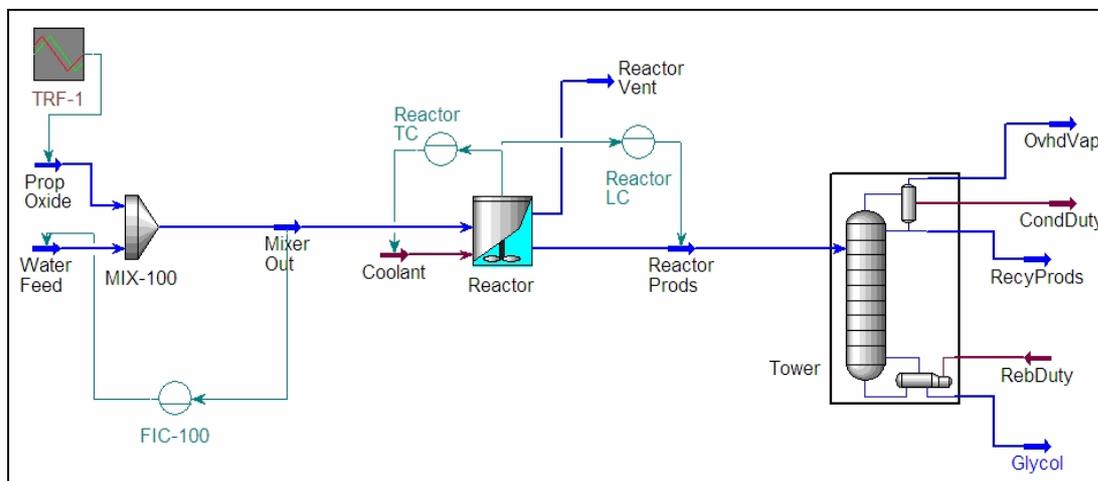


Figure 4.19. Schéma de l'unité de production de propylène glycol [ORA05]

Pour l'application de la méthode d'optimisation, nous avons utilisé le même ensemble de données d'apprentissage que celui qui a été employé pour la méthode de placement capteurs [ORA07a]. Des perturbations/défaillances ont été appliquées à différents points du procédé chimique. Les différentes perturbations simulées sont données dans le Tableau 4.3., avec l'intervalle des valeurs pour lesquelles le système

est capable de revenir en opération normale, ainsi que la valeur nominale de chaque variable perturbée.

Nom de la perturbation	Intervalles de variation			Unités
	Minimum	Nominal	Maximum	
Variation de l'Oxyde de Propylène (Oxyde)	50	68.04	80	Kgmole/h
Variation de la température du fluide de refroidissement du réacteur (TinletCool)	11	15	25	°C
Variation température du liquide refroidissement de la colonne (TinletCond)	11	15	19	°C
Modification de la vitesse de la réaction : Variation du facteur de fréquence (Freq)	1.20E+13	1.7E+13	2.2E+13	s/u

Tab. 4.3. – Intervalles de variations des perturbations pour lesquels le système reste stable.

A partir des perturbations données dans le Tableau 4.3., 8 défauts correspondant (1.↑Oxyde, 2.↓Oxyde, 3. ↑TinletCool, 4. ↓TinletCool, 5. ↑TinletCond, 6. ↓TinletCond, 7.↑Freq., 8. ↓Freq.) aux 8 valeurs extrêmes des perturbations ont été simulés. Chaque perturbation/défaillance est constituée de deux parties : la hausse puis la baisse en amplitude de la variable perturbée. Les perturbations ont été choisies de manière à ce que le système puisse revenir d'un état de défaillance vers un fonctionnement normal et se maintenir stable. Cette simulation a fourni 6337 points de mesure qui constituent les individus à classer.

4.3.2 Application de la méthode d'optimisation avec tous les descripteurs

Les descripteurs utilisés dans la première étape de la méthodologie proposée dans [ORA07a] comprennent des capteurs usuels: de température, de pression, de niveau et aussi de débits placés sur les appareils et le % d'ouverture des vannes. Les capteurs de concentration n'ont pas été inclus puisqu'en pratique, il est très difficile et onéreux de trouver un moyen de mesurer en-ligne la concentration d'un constituant dans un mélange multi-constituant. Ces travaux montrent qu'un placement correct de certains capteurs de température permet d'avoir accès à une défaillance affectant la qualité du produit fabriqué sur cette installation sans avoir recours à une mesure de cette qualité en ligne. Le Tableau 4.4. correspond à la liste des descripteurs possibles.

Numéro Descrip.	Nom	Numéro Descrip.	Nom	Numéro Descrip.	Nom
1	Pression Réacteur (Kpa)	8	Température plateau 2 (°C)	15	Température plateau 9 (°C) (Variable régulée)
2	Température Réacteur (°C) (Variable régulée)	9	Température plateau 3 (°C)	16	Température plateau 10 (°C)
3	Niveau Liquide Réacteur (%) (Variable régulée)	10	Température plateau 4 (°C)	17	Puissance délivrée au bouilleur (Kj/h) (Variable manipulée)
4	Vanne pour régler le débit de Refroidissement Réacteur (%) (Variable manipulée)	11	Température plateau 5 (°C)	18	Vanne pour régler le débit de Refroidissement Condenseur (%) (Variable manipulée)
5	Vanne pour régler le Niveau du Liquide Réacteur (%) (Variable manipulée)	12	Température plateau 6 (°C)	19	Pression Condenseur (Variable régulée)
6	Température de sortie du fluide Refroidissement Réacteur (°C)	13	Température plateau 7 (°C)	20	Température de sortie du fluide de refroidissement du condenseur (°C)
7	Température plateau 1 (°C)	14	Température plateau 8 (°C)	21	Température sortie du Glycol (°C)

Tab. 4.4. – Liste des descripteurs possibles pour le processus de propylène glycol.

Nous avons utilisé l'ensemble de données constitué par les 21 descripteurs possibles en appliquant les perturbations décrites dans la section précédente. En utilisant la méthode de classification LAMDA avec apprentissage non supervisé et les mêmes paramètres de l'algorithme employé par Orantes [ORA05] et Kempowsky et Le Lann [KEM06], on obtient la classification initiale qui servira de point de départ pour l'application de la méthode d'évaluation et d'optimisation que nous proposons. La Figure 4.20.a. montre la classification initiale obtenue avec les données des 21 descripteurs possibles. Les valeurs des variables normalisées sont présentées sur la Figure 4.20.b.

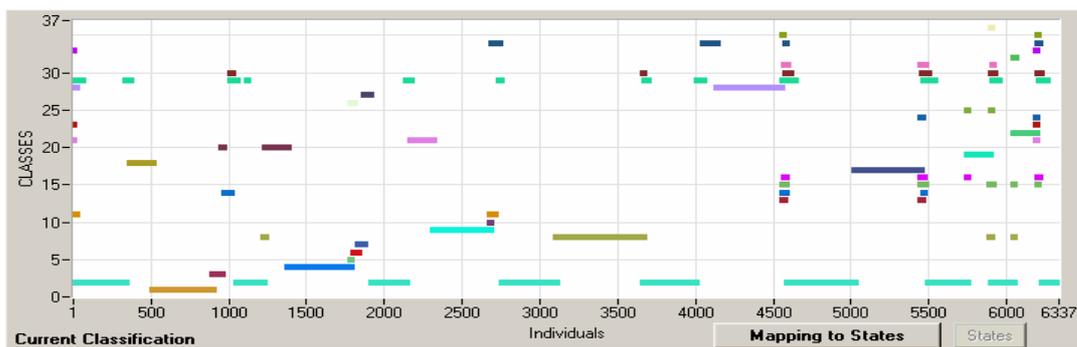


Figure 4.20.a. Classification initiale

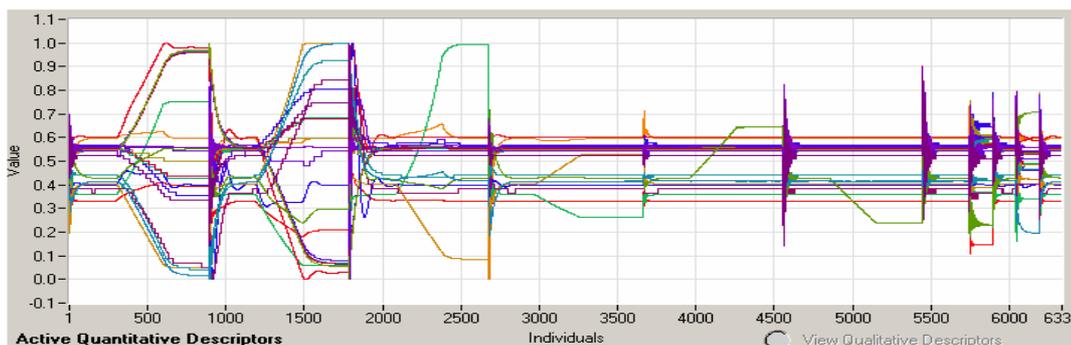


Figure 4.20.b. Variables du procédé

Figure 4.20. Classification initiale -21 descripteurs, Procédé de production du propylène glycol

Le nombre de classes identifiées est de 36. Ce nombre est très élevé pour établir le modèle de comportement utile pour le diagnostic. Normalement la procédure à suivre est de regrouper et d'associer les classes à des états [ORA05][KEM06]. Dans les travaux précédents, cette association a été faite uniquement à partir de l'analyse de l'expert du procédé. Avant de donner cette partition à l'expert du procédé, nous proposons d'améliorer cette partition en utilisant la méthode d'évaluation et d'optimisation de partitions (Chapitre 3), de sorte à fournir à l'expert une partition optimale ainsi qu'une évaluation quantitative de la qualité de la partition. La Figure 4.21. présente l'évolution de l'indice de qualité des partitions proposé (CV) au fur et à mesure des itérations.

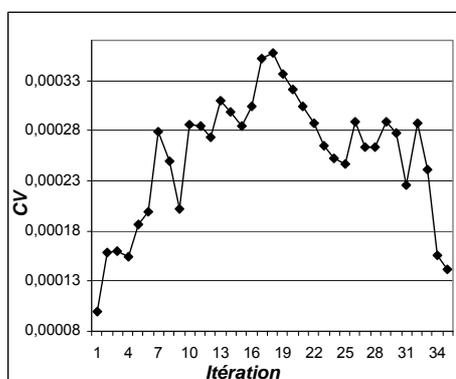


Figure 4.21. Indice de qualité des partitions CV -21 descripteurs, Procédé de production du propylène glycol

La meilleure partition a été obtenue pour l'itération 18. La Figure 4.22. correspond à la classification optimale suivant l'indice CV.

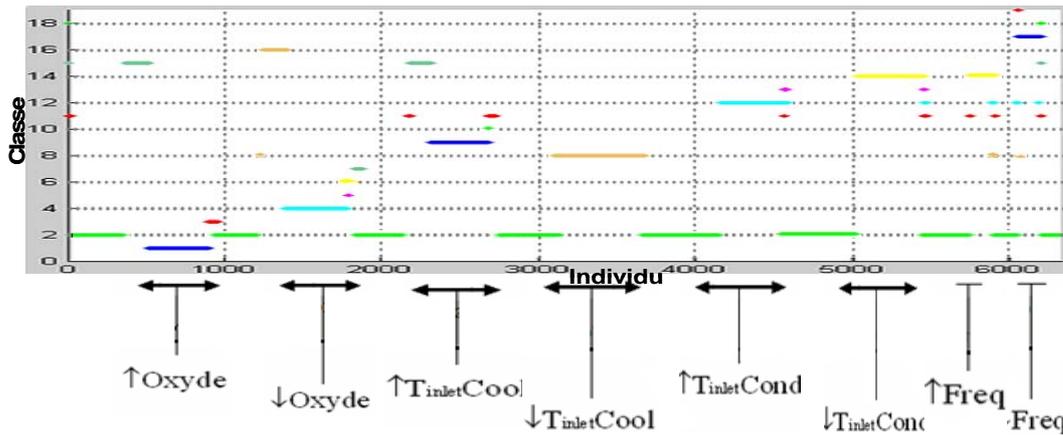


Figure 4.22. Classification optimale -21 descripteurs, Procédé de production du propylène glycol

Cette partition permet d'associer automatiquement la plupart des classes à des états fonctionnels. Nous avons ainsi défini automatiquement 19 classes qui permettent d'identifier les situations de défaillance. Mais, les deux classes associées aux défaillances : diminution de la température du liquide de refroidissement du condenseur (*TinjetCond Bas*) et augmentation de la vitesse de réaction (facteur de fréquence dans la loi d'Arrhenius, (*Freq Haute*)) ont été réunies par l'algorithme parce que les caractéristiques de ces deux classes sont très semblables, elles se manifestent par les mêmes symptômes. Cette partition permet aussi d'identifier, pour certaines situations, les dérives, i.e. les situations d'alarme vers un défaut, ainsi que des transitoires vers le fonctionnement normal. La détection au plus tôt de ces dérives est primordiale pour pouvoir réagir avant l'installation de la défaillance elle-même.

Dans les travaux concernant le placement de capteurs, la classification initiale (Figure 4.20.) a été analysée par l'expert du procédé pour établir les états fonctionnels. Celui-ci a regroupé les classes, de manière empirique en fonction de son expertise sur le procédé et a établi un modèle de comportement avec 16 états, mais il a regroupé plusieurs classes dans un état appelé "NON STATE" (numéro zéro). Pour ces classes, il n'a pas trouvé de relation exploitable avec des états du procédé. La partition proposée par l'expert est présentée sur la Figure 4.23.

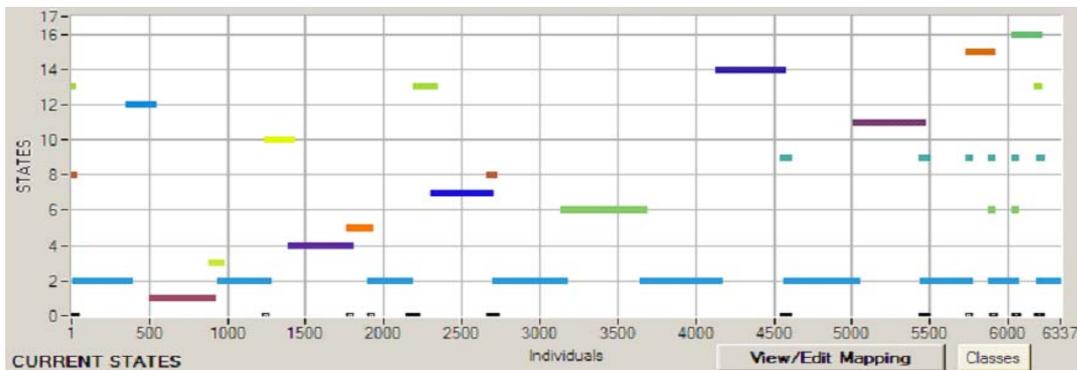


Figure 4.23. Classification proposée par l'expert -21 descripteurs, Procédé de production du propylène glycol

Nous considérons que regrouper les classes non identifiées, peut être considéré comme une perte d'information. Nous avons donc associé un état à chaque classe qui appartient à l'état "NON STATE" de la partition proposée par l'expert. Nous utilisons cette partition comme référence pour la comparer avec la partition optimale obtenue automatiquement (Figure 4.22). La Figure 4.24 correspond aux états établis par l'expert mais sans l'état "NON STATE", les données appartenant à cet état sont marquées par un carré. De cette manière, la partition de référence a 25 classes.

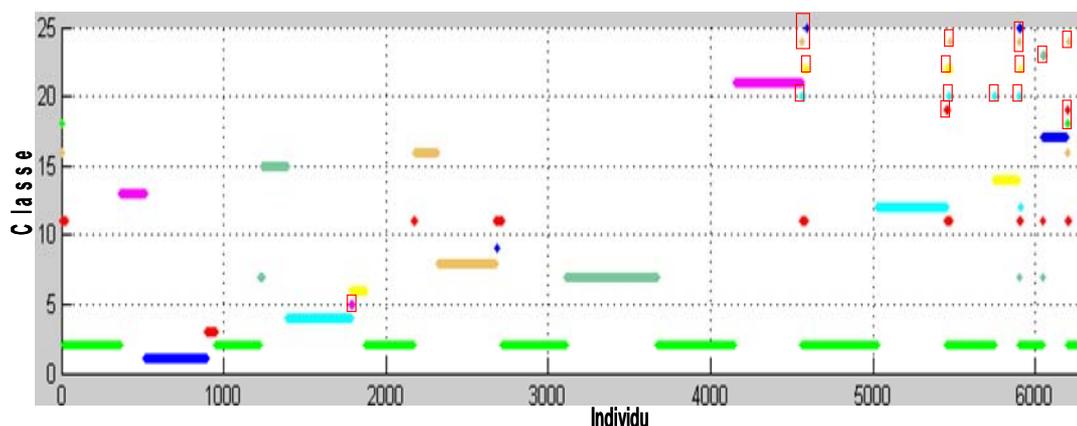


Figure 4.24. Classification de référence -21 descripteurs, Procédé de production du propylène glycol

Pour comparer les deux partitions, nous avons utilisé la mesure de similitude des partitions proposée par Lopez de Mantaras [LOP77]. La première comparaison a été établie en calculant la Matrice de Contingence (Chapitre 3) entre les deux partitions:

		Partition de Référence																									
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
Partition Automatique	1	0,06	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0,44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,002	0	6E-04	8E-04
	3	0	0	0,01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0,062	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	5E-04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	6	0	0	0	0	0	0,01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	7	0	0	0	0	0	0,01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	8	0	0	0	0	0	0	0,09	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	9	0	0	0	0	0	0	0	0,06	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	10	0	0	0	0	0	0	0	0	9E-04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	11	0	0	0	0	0	0	0	0	0	0,01	0	0	0	0	0	0	0	0	0	0	6E-04	0,001	0	0	0	0
	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,065	0	0	0	0
	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	14	0	0	0	0	0	0	0	0	0	0	0,07	0	0,02	0	0	0	0	0	0	0	0	0	0	0	0	0
	15	0	0	0	0	0	0	0	0	0	0	0	0,02	0	0	0,02	0	0	0	0	0	0	0	0	0	0	0
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0,024	0	0	0	0	0	0	0	0	0	0	0	0
	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,022	0	0	0	0	0	0	0	0	0
	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5E-04	0	0	0	0	0	0	0
	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9E-04	0	0

Tab. 4.5. – Matrice de contingence-21 descripteurs, Procédé de production du propylène glycol.

En permutant des colonnes, la matrice de contingence n'est pas exactement une matrice "quasi-diagonalisable" (il y a quelques valeurs hors de la diagonale) (Chapitre 3). En conséquence, les partitions ne sont pas "compatibles" mais pour la plupart des classes il y a une correspondance entre les deux partitions. L'indice de dissemblance qui permet d'établir la différence entre les deux partitions (Chapitre 3), est calculé à

partir de cette matrice, sa valeur est de 0.0131. Cet indice varie entre 0 et 1. Quand les deux partitions comparées sont exactement égales ou dites encore "compatibles" l'indice est 0 [LOP77]. On peut déduire de la valeur de cet indice, que les deux partitions sont très semblables.

Nous avons identifié, en utilisant un critère de qualité et de façon automatique, un modèle de comportement du procédé de propylène glycol très semblable à celui qui avait été établi par l'expert du procédé.

En renommant les classes de la partition automatique, on peut illustrer la similitude entre les partitions, la correspondance des classes entre les deux partitions et l'assignation d'états, comme le montre le Tableau suivant. La Matrice de Contingence (Tab. 4.5) a été utilisée pour renommer les classes, chaque classe de la partition de référence a été associée avec la classe de la partition automatique ayant la valeur la plus grande de probabilité de l'intersection ($P(c_i \cap f_j)$, eq.3.37).

CLASSES REFERENCE	CLASSES AUTOMATIQUE	N. ETAT [KEM06]
1	1	H-OXYDE
2	2	NORMAL
3	3	REC_H-OXY_N
4	4	B-OXYDE
5	5	SI1
6	6	REC_B-OXY_N
7	8	B-TCOOL
8	9	H-TCOOL
9	10	REC_H-COOL_N
10	SANS ELEMENTS	
11	11	REC-COND_N
12	14	B-TCOND
13	15	AL_H-OXYDE
14	14	H-FREQV
15	16	AL_B-OXYDE
16	15	AL_H-TCOOL
17	17	B-FREQV
18	18	SI4
19	11	SI5
20	11	SI6
21	12	H-TCOND
22	2	SI7
23	19	SI8
24	2	SI9
25	2	SI10

Tab. 4.6. – Correspondance des classes et états-21 descripteurs, Procédé de production du propylène glycol.

Les états appelés SI#, correspondent aux classes incluses par l'expert dans l'état "NON STATE". Dû le type d'apprentissage et la reconnaissance à *posteriori* des éléments avec la méthode LAMDA, on peut avoir des classes sans aucun élément. Les éléments d'une classe créée lors d'une classification initiale peuvent être ensuite classés dans une autre classe et laisser la première sans élément. C'est ainsi qu'on ne

tiendra pas compte de la classe 10 pour le modèle de comportement du procédé. La Figure 4.25. présente une comparaison des deux partitions une fois que les classes ont été renommées.

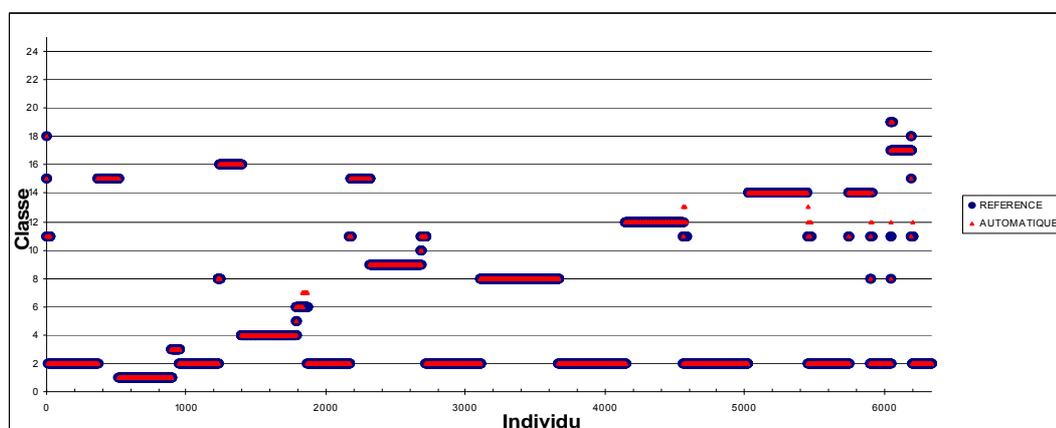


Figure 4.25. Comparaison des classifications optimale et de référence -21 descripteurs, Procédé de production du propylène glycol

Avec cette méthode, nous avons fourni une partition facilement associable aux états du procédé. Cette partition est très semblable à celle proposée par l'expert du procédé après une analyse difficile et exhaustive de la partition initiale. Ainsi nous contribuons à l'amélioration de la méthode de placement des capteurs consistant à choisir les descripteurs les plus significatifs d'un point de vue de la quantité d'information issue du procédé.

4.3.3 Application de la méthode d'optimisation avec les descripteurs sélectionnés.

La deuxième étape de la méthode proposée dans [ORA07a], consiste à déterminer les descripteurs les plus importants pour chaque état qui a été identifié dans l'étape précédente. En utilisant une analyse de la quantité d'information apportée par chaque descripteur à chaque état, 7 descripteurs (capteurs) ont été considérés nécessaires pour identifier le modèle de comportement du procédé [ORA05][KEM06]. Le Tableau 4.7. présente les capteurs tels qu'ils ont été sélectionnés:

Numéro Descrip.	Nom	Numéro Descrip.	Nom	Numéro Descrip.	Nom
1	Pression Réacteur (Kpa)	5	Vanne pour régler le Niveau du Liquide Réacteur (%) (Variable manipulée)	18	Vanne pour régler le débit de Refroidissement Condenseur (%) (Variable manipulée)
4	Vanne pour régler le débit de Refroidissement Réacteur (%) (Variable manipulée)	6	Température de sortie du fluide de Refroidissement de Réacteur (°C)	20	Température de sortie du fluide de refroidissement du condenseur (°C)

Tab. 4.7. – Descripteurs sélectionnés pour le procédé de production du propylène glycol.

La température sur le plateau 10 de la colonne de distillation a été ajoutée comme descripteur. Kemposky et Le Lann [KEM06] ont montré qu'en ajoutant ce descripteur (*10_MainTS*), il est possible de mieux détecter les perturbations inconnues *a priori*.

En utilisant uniquement ces 7 descripteurs, la troisième étape de la méthode de sélection des capteurs, consiste à établir le modèle de comportement du procédé. Pour ceci, la procédure est similaire à celle réalisée pour obtenir la première classification. Le scénario de comportement inclut 8 défauts. L'ordre d'application des défauts est le même que dans le cas avec les 21 descripteurs : variation de l'oxyde d'éthylène, variation de la température du liquide de refroidissement du réacteur, variation de la température du liquide de refroidissement du condenseur et variation de la vitesse de réaction (variation du facteur de fréquence dans la loi d'Arrhenius).

Nous présentons la partition initiale obtenue (Figure 4.26.a.) avec apprentissage non supervisé et en utilisant le même ensemble de données et les mêmes paramètres de LAMDA que ceux qui ont été employés dans les travaux précédents. Les évolutions des 7 descripteurs sélectionnés sont données sur le graphe du bas.

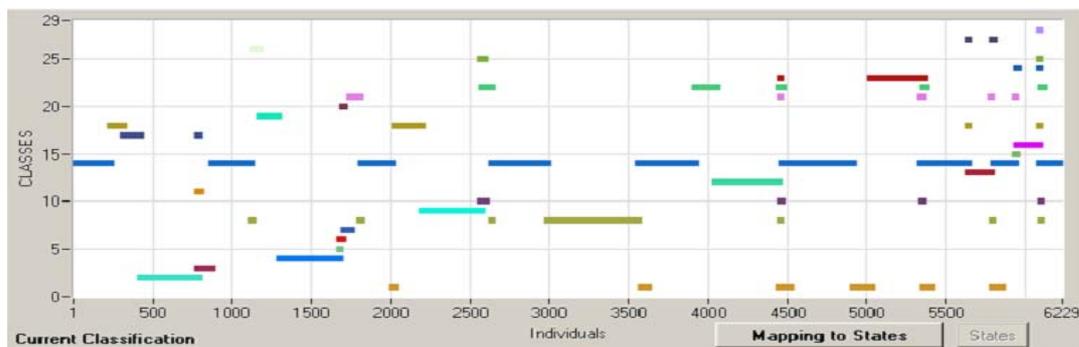


Figure 4.26.a. Classification initiale

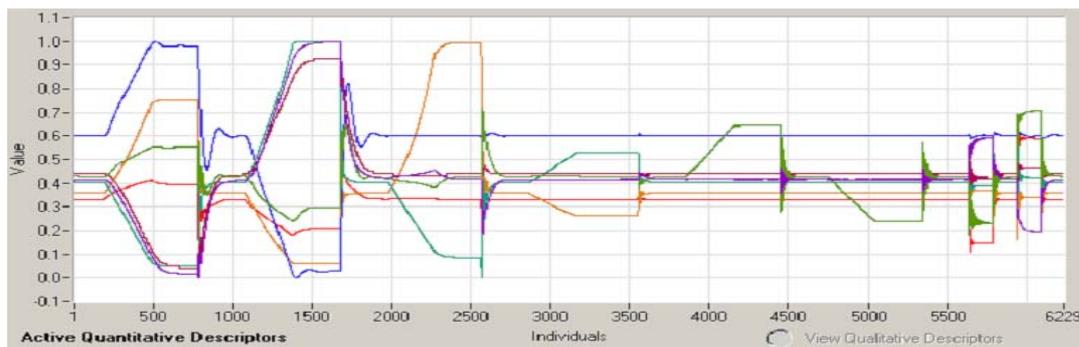


Figure 4.26.b. Variables sélectionnées

Figure 4.26. Classification initiale -7 descripteurs, Procédé de production de propylène glycol

De la même manière que pour le cas avec 21 descripteurs, nous utilisons la méthode d'évaluation et d'optimisation des partitions pour donner une partition plus facilement utilisable pour établir les états fonctionnels. Nous avons comparé avec le modèle d'états proposé par l'expert mais comme précédemment, sans y inclure l'état

qualifié de "NON STATE". La Figure 4.27. présente l'évolution de l'indice de qualité des partitions CV.

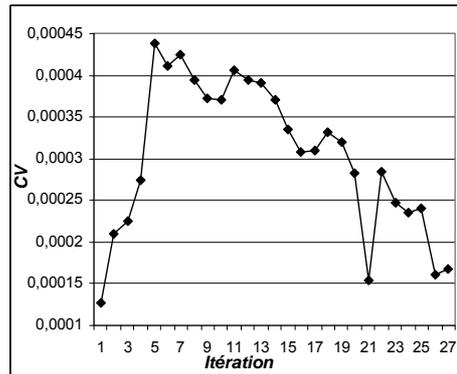


Figure 4.27. Indice de qualité des partitions CV -7 descripteurs, Procédé de production de propylène glycol

Nous obtenons la meilleure partition pour l'itération 5. La classification optimale en fonction de l'indice CV est présentée sur la Figure 4.28.

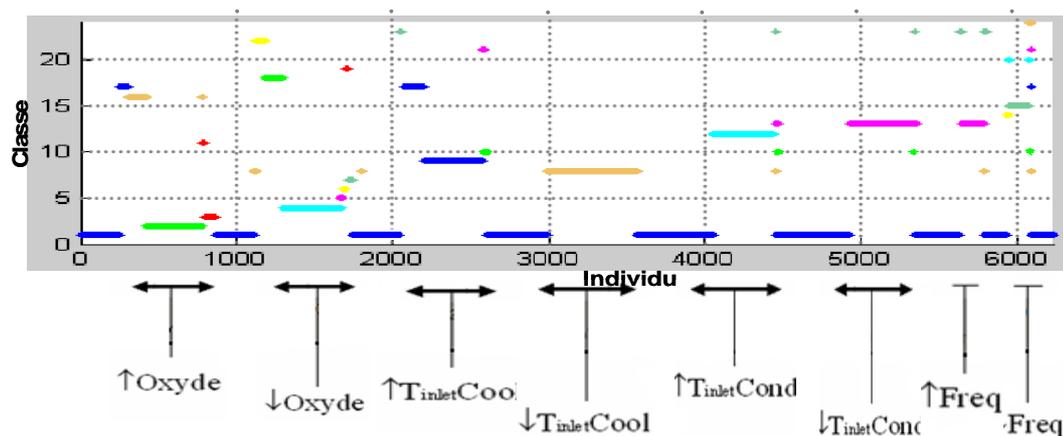


Figure 4.28. Classification optimale -7 descripteurs, Procédé de production de propylène glycol

La méthode d'optimisation proposée permet de réunir les classes pour obtenir une classification plus simple, et donc plus facile à interpréter. Mais elle permet aussi de garder des classes très petites qui peuvent correspondre à des états de transition et qui peuvent constituer des "états d'alarmes" coïncidant avec une dérive ou un état de "pré-défaillance", états qu'il est crucial de détecter dans le cas de procédés industriels (ceci permet généralement de déclencher des maintenances préventives par exemple). Dans ce cas, la méthode d'optimisation n'a pas réuni beaucoup de classes, ceci montre que les "petites" classes (réunies par l'expert dans l'état "NON STATE") sont bien caractérisées et que donc, elles peuvent être conservées comme telles pour représenter des états de transition ou d'alarme non considérés d'abord par l'expert. Lorsque il n'y a plus regroupement de classes sous l'état "NON STATE", on trouve un nombre de groupes pratiquement identique à celui de la partition de référence (23 groupes, Figure 4.29.). La partition optimale possède 24 classes. Cette partition permet d'associer les états fonctionnels correspondants aux défaillances et états de transition et d'alarmes.

Dans ce cas, les classes associées aux défaillances : *TinjetCond Bas* et *Freq Haute* ont été elles aussi réunies.

De manière similaire au cas avec 21 descripteurs, nous avons comparé la partition optimale (Figure 4.28) avec la partition de référence. Rappelons que la partition de référence a été obtenue en utilisant le modèle de comportement donné par l'expert du procédé mais sans inclure l'état "NON STATE". Cette partition est présentée sur la Figure 4.29. Les classes qui avaient été regroupées par l'expert dans l'état "NON STATE" sont marquées par un carré.

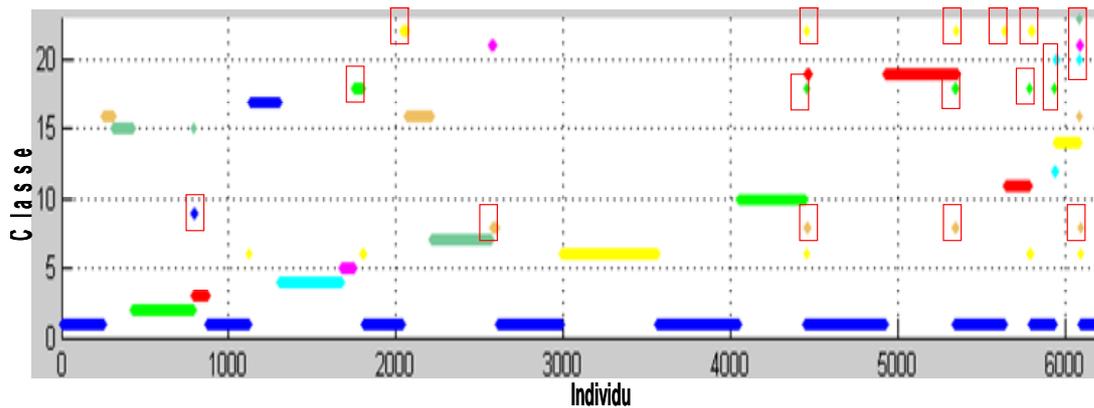


Figure 4.29. Classification de référence -7 descripteurs, Procédé de production du propylène glycol

La Matrice de Contingence (Chapitre 3, [LOP77]) permet de comparer les deux partitions et de constater qu'elles sont très semblables. Cette matrice n'est pas quasi-diagonalisable, parce que les partitions ne sont pas parfaitement égales. Nonobstant, l'indice de dissemblance normalisé (0.0442) met en évidence que les partitions sont très semblables. Les classes qui sont marquées en gris ont des éléments communs avec plus d'une classe de la partition de référence, ceci ne permet pas d'obtenir une matrice quasi-diagonalisable.

		Partition de Référence																							
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
Partition Automatique	1	0,424	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,009	0	0	0	0	0	
	2	0	0,0581	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0,022	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0,0599	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	0,00048	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	6	0	0	0	0	0,00225	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	7	0	0	0	0	0,00626	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	8	0	0	0	0	0	0,0946	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	9	0	0	0	0	0	0	0,05795	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	10	0	0	0	0	0	0	0	0,0047	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	11	0	0	0	0	0	0	0	0	0,0012	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	12	0	0	0	0	0	0	0	0	0	0,064	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	13	0	0	0	0	0	0	0	0	0	0	0,023	0	0	0	0	0	0	0	0	0,068	0	0	0	0
	14	0	0	0	0	0	0	0	0	0	0	0	0,0008	0	0	0	0	0	0	0	0	0	0	0	0
	15	0	0	0	0	0	0	0	0	0	0	0	0	0,0215	0	0	0	0	0	0	0	0	0	0	0
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0,019	0	0	0	0	0	0	0	0	0	0
	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,0323	0	0	0	0	0	0	0	0	0
	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,01878	0	0	0	0	0	0	0	0
	19	0	0	0	0	0,00241	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,0013	0	0	0	0
	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,0022	0	0	0
	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,00995	0	0	0	0	0	0	0
	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,0063	0
	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,00032

Tab. 4.8. - Matrice de contingence-7 descripteurs, Procédé de production du propylène glycol.

En utilisant les probabilités de l'intersection (Matrice de Contingence, Tab.4.8), la correspondance des classes entre les deux partitions et l'assignation des états (modèle de comportement) ont été établies, cette association est présentée sur le Tableau 4.9. Les états étiquetés SI#, correspondent aux classes incluses par l'expert dans l'état "NON STATE". En renommant les différentes classes, nous obtenons une classification semblable à celle proposée par l'expert. La Figure 4.30. présente le graphe de la comparaison des deux partitions.

CLASSES REFERENCE	CLASSES AUTOMATIQUE	N. ETAT [ORA05]
1	1	B-TCOND
2	2	H-OXYDE
3	3	REC_H-OXY_N
4	4	B-OXYDE
5	7	REC_B-OXY_N
6	8	B-TCOOL
7	9	H-TCOOL
8	10	REC-HAUSSE
9	11	H-TCOND
10	12	NORMAL
11	13	H-FREQV
12	14	B-FREQV
13	SANS ELEMENTS	
14	15	AL_H-OXYDE
15	16	AL_B-OXYDE
16	17	SI1
17	18	SI2
18	1	SI3
19	13	SI4
20	20	SI5
21	21	SI6
22	23	SI7
23	24	SI8

Tab. 4.9. - Correspondance de classe et états-7 descripteurs, Procédé de production du propylène glycol.

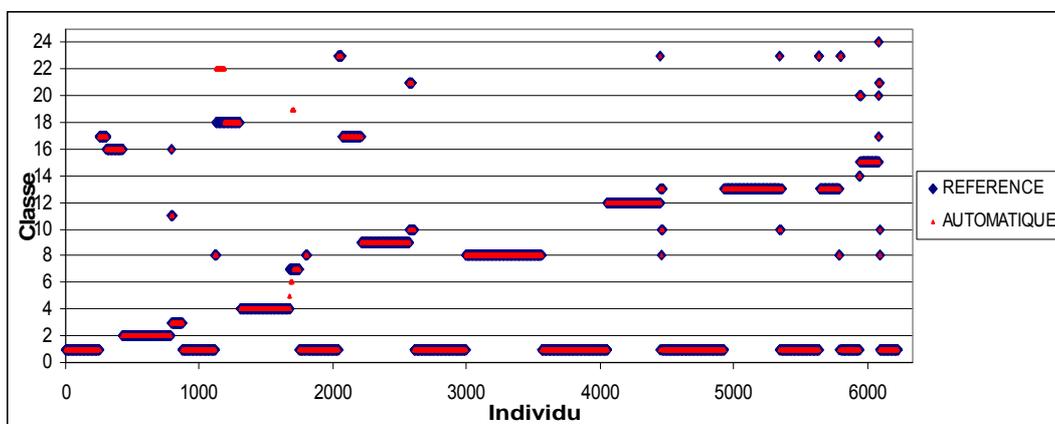


Figure 4.30. Comparaison des classifications optimale et de référence -7 descripteurs, Procédé de production du propylène glycol

En appliquant la méthode d'optimisation de la partition, nous avons, dans le cas de ce procédé, obtenu de manière automatique une classification que l'on peut qualifier d' "optimale" (en terme de compacité et de séparation des classes) qui est d'une aide très précieuse pour orienter l'expert du procédé dans l'étape d'établissement d'un modèle comportemental.

4.4 Réaction d'Estérification utilisant la technologie "Open Plate Reactor"

La mise en œuvre de réactions chimiques dans des réacteurs discontinus (batch) ou semi-continus est fortement limitée par des contraintes liées à l'évacuation de la chaleur engendrée par les réactions lorsqu'elles sont exothermiques. Un nouveau concept de réacteur intégrant les caractéristiques fondamentales des échangeurs de chaleur devrait permettre d'améliorer grandement la maîtrise des échanges thermiques dans le cas de réacteur fonctionnant en continu. C'est dans cet objectif que l'Open Plate Reactor (OPR) a été développé par le Laboratoire de Génie Chimique en partenariat avec ALFA-LAVAL [KEM06][DEV03].

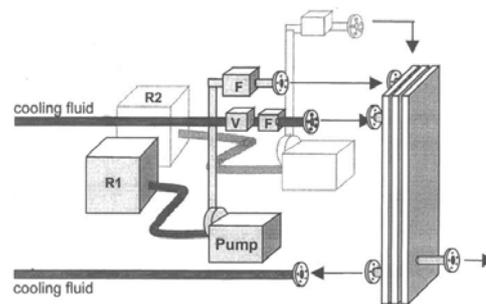


Figure 4.31. L' "Open plate reactor"

La Figure 4.31 montre l'unité pilote conçue et construite dans ce laboratoire autour de ce principe. Ce réacteur reprend les caractéristiques d'un échangeur de chaleur à plaques. Des équipements de mesure de température sont installés sur les admissions et les sorties du fluide. La température du procédé à l'extrémité du premier bloc est de même disponible. Le dispositif de mesure de pression est installé sur l'alimentation générale du réacteur. Toutes les mesures sont enregistrées sur un PC et une représentation graphique en ligne des variations de toutes les variables enregistrées est effectuée. La procédure de mise en régime de l'unité est la suivante : le fluide de refroidissement doit avoir circulé pendant une heure avant le début de l'expérience. Le réacteur est rempli d'eau distillée et une circulation est assurée pour éliminer les bulles d'air résiduelles. Les réactions sont effectuées avec un excès d'un des réactifs qui est présent au préalable dans le réacteur. Quand l'état d'équilibre est atteint, le deuxième réactif est injecté.

La méthode sur le placement de capteurs a été appliquée initialement sur des données expérimentales, résultats d'expériences classiques effectuées sur ce type

d'unité pilote. Quelques dysfonctionnements ou changements de conditions opératoires étaient présents dans ce scénario mais pas totalement représentatifs de défaillances majeures pouvant affecter le fonctionnement de cette unité [ORA05][KEM06]. C'est pourquoi, pour pouvoir diagnostiquer un certain nombre de défaillances dont certaines difficilement réalisables en pratique pour des raisons de sécurité, ces scénarii ont été générés à l'aide d'un simulateur du fonctionnement de l'unité pilote qui a été au préalable validé par le LGC [DEV03].

Dans les travaux cités, on a étudié d'une part la réaction d'estérification de l'anhydride propionique par le 2Butanol que nous analysons ici, ainsi que la réaction d'oxydation du thiosulfate de sodium ($\text{Na}_2\text{S}_2\text{O}_3$) par l'eau oxygénée (H_2O_2), en milieu homogène qui sera analysée dans la section suivante.

Pour ces deux réactions, nous avons appliqué la technique d'amélioration sur la partition obtenue une fois que les descripteurs les plus pertinents en terme de quantité d'information aient été déterminés par la méthode telle qu'elle a été décrite au paragraphe précédent. Nous présentons pour chaque procédé réactionnel, la comparaison entre la partition optimale obtenue automatiquement et la partition proposée par l'expert. Dans les deux cas, le nombre de classes de la partition initiale est très élevé et l'analyse de l'expert a été effectuée sans aucune idée de la qualité de la partition, en se basant uniquement sur des connaissances de la physique et des phénomènes mis en jeu sur ce procédé. Il est à noter que cette étape cruciale demande une connaissance très approfondie du fonctionnement de l'installation et que, la méthode d'optimisation de la partition que nous avons développée, apporte à l'expert une aide précieuse permettant d'établir les états fonctionnels sur une donnée quantifiée de la qualité de la partition et non plus uniquement sur une connaissance abstraite du procédé.

La réaction d'estérification est une réaction lente et faiblement exothermique. Pour l'accélérer, il faut chauffer le milieu réactionnel. Dans ce cas, le fluide "utilité" sert de fluide de chauffage. Il y a moins de contraintes de sécurité que dans le cas de la réaction du thiosulfate [ORA05].

L'estérification de l'anhydride propionique par le 2-Butanol, est largement citée dans les publications relatives à la sécurité des procédés physico-chimiques. Cette réaction est utilisée dans le cadre de la validation expérimentale d'études théoriques concernant par exemple la prédiction d'emballement thermique au sein de réacteur batch ou la détermination rapide de données cinétiques [GALV96].

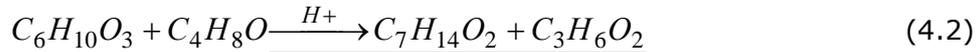
Plus récemment, cette réaction a également été utilisée pour tester une nouvelle génération du simulateur dynamique permettant la mise en place d'une stratégie de contrôle thermique performante [ORA05].

Cette réaction présente certains avantages :

- Elle s'effectue en milieu homogène [BEN91].
- Elle est relativement simple à mettre en œuvre.

- Elle est modérément exothermique.
- Le comportement cinétique change selon que l'on opère avec ou sans catalyseur: la vitesse de réaction est du second ordre en l'absence d'acide fort et est de type auto-catalysée en présence d'acide sulfurique.

Le schéma réactionnel est le suivant :



Où



La technique de placement des capteurs a été appliquée à ce procédé. Dans la première étape, de la même manière que pour le procédé de production de propylène glycol, on a procédé à une assignation des classes aux états pour établir les descripteurs les plus pertinents. Des 27 capteurs possibles, une liste de 8 capteurs pertinents a été établie. Cette liste est présentée dans le Tableau 4.10.

Numéro Descrip.	Nom	Numéro Descrip.	Nom	Numéro Descrip.	Nom
13	Temp. Fluide procédé bloc 2 cellule 5	29	Pression fluide procédé en sortie	3	Temp. Fluide procédé bloc 1 cellule 1
32	Temp. Fluide utilité sortie cellule 1	4	Temp. Fluide procédé bloc 1 cellule 2	26	Débit fluide procédé en sortie
14	Temp. Fluide procédé bloc 2 cellule 10	5	Temp. Fluide procédé bloc 1 cellule 3		

Tab. 4.10. – Descripteurs sélectionnés pour la réaction d'estérification.

Une fois les descripteurs les plus pertinents (en termes de quantité d'information fournie) sélectionnés, l'optimisation de la partition a été effectuée sur les données en présence de défaillances et de perturbations. Au total, ce sont 16 défauts qui ont été appliqués au réacteur. Les défaillances ont été simulées sous forme de perturbations sur les températures et les débits du réactif principal (C_4H_8O), du réactif secondaire ou injecté ($C_6H_{10}O_3$) et sur le système de refroidissement (système utilité), ainsi que sur la composition des réactifs principal et secondaire.

Avec ces données, une partition initiale a été obtenue en utilisant l'algorithme LAMDA avec apprentissage non supervisé. La Figure 4.32 présente cette partition et l'évolution des descripteurs normalisés.

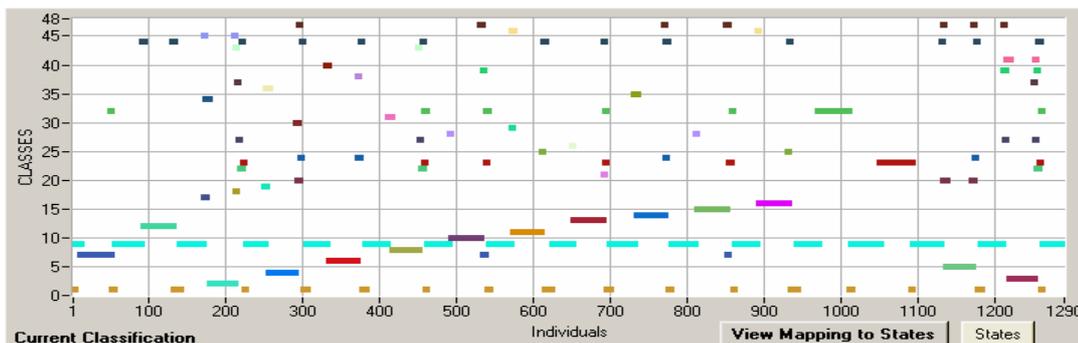


Figure 4.32.a. Classification initiale

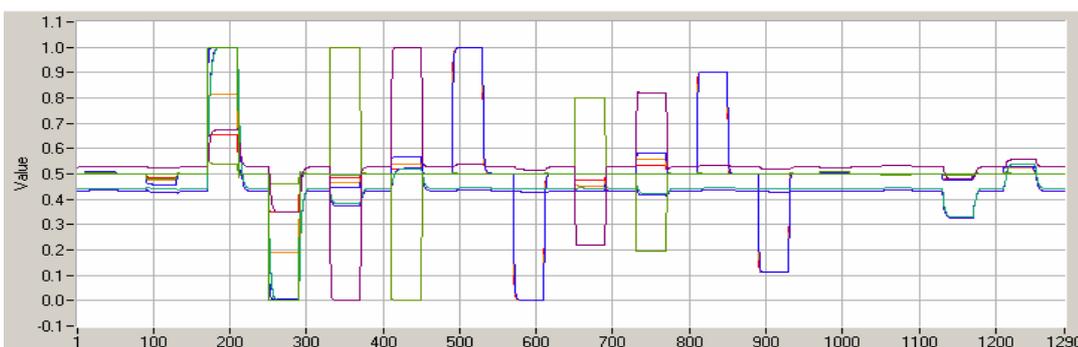


Figure 4.32.b. Variables sélectionnées

Figure 4.32. Classification initiale -8 descripteurs, Réaction d'estérification

Par cette classification, 47 classes ont été identifiées. En appliquant la technique d'optimisation de la partition, nous avons obtenu une partition optimale, avec des classes mieux définies. L'évolution de l'indice de qualité des partitions CV est présentée sur la Figure 4.33.

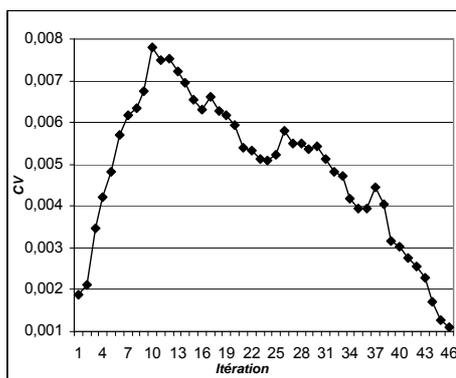


Figure 4.33. Indice de qualité des partitions CV -8 descripteurs, Réaction d'estérification

La meilleure partition a été obtenue à la 10^{ème} itération. La Figure 4.34 présente la classification optimale.

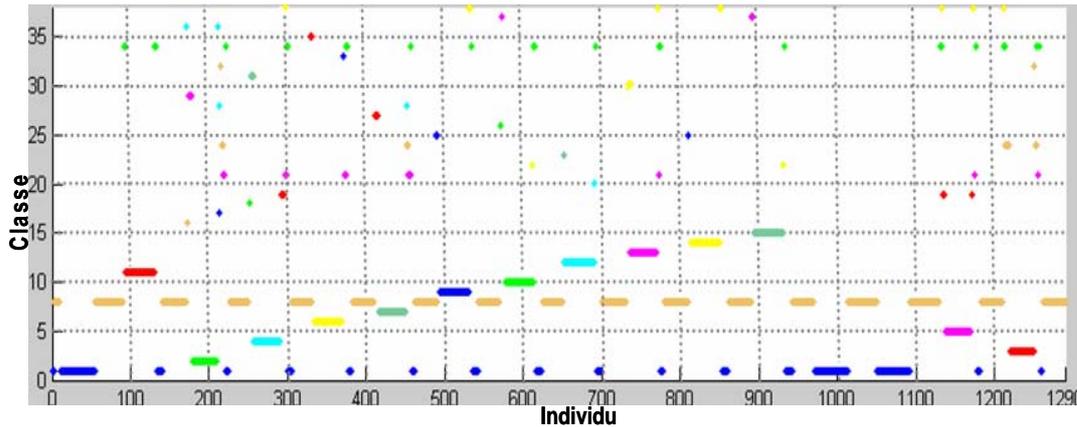


Figure 4.34. Classification optimale -8 descripteurs, Réaction d'estérification

Cette classification est constituée de 38 classes, elle permet de différencier les états normaux et de défaillance de la réaction, et permet aussi d'identifier des états d'alarme.

Nous avons effectué une comparaison avec les états proposés par l'expert du procédé. Dans les travaux de positionnement des capteurs, la partition initiale comportant 47 classes (Figure 4.32) a été analysée par l'expert pour regrouper les classes et identifier les états fonctionnels. Il a établi 19 états et un état appelé "NON STATE". Ce dernier état, comme dans le cas du procédé de production de propylène glycol, regroupe beaucoup de classes associées à aucun état précis du procédé.

Nous prenons comme référence la partition de l'expert mais nous avons transformé l'état "NON STATE" en plusieurs états correspondant chacun à une classe. La partition que nous employons comme référence pour comparer la partition optimale est présentée sur la Figure 4.35.

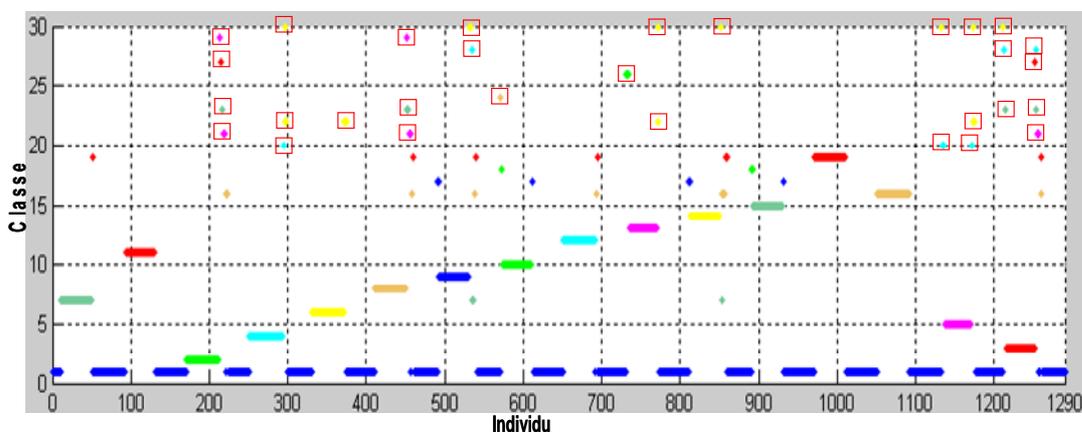


Figure 4.35. Classification de référence -8 descripteurs, Réaction d'estérification

La matrice de contingence (Annexe 1) permet de démontrer que les classifications sont proches même si elles ne sont pas égales. L'indice de dissemblance est de 0.0938

ce qui dénote une bonne similitude. La relation entre les deux partitions et les états associés est donnée dans le Tableau 4.11. La Figure 4.36 présente la partition de référence et la partition optimale.

REFERENCE	AUTOMATIQUE	N. ETAT [ORA05]	REFERENCE	AUTOMATIQUE	N. ETAT [ORA05]
1	8	NORMAL	16	1	↓X_MAIN
2	2	↑TU_B1_IN	17	25	Tendance ↑TP_IN
3	3	↓X_INJ	18	37	Tendance ↓TP_IN
4	4	↓TU_B1_IN	19	1	↑X_MAIN
5	5	↑X_INJ	20	19	SI1
6	6	↑FP_IN_1	21	21	SI2
7	1	↑FU_IN	22	21	SI3
8	7	↓FP_IN_1	23	24	SI4
9	9	↑TP_IN_1	24	26	SI5
10	10	↓TP_IN_1	25	SANS ELEMENTS	
11	11	↓FU_IN	26	30	SI7
12	12	↑FP_IN_2	27	32	SI8
13	13	↓FP_IN_2	28	34	SI9
14	14	↑TP_IN_2	29	28	SI10
15	15	↓TP_IN_2	30	38	SI11

Tab. 4.11. – Correspondance des classes et des états-8 descripteurs, Réaction d'estérification.

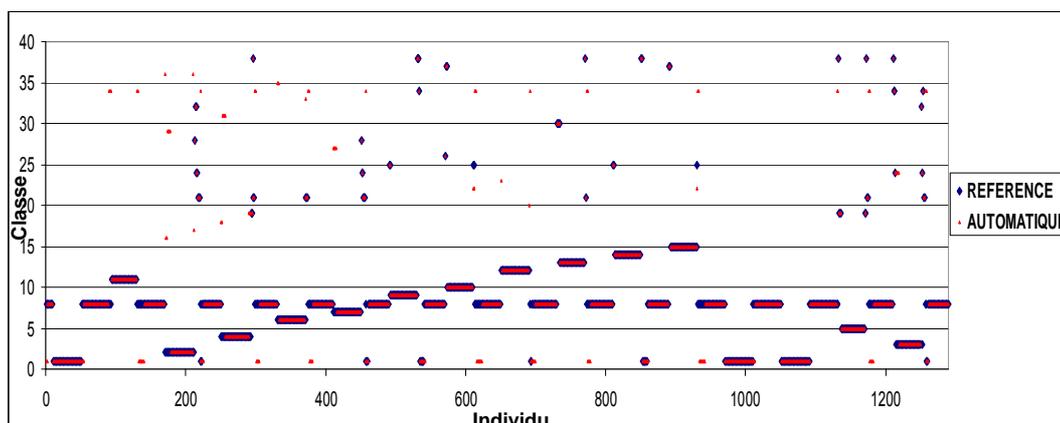


Figure 4.36. Comparaison des classifications optimale et de référence -8 descripteurs, Réaction d'estérification

4.5 Réaction du thiosulfate de sodium utilisant la technologie "Open Plate Reactor"

La réaction du thiosulfate de sodium avec l'eau oxygénée est une réaction très rapide, fortement exothermique et aussi très sensible à la température. Elle nécessite un refroidissement permanent par un fluide utilisé (de l'eau). La forte exothermicité entraîne de nombreuses contraintes de sécurité [ORA07b].

La réaction d'oxydation du thiosulfate de sodium avec l'eau oxygénée présente différentes caractéristiques : une stœchiométrie connue, elle est irréversible, rapide et fortement exothermique. L'eau oxygénée est utilisée en excès pour éviter toute réaction secondaire. L'énergie générée lors de la réaction est connue et vaut :

$$\Delta H = -586.2 \text{ KJ.mol}^{-1}$$

Le schéma réactionnel est le suivant :



La technique de placement de capteurs a été appliquée à la réaction du thiosulfate de sodium. De la même manière que pour le procédé de production de propylène glycol et que pour la réaction d'estérification, Orantes et al. [ORA07b] ont proposé les descripteurs les plus pertinents pour établir le modèle de comportement de la réaction et pouvoir identifier les états normaux et de défaillance. L'ensemble des données d'apprentissage est constitué des évolutions des variables. Dans les travaux précédents, les variables dont les conditions opératoires ont été modifiées ont été éliminées en tant que descripteurs afin de caractériser et identifier les situations de façon indirecte, et le démarrage de la réaction n'a pas été pris en compte. Ils ont établi que des 27 capteurs possibles, 9 capteurs sont pertinents. La liste des variables sélectionnées est présentée dans le Tableau 4.12.

Numéro Descrip.	Nom	Numéro Descrip.	Nom	Numéro Descrip.	Nom
14	Temp. Fluide procédé bloc 2 cellule 10 TP_B2_10	16	Temp. Fluide procédé bloc 2 cellule 20 TP_B2_20	3	Temp. Fluide procédé bloc 1 cellule 1 TP_B1_1
32	Temp. Fluide utilité sortie bloc 1 TU_B1_OU	26	Débit fluide procédé en sortie FP_OUT	6	Temp. Fluide procédé bloc 1 cellule 4 TP_B1_4
8	Temp. Fluide procédé bloc 1 cellule 10 TP_B1_10	29	Pression Fluide procédé en sortie PP_OUT	7	Temp. Fluide procédé bloc 1 cellule 5 TP_B1_5

Tab. 4.12. – Descripteurs sélectionnés pour la réaction du thiosulfate de sodium.

De la même façon que pour le procédé d'estérification, au total 17 défauts ont été simulés sur le réacteur. Les défaillances introduites sur le procédé sont :

- Variation de la température du fluide principal
- Variation de la température du fluide secondaire
- Variation du débit du fluide principal (L/h)
- Variation du débit du fluide secondaire (L/h)
- Variation du débit du fluide utilité (m³/h)
- Variation de la température du fluide utilité (m³/h)
- Fraction molaire du thiosulfate de sodium dans l'alimentation du fluide principal
- Fraction molaire de l'eau oxygénée dans l'alimentation du fluide secondaire

Pour chaque cas, deux types de variations ont été considérées : augmentation et diminution de la variable. Une dernière perturbation correspondant à un arrêt total du fluide utilisé a été ajoutée.

Nous avons utilisé comme partition initiale (Figure 4.37) la classification obtenue en appliquant LAMDA avec apprentissage non supervisé aux données d'apprentissage de la même manière que ce qui avait été fait dans les travaux précédents.

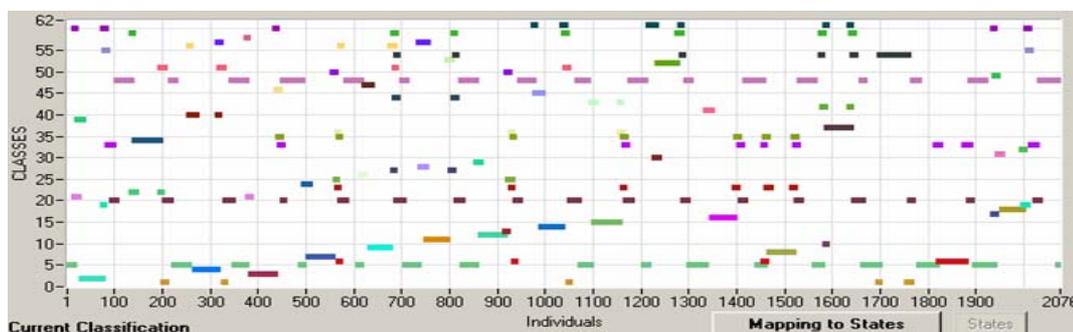


Figure 4.37.a. Classification initiale

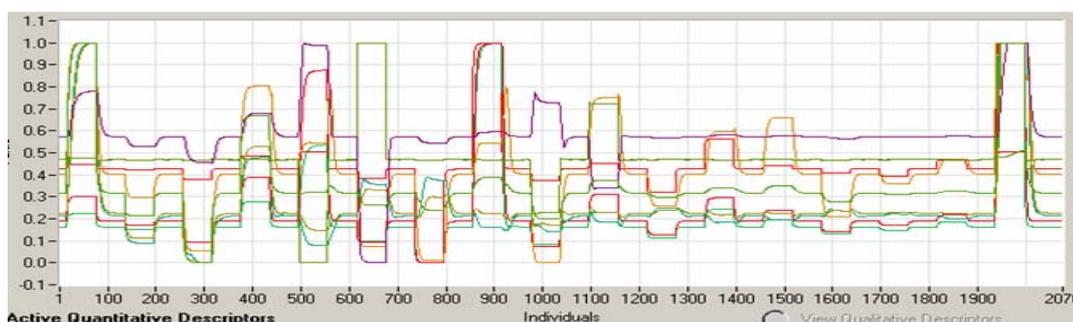


Figure 4.37.b. Variables sélectionnées

Figure 4.37. Classification initiale -9 descripteurs, Réaction du Thiosulfate de Sodium

Dans les travaux précédents, cette classification a été donnée à l'expert du procédé afin qu'il l'analyse et regroupe les classes. De cette manière, un modèle de comportement du système a été établi. Nous avons appliqué la technique d'amélioration des partitions pour fournir à l'expert une partition plus facilement interprétable. La Figure 4.38 correspond à l'évolution de l'indice CV au fur et à mesure des itérations.

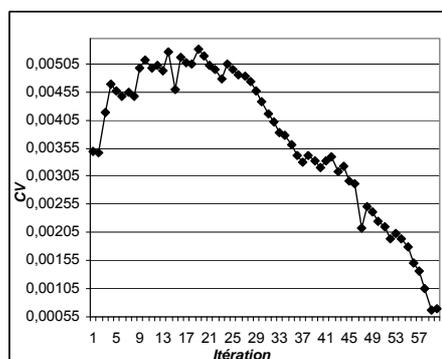


Figure 4.38. Indice de qualité des partitions CV -9 descripteurs, Réaction du Thiosulfate de Sodium

La meilleure partition (Figure 4.39) est obtenue à l'itération 19.

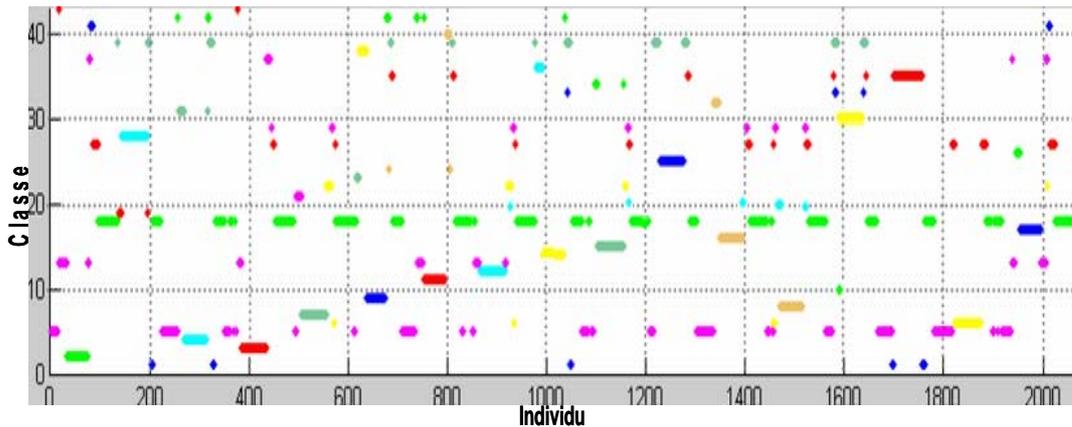


Figure 4.39. Classification optimale -9 descripteurs, Réaction du Thiosulfate de Sodium

De la même façon que pour la réaction d'estérification, la partition optimale permet plus facilement que la partition initiale, d'identifier les 17 défauts et le fonctionnement normal ainsi que des états transitoires qui peuvent être interprétés comme des alarmes.

De manière identique au cas de la réaction d'estérification, nous avons comparé les résultats de l'algorithme d'optimisation des partitions, avec la partition proposée par l'expert du procédé mais en utilisant une partition de référence qui inclut les états définis par l'expert mais dans laquelle l'état "NON STATE" a été représenté à chaque fois par une des classes. La partition de référence est représentée sur la Figure 4.40, les carrés représentent les classes correspondant à l'état "NON STATE" »

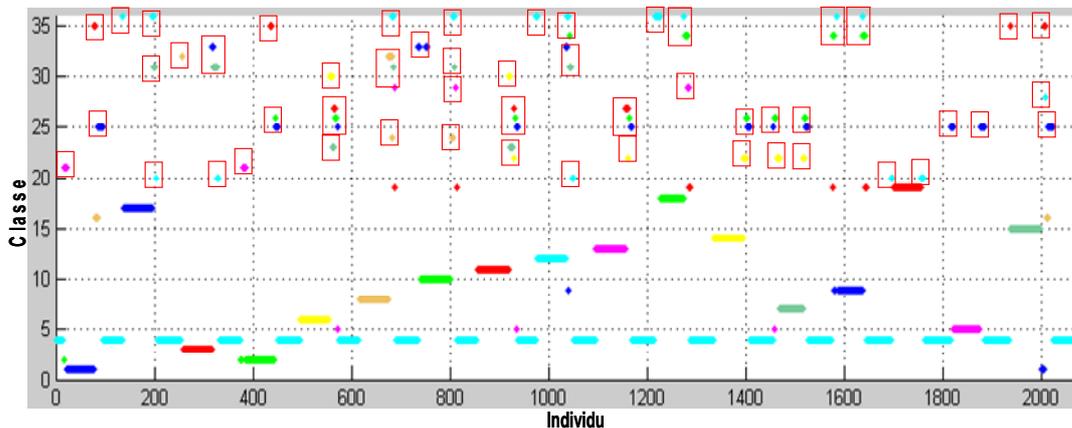


Figure 4.40. Classification de référence -9 descripteurs, Réaction du Thiosulfate de Sodium

La valeur de similitude entre les partitions est mesurée à partir de la matrice de contingence (Annexe 2) pour estimer l'indice de dissemblance. Dans ce cas, la valeur est aussi basse (0.0387) indiquant ainsi que les deux partitions sont semblables.

Le modèle de comportement (association d'états aux classes) et la relation entre les deux partitions sont présentés dans le Tableau 4.13. La Figure 4.41. présente la

partition de référence et la partition optimale avec les classes renommées afin de montrer la similitude.

REFERENCE	AUTOMATIQUE	N. ETAT [ORA05]	REFERENCE	AUTOMATIQUE	N. ETAT [ORA05]
1	2	↓FU_IN	19	35	↑X_INJ
2	3	↑TU_B1_IN	20	1	SI2
3	4	↓TU_B1_IN	21	13	SI3
4	18	NORMAL	22	20	SI4
5	6	↓X_INJ	23	22	SI5
6	7	↓FP_IN_1	24	24	SI6
7	8	↑X_MAIN	25	27	SI7
8	9	↑FP_IN_1	26	29	SI8
9	30	↓X_MAIN	27	22	SI9
10	11	↓TP_IN_1	28	22	SI10
11	12	↑TP_IN_1	29	35	SI11
12	14	↓FP_IN_2	30	22	SI12
13	15	↑FP_IN_2	31	39	SI13
14	16	↑TP_IN_2	32	42	SI14
15	17	Arrêt FU_IN	33	42	SI15
16	41	SI1	34	39	SI16
17	28	↑FU_IN	35	37	SI17
18	25	↓TP_IN_2	36	39	SI18

Tab. 4.13. – Correspondance des classes et des états-9 descripteurs, Réaction du Thiosulfate de Sodium.

Avec la partition optimale, plusieurs données qui avaient été classées comme "NOM STATE" sont identifiées comme faisant partie d'un état défini automatiquement par le regroupement des classes semblables. La classe 5 de la partition automatique peut être associée à un état de pre-défaillance, l'expert a regroupé les éléments de cette classe dans la classe normale. La méthode d'amélioration proposée laisse cette classe séparée en donnant une information très utile pour la prévention des défaillances.

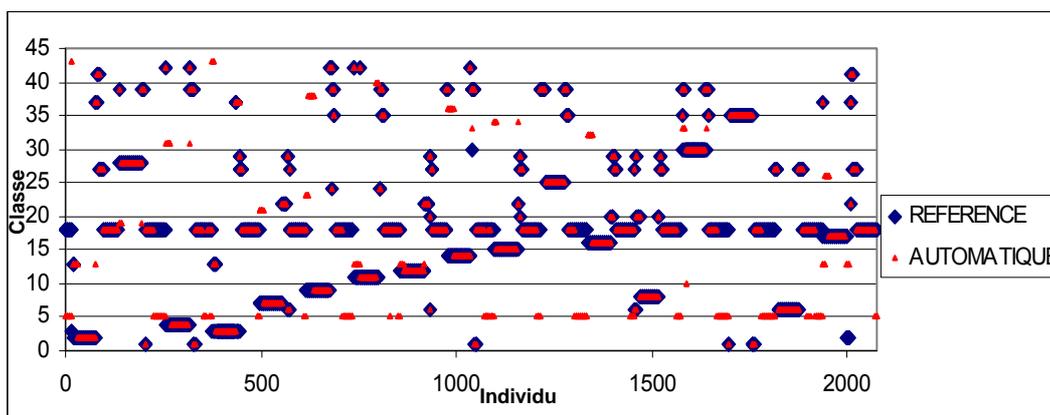


Figure 4.41. Comparaison des classifications optimale et de référence -9 descripteurs, Réaction du Thiosulfate de Sodium.

4.6 Conclusion

Au cours de ce chapitre, nous avons présenté l'application de la méthode d'évaluation et d'optimisation des partitions proposée dont l'objectif est de donner une partition plus facilement utilisable pour établir les états fonctionnels à partir de données d'historiques et un apprentissage non supervisé. Le but est l'obtention d'un nombre optimal de classes, et de fournir, de cette manière, une aide très précieuse pour orienter l'expert du procédé dans l'étape d'établissement d'un modèle comportemental. La partition optimale est établie en fonction du critère mathématique de séparation et de compacité des classes.

En fournissant une partition plus facilement associable aux états du procédé, nous avons contribué à l'amélioration de la technique de placement de capteurs proposée dans [ORA07a] consistant à établir les descripteurs les plus importants d'un point de vue de la quantité d'information dans le procédé.

Nous avons identifié, en utilisant un critère de qualité et de façon automatique, des modèles de comportement pour trois procédés de l'industrie chimique. Les résultats sont très semblables à ceux qui avaient été établis par l'expert du procédé. La méthode d'optimisation proposée permet de réunir les classes pour obtenir une classification plus simple, et donc optimale dans ce sens, et elle permet aussi de garder des classes très petites qui peuvent correspondre à des états de transition pouvant constituer des "états d'alarmes" coïncidant avec une dérive ou un état de "pré-défaillance", états qu'il est crucial de détecter dans le cas de procédés industriels, ceci pouvant permettre alors de déclencher des maintenances préventives.

5. CONCLUSIONS ET PERSPECTIVES

Tout au long de ce travail de thèse, nous avons principalement traité le problème de l'évaluation et de l'amélioration des partitions floues. Nous avons abordé ce problème afin d'améliorer de manière automatique, la partition obtenue par des méthodes de fouille de données. Pour la surveillance de processus, une meilleure caractérisation des états fonctionnels du système permet de mieux diagnostiquer le procédé, c'est-à-dire de mieux détecter des états anormaux, ainsi que des dérives éventuelles dans le fonctionnement du procédé.

Pour le cas de processus complexes (e.g. chimiques et biochimiques) qui présentent des dynamiques hautement non linéaires et possèdent un grand nombre de variables, pour lesquels il est parfois difficile d'obtenir un modèle mathématique précis qui permette de détecter les états anormaux, les méthodes d'apprentissage et de reconnaissance de formes ont montré de très bonnes performances. En particulier, les méthodes issues de la logique floue (méthodes de classification floue), grâce à leur possibilité d'exprimer l'appartenance d'un individu à plusieurs classes, permettent d'obtenir une description plus précise (parce que plus complète) de la situation actuelle du système. Vu le grand nombre de méthodes qui ont été proposées pour la classification floue nous avons procédé, dans le Chapitre 2, à une comparaison de ces méthodes sur des données issues de *benchmarks* reconnus dans le domaine de la classification. Cette comparaison nous a permis de montrer que les performances des algorithmes peuvent être très semblables si les paramètres propres à chaque algorithme ont été correctement choisis. Toutefois leurs performances sont fortement dépendantes de la bonne sélection des paramètres et du nombre de classes lorsque celui-ci doit être connu *a priori*. Devant ces difficultés, nous avons voulu proposer une méthode, générique, applicable à toutes les méthodes de classification floue, qui permette de diminuer la dépendance des algorithmes d'apprentissage aux paramètres propres à chacun d'eux et de déterminer de manière automatique la meilleure partition qui caractérise les états fonctionnels, ainsi que le nombre de classes.

Nous avons proposé une méthode d'évaluation et d'amélioration de la partition obtenue par des techniques de classification floue. Cette méthode permet de trouver automatiquement, et à partir uniquement de la matrice de partition floue qui est associée à une partition initiale, une partition optimale en termes de compacité et de séparation de classes. Cette méthode permet d'évaluer et d'améliorer l'apprentissage des états fonctionnels qui sont établis à partir de la classification des données historiques du système en facilitant l'entraînement des méthodes d'apprentissage basées sur l'analyse de données. Cette méthode fournit donc une orientation et une aide précieuse aux experts du procédé pour établir les états fonctionnels lors de l'implantation d'une technique de surveillance basée sur des méthodes de classification floue.

La méthode proposée est basée sur un algorithme qui améliore de manière itérative la partition. A chaque itération, la qualité de la partition courante est tout d'abord

évaluée et les deux classes les plus semblables sont réunies pour obtenir une nouvelle partition. En prenant en compte que le nombre initial de classes K est obligatoirement fini, le nombre de partitions obtenues sera par conséquent égal ou inférieur à $K-1$. L'indice maximum de qualité sera donc toujours atteint en un nombre fini d'itérations. En conséquence, nous avons choisi d'arrêter l'algorithme jusqu'à obtenir une partition avec deux classes, et la partition optimale correspond à la partition avec l'indice de qualité le plus élevé.

Pour la définition du critère de qualité, nous avons étudié différentes approches. En général, les approches proposées dans la littérature ont été définies pour un type d'algorithme d'apprentissage et leur application nécessite généralement des valeurs spécifiques à un type d'algorithme (e.g. description géométrique des classes). Ces indices ne sont donc pas applicables à tous les algorithmes de classification floue. Dans la littérature, nous trouvons aussi des indices génériques applicables à différentes méthodes de classification mais ils incluent dans leur calcul, les valeurs des données classées, en restreignant leur application aux données de type quantitatif. C'est pour cela que nous avons proposé un indice de qualité des partitions qui est estimé uniquement à partir des degrés d'appartenance qui sont des informations fournies par toute méthode de classification floue.

Pour l'évaluation de la qualité, nous avons suivi la tendance de la plupart des indices de qualité déjà proposés et qui ont montré de bonnes performances pour trouver la meilleure partition entre plusieurs partitions. Comme critère de qualité, nous avons conservé une relation entre la compacité et la séparation des classes (dispersion et distance entre classes). De cette manière, nous cherchons une partition avec des classes compactes et séparées. Ces deux caractéristiques, dans le cas de la surveillance de procédés industriels, nous permettent d'obtenir une classification plus simple qui facilite l'identification des états fonctionnels du procédé tout en permettant de conserver des classes très petites qui peuvent correspondre à des états de transition. Ce dernier point est très important car la tentation lors d'un regroupement « manuel » des classes en état (comme cela était fait auparavant), est la diminution du nombre d'états. Cependant, l'expert devait lors de cette opération s'assurer de ne pas faire disparaître des classes petites mais importantes qui pouvaient constituer des "états d'alarmes" coïncidant avec une dérive ou un état de "pré-défaillance", états qu'il est crucial de détecter dans le cas de procédés industriels pour pouvoir déclencher des maintenances préventives.

Pour l'estimation de la similitude entre ensembles flous, nous avons proposé une mesure qui permet d'analyser la similitude entre deux ensembles flous en même temps en termes de séparation entre classes et d'entropie des classes. L'entropie floue permet de mesurer la quantité d'information d'un ensemble flou; les classes qui apportent la même quantité d'information peuvent être considérées comme similaires. De la même manière que pour l'indice de qualité, cette mesure est calculée uniquement à partir des degrés d'appartenance (matrice de partition floue) et non pas des valeurs des individus eux-mêmes.

Pour l'application des mesures de distance et de similitude, chaque classe est représentée comme un ensemble flou défini dans l'univers de discours discret $[1 N]$ où N correspond au nombre des individus classés. Cette représentation confère à la méthode d'évaluation et d'amélioration des partitions une indépendance vis-à-vis du nombre de variables (descripteurs). Cette caractéristique constitue un avantage indéniable, parce que les performances de cette méthode ne dépendent pas de la quantité de descripteurs. C'est un point clé, spécialement intéressant dans le cas du diagnostic en-ligne des systèmes complexes qui ont un grand nombre de variables à analyser en-ligne pour trouver une description complète de l'état courant du procédé.

La méthode que nous proposons est une aide essentielle dans l'étape d'apprentissage des méthodes de classification non supervisées parce qu'elle permet d'améliorer une partition en fonction de critères clairs. Dans le cas d'applications dans le domaine de la supervision, comme nous l'avons illustré dans le Chapitre 4 à travers des exemples issus du domaine industriel, la méthode est un complément aux outils de supervision parce qu'elle offre à l'opérateur une sorte de « modèle de comportement » lui permettant de suivre le comportement du système. D'après les résultats obtenus, nous pouvons conclure que l'approche proposée permet effectivement de donner à l'expert du système une partition plus simple facilitant ainsi une association plus directe aux états fonctionnels du système. De cette manière, la méthode constitue une amélioration réelle des stratégies pour la construction de modèles comportementaux des processus utiles pour la supervision.

Nous voulons faire remarquer que le principal avantage de notre approche est qu'elle inclut l'évaluation et l'amélioration des partitions floues, sans avoir de limitation sur le type d'algorithme de classification, le type de données et sans avoir besoin de relancer l'algorithme d'apprentissage. C'est donc une méthode tout à fait générique applicable à toute technique de classification. Ces caractéristiques permettent d'avoir une méthode rapide qui diminue la dépendance de ces techniques aux paramètres des algorithmes de classification. La méthode estime, sans ajouter de paramètres supplémentaires à choisir par l'utilisateur et à partir d'une seule partition initiale adéquate bien que non optimale, le nombre adéquat de classes. Cette méthode constitue donc un complément très utile aux méthodes de classification de données qui n'ont pas besoin de connaissance *a priori* de la distribution de données.

Nous pouvons regretter le fait de ne pas disposer d'un critère permettant d'arrêter l'algorithme une fois que le maximum ait été trouvé et qui assure dans un même temps que l'algorithme ne tombe pas dans un maximum local. Nous avons préféré faire une analyse exhaustive jusqu'à obtenir une partition avec deux classes uniquement. Dans la procédure actuelle, la partition dite « optimale » est celle qui fournit l'indice de qualité maximal. Des études permettant d'arrêter l'algorithme dès que cette partition optimale ait été trouvée pourraient être envisagées. Cependant, le fait de disposer des indices de qualité pour toutes les partitions possibles est une information qui peut être elle-même très utile pour l'expert. Il peut en effet décider de passer outre le résultat préconisé par la méthode et imposer le nombre de classes (pour des raisons liées par exemple aux connaissances particulières qu'il a du procédé) et il dispose donc avec

cette information d'un critère lui permettant d'évaluer si son choix est vraiment pertinent en termes de qualité de la partition. De plus, le nombre d'itérations de la méthode est fixé par le nombre de classes initiales et donc le plus souvent ce nombre sera faible au regard de celui d'une méthode d'optimisation non linéaire qu'il faudrait mettre en œuvre pour assurer l'arrêt automatique à l'optimum global.

Ce travail laisse cependant un certain nombre de questions ouvertes qui pourraient être approfondies tant sur le plan mathématique que sur le plan méthodologique. Dans ce qui suit, nous donnons quelques points qui nous paraissent pertinents pour des développements futurs.

Il serait par exemple intéressant d'analyser l'influence du type d'opérateurs employés pour la *S-Norme* et la *T-Norme* dans le calcul de l'indice de qualité et la mesure de similitude entre classes. Pour le moment, l'approche proposée utilise comme *S-Norme* et *T-Norme* le **maximum** et le **minimum**, qui sont les opérateurs les plus généraux.

Comme le temps de calcul nécessité par l'algorithme est faible, il est envisageable de penser à une application de reconfiguration du système de diagnostic en ligne, (sans avoir à arrêter le procédé). En effet, si le système de diagnostic est basé sur une technique de classification qui permet de créer des classes au fur et à mesure que l'on détecte un état non identifié auparavant ou inconnu (e.g. LAMDA, RNA proposé par Lurette [LUR03]), une validation de la nouvelle partition par rapport à la précédente serait très utile pour réaliser correctement l'adaptation (évolution) du modèle comportemental du procédé.

Les travaux effectués ont permis de développer une mesure de la similitude des classes. A partir de cette information il est donc possible d'établir l'état le plus « proche » d'un état courant et de déterminer ainsi l'état suivant le plus « probable » s'il y a changement d'état à venir. Cette possibilité pourrait donc permettre d'effectuer un pronostic sur l'évolution future du procédé qui peut être intégré dans le diagnostic en ligne et être utilisé par l'opérateur pour déclencher des maintenances préventives.

Dans la méthode de placement de capteurs proposée par Orantes A. [ORA07a], ceux-ci sont choisis parmi un ensemble de capteurs possibles en référence à la quantité d'information qu'ils fournissent. On peut imaginer d'utiliser les indices de qualité des partitions pour sélectionner ces capteurs. En effet, ce que l'on recherche c'est l'ensemble minimal des capteurs qui permettent d'identifier correctement tous les états de défaillance mais aussi qui permettent de mieux les discriminer : on retrouve donc ici le principe d'obtenir une partition avec des classes compactes et séparées. Cependant, il reste en suspens la question de la validité mathématique de la comparaison des indices de qualité pour des partitions qui font référence à quantités différentes de descripteurs (capteurs).

Dans l'immédiat, la poursuite de ces travaux s'effectue en coopération avec le groupe de recherche sur l'automatisation des puits de pétrole du secteur PDVSA-

Barinas et concerne l'implantation d'un système de surveillance des puits de pétrole du secteur Barinas-Venezuela. Les caractéristiques de ce problème font qu'il est amplement traitable avec une méthode de surveillance incluant pour l'entraînement, la méthode proposée dans cette thèse. Précédemment, l'utilisation de réseaux de neurones pour le système de diagnostic a été étudiée, mais l'expert du processus n'est pas parvenu aux performances désirées à cause de la difficulté, d'une part du choix correct des paramètres et d'autre part d'interprétation des classes. Le groupe de recherche de PDVSA (*Petrolera de Venezuela S.A.*) a donc un intérêt tout particulier pour inclure les méthodes de validation automatique des partitions dans l'étape d'entraînement pour faciliter l'utilisation de méthodes de diagnostic basée sur l'analyse des données.

BIBLIOGRAPHIE

- [AGU82] Aguilar J., Lopez De Mantaras R., *The process of classification and learning the meaning of linguistic descriptors of concepts*. Approximate reasoning in decision analysis, p. 165-175. North Holland, 1982.
- [AKB01] Akbaryan F., Bishnoi P.R., *Fault diagnosis of multivariate systems using pattern recognition and multisensor data analysis technique*. Computers and chemical engineering, Elsevier, 2001.
- [ALE04] Alexiuk M.D., Pizzi N.J., *Cluster validation indices for fMRI data: fuzzy C-means with feature partitions versus cluster merging strategies*. IEEE NAFIPS conference, 2004.
- [ALS01] Al-Sharhan S., Karray F., Gueaieb W., Basir O., *Fuzzy Entropy: a Brief Survey*. IEEE International Fuzzy Systems Conference, 2001.
- [ATSDR] ATSDR, Agency for Toxic Substances and Disease Registry disponible sur: <http://www.atsdr.cdc.gov/>.
- [BAB98] Babuska R., *Fuzzy modeling for control*. Control engineering technology and systems. Delft University of Technology, Kluwer Academic Publishers, 1998.
- [BEN91] Benuzzi A., Zaldivar J.M., *Safety of chemical batch reactors and storage tanks*, Eds Kluwer Academic Publishers, Dordrecht and Norwell, Massachusetts, 1991.
- [BEZ05] Bezdek J., Keller J., Krisnapuram R., Pal N., *Fuzzy models and algorithms for pattern recognition and image processing*, Springer, 2005.
- [BEZ81] Bezdek J., *Pattern recognition with fuzzy objective function*, Ed. Plenum Press, NY, USA, 1981.
- [BEZ98] Bezdek James C., Pal Nikhil R., *Some New Indexes of Cluster Validity*. IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics, Vol. 28, No. 3, 1998.
- [BIS95] Bishop C., *Neural Networks for Pattern Recognition*, New York: Oxford University Press Inc, 1995.
- [BOI03] Boilleraux L., Flaus J.-M., *Commande et supervision des Procédés Agroalimentaires*, Hermes Paris, 2003.
- [BRE84] Breiman L., Friedman J., Olshen R., Stone C., *Classification and Regression Trees*. Statistics/Probability Series, Wadsworth & Brooks/Cole Advanced Books and Software, 1984.

- [CAS03] Casimir R., *Diagnostic des Défaits des Machines Asynchrones par Reconnaissance des Formes*. Thèse de Doctorat, Ecole Centrale de Lyon, 2003.
- [CHE91] Chen S., Cowan C., Grant P., *Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks*. IEEE on Neural Networks, vol.2, 1991.
- [CHI01] Chiang L., Russell E., Bratz R. *Fault Detection and Diagnosis in Industrial Systems*. Springer, 2001.
- [DEV03] Devatine A., Prat L., Cognet P., Cabassud Michel, Gourdon C., Chopard F., *Process Intensification, performances evaluation of a new concept « Open Plate Reactor »*. ECCE 4, 2003.
- [DUB01] Dubuisson B., *Diagnostic, intelligence artificielle et reconnaissance des formes*, Hermes science ISBN 2-7462-0249-2, 2001.
- [DUD01] Duda R., Hart P., Stork D., *Pattern Classification*, Willey Interscience Second edition, 2001.
- [FIS36] Fisher R.A., *The use of multiple measurements in taxonomic problems*. Annals of Eugenics, v.7, p. 179-188 Cambridge University Press, 1936
- [FIX51] Fix E., Hodges, J.L., *Discriminatory analysis. Non-parametric discrimination*. USAF School of Aviation Medicine, Randolph Field, TX, 4 Technical Report 4, 1951.
- [FRA02] Franco Claudia, SILVA Leonardo, OLIVEIRA Adriano, *A Validity Measure for Hard and Fuzzy clustering derived from Fisher's Linear Discriminant*. International Conference on Fuzzy Systems, 2002.
- [GALI02] Galindo M., Aguilar-Martin J., *Interpretación Secuencial de Encuestas con Aprendizaje LAMDA. Aplicación al Diagnostico en Psicopatología*. Workshop IBERAMIA'02, 2002.
- [GALV96] Galvan M., Zaldivar J., Hernández H., Molga E., *The use of neural networks for fitting complex kinetic data*. Computers and Chemical Engineering, vol. 20, no. 12, pp. 1451-1465, 1996.
- [GAT89] Gath I., Geva A., *Unsupervised Optimal Fuzzy Clustering*. IEEE transaction on Pattern Analysis and Machine Intelligence Vol.11, 1989.
- [GEN07] Gentil S., Celse B., Charbonnier S., Cocquempot V., Hamelin F., Leseq S., Maquin D., Montmain J., Ragot J., Sauter D., *Supervision des Procédés Complexes*, Lavoisier, 2007.
- [GOR92] Gordon H.L. and Somorjai R.L., *Fuzzy cluster analysis of molecular dynamics trajectories*. Proteins Functions and Genetics, Vol. 14, 1992.
- [GOU05] Goushun H., Yunsheng L., *New Subsethood Measures and Similarity Measures of Fuzzy Sets*. IEEE International Conference on Communications, Circuits and Systems, 2005.
- [GRI02] Grisales V., Gauthier A., Isaza C., Villamarin, G.A., *Identificación y modelado difuso de Takagi-Sugeno de un Bioreactor Anaerobio*. X Congreso Latinoamericano de Control, 2002.
- [GUS92] Gustafson D., Kessel W., *Fuzzy Clustering with a Fuzzy Covariance Matrix*. IEEE Fuzzy Models for Pattern Recognition, 1992 reprinted from Proc. IEEE CDC 1979
- [HAL02] Halkidi M., Batistakis Y., Vazirgiannis M., *Clustering validity checking*

- methods: part I et II*,. ACM SIGMOD Record, v.31 n.3, 2002.
- [HAM03] Hamilton H., University of Regina, Canada, Dpt. of Computer Science in:http://www2.cs.uregina.ca/~hamilton/courses/831/notes/ml/dtrees/4_dtrees1.html.
- [HAN01] Han J., Kamber M., *Data Mining: Concepts and Techniques*, Ed. Morgan Kaufmann, 2001.
- [HER06] Hernandez H., *Supervision et diagnostic des procédés de production d'eau potable*. Thèse de Doctorat, INSA de Toulouse, 2006.
- [HUL05] Hüllermeier Eyke, *Fuzzy methods in machine learning and data mining: Status and prospects*, Fuzzy Sets and Systems, pg. 387-406, 2005.
- [HYSYS] HYSYS, Manuels Get Started and Customization Guide versions 3.1, Hyprotech. disponibles sur:
<http://itll.colorado.edu/HYSYS/Doc/Doc/HYSYS/GetStart.pdf> and
<http://itll.colorado.edu/HYSYS/Doc/Doc/HYSYS/CustomGD.pdf>, 2002.
- [ISA04] Isaza C., Kempowsky T., Aguilar-Martin J., Gauthier A., *Qualitative data Classification Using LAMDA and other Soft-Computer Methods*. Recent Advances in Artificial Intelligence Research and Development, IOS Press, 2004.
- [ISA06] Isaza C., Aguilar-Martin J; LeLann M.V., Aguilar J., Rios A., *An Optimization Method for the Data Space Partition Obtained by Classification Techniques for the Monitoring of Dynamic Processes*. Artificial Intelligence Research and Development, IOS Press, 2006.
- [ISA07] Isaza C., Diez-Lledo E., Hernández de Leon H., Aguilar-Martin J.; LeLann M.V., *Decision Method for Functional States Validation in a Drinking Water Plant*. 10th Computer Application in Biotechnology (CAB), IFAC, 2007.
- [JAI88] Jain A.K., *Handbook of Pattern Recognition and Image Processing*, New York: Academic, 1988.
- [JAN97] Jang J., Sun C., Mizutani E., *Neuro-Fuzzy and Soft Computing*, Prentice Hall, 1997.
- [KAU75] Kaufmann, *Introduction to the theory fuzzy subsets*, New York: Academic, Vol.1, 1975.
- [KAY95] Kaymak U., Babuska R, *Compatible Cluster Merging for Fuzzy Modelling*, IEEE Fuzzy Systems 1995.
- [KAY02] Kaymak U., Setnes M., *Fuzzy Clustering With Volume Prototypes and Adaptive Cluster Merging*. IEEE Transaction on Fuzzy Systems, Vol. 10 No. 6, 2002.
- [KEM03] Kempowsky T., Aguilar-Martin J., Subias A., Le Lann M.V., *Classification tool based on interactivity between expertise and self-learning techniques*, 5th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes , 2003.
- [KEM04a] Kempowsky T., *Surveillance de Procédés à Base de Méthodes de Classification*. Thèse de doctorat, INSA Toulouse, 2004.
- [KEM04b] Kempowsky T., *SALSA user's manual*. Rapport LAAS-CNRS No. 04160, 2004.
- [KEM06] Kempowsky T. and Le Lann M.V., *Nouvelles applications dans le domaine*
-

de la sécurité industrielle et du REX de systèmes d'aide à la conduite supervisée, d'abstraction d'informations, d'analyse et de classification de données. Rapport LAAS No. 06391, 2006.

- [KIO94] Koivo H.N., *Artificial Neural Networks in fault diagnosis and control, Control.* Engineering Practice, Vol. 2, Nr.1, 1994.
- [KOS86] Kosko, *Fuzzy entropy and conditioning*, Informatic Science, Vol.40, 1986.
- [KOU02] Kourtí T., *Process Analysis and Abnormal Situation Detection : From Theory to Practice*, IEEE Control Systems Magazine, Vol. 22, No. 5, 2002.
- [KRI00] Krishnapuram R. and Kim J., *Clustering algorithms based on volume criteria*, IEEE Trans. Fuzzy Syst., vol. 8, pp. 228–236, Apr. 2000.
- [KRI92] Krishnapuram R., Freg C., *Fitting an unknown number of lines and planes to image data through compatible cluster merging*, Pattern Recognition, 1992.
- [LEE01] Lee H., Chen C., Chen J., Jou Y., *An Efficient Fuzzy Classifier with Feature Selection Based on Fuzzy Entropy*, IEEE Transactions on Systems, Man and Cybernetics, Vol. 31, No. 3, June 2001.
- [LIM00] Lim S., Loh W., Shih Y., *A comparison of Prediction of Accuracy Complexity and Training Time of Thirty-three Old and New Classification Algorithms*. Machine Learning, Kluwer Academic Publishers, 2000.
- [LOP77] López de Mántaras R., *Auto apprentissage d'une partition : Application au classement itératif de données multidimensionnels*. Thèse de Doctorat, UPS de Toulouse, 1977.
- [LOP89] López de Mántaras, R., *ID3 revisited: a distance-based criterion for attribute selection.*, Methodologies for Intelligent Systems, 4. Ed. Ras, Z.W. North-Holland,1989.
- [LUC72] Luca D. and Termini S., *A Definition of Non Probabilistic Entropy in the Setting of Fuzzy Set Theory*. Information and Control, 20:301-312,1972.
- [LUR03] Lurette C., *Développement d'une technique neuronale auto adaptative pour la classification dynamique de données évolutives Application a la supervision d'une presse hydraulique*, Thèse de Doctorat de l'Université des Sciences Technologies de Lille, Lille France,2003.
- [MARI03] Marie-Joseph I., *Méthodologie de diagnostic appliquée á la maintenance Préventive d'unités de production d'électricité en sites isolés*. Thèse de Doctorat de l'Université des Antilles et de la Guyane, 2003.
- [MARK96] Mark J., Orr L., *Introduction to Radial Basis Function Networks*. Centre for Cognitive Science, University of Edinburgh, 1996.
- [MART02] Martin del Brio B., Sanz A., *Redes Neuronales y Sistemas Difusos*, RAMA, 2002.
- [MAU02] Maulik U. and Bandyopadhyay S., *Performance Evaluation of some clustering algorithms and Validity Indices*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 24 No 12, 2002.
- [MER96] Merz C.J., Murphy, P.M., *UCI Repository of Machine Learning Databases*. Dpt. Information and computer science U.California <http://www.ics.uci.edu/~mlearn/MLRepository.html>, 1996.
- [MIC94] Michie D., Spiegelhater D.J., Taylor C.C., *Machine Learning, Neural and Statistical Classification*, Ellis Horwood series in Artificial Intelligence,

- 1994.
- [ORA05] Orantes A., *Méthodologie pour le placement des capteurs a base de méthodes de classification en vue du diagnostic*. Thèse de Doctorat, INSA de Toulouse, 2005.
- [ORA07a] Orantes A., Kempowsky T., Le Lann M. V., Aguilar-Martin J., *A new support methodology for the placement of sensors used for fault detection and diagnosis*, Chemical Engineering Process, Elsevier, In Press 2007.
- [ORA07b] Orantes A., Kempowsky T., M.V. Le Lann, Prat L., Elgue S., Gourdon C., Cabassud M., *Selection of sensors by a new methodology coupling a classification technique and entropy criteria*, Trans IChemE, Part A, Chemical Engineering Research and Design, Vol 85 (A6), 2007.
- [PAL92] Pal N. R. et Pal S. K., *Higher order fuzzy entropy and hybrid entropy of a set*, Inform. Sei., vol. 61, no. 3, pp. 211-231, 1992
- [PIE91] Piera N., Aguilar-Martin J., *Controlling Selectivity in Non-Standard Pattern Recognition Algorithms*. IEEE Transactions on systems, man and cybernetics, Vol. 21, 1991.
- [POW87] Powell M.J.D., *Radial Basis Functions for Multivariable Interpolation: A Review*. Algorithms for Approximation, J.C. Mason and M.G. Cox Eds., 1987.
- [QUI86] Quinlan J.R., *Induction of Decisions Trees*. Machine Learning 1,81-106, 1986.
- [RIV96] Rivière J.M., Bayart M/, Thiriet J.M., Bouras A., Robert M., *Intelligent Instruments: Some modelling approaches*. Measurement and Control, vol. 29, 1996.
- [SHA48] Shannon C. E., *A mathematical theory of communication*. Bell Syst. Tech. J., vol. 27, pp. 379-423, 1948.
- [TAN05] Tang Y., Sun F., Sun Z., *Improved Validation Index for Fuzzy Clustering*. American Control Conference, 2005.
- [TRI78] Trillas E., Riera T., *Entropies in finite fuzzy sets*. Information Sciences 15, 2, pp. 159-168, 1978.
- [TRI79] Trillas E., Alsina C., *Sur les mesures du degré du flou*. Stochastica, vol.III pp. 81-84, 1979.
- [VEN03] Venkatasubramanian V., Rengaswamy R., Yin K., Kavuri S. N., *A review of process fault detection and diagnosis*. Computers and Chemical Engineering. Elsevier Science, 2003.
- [WAI00] Weissman J., *Construction d'un modèle comportemental pour la supervision de procédés: Application a une station de traitement des eaux* Thèse de Doctorat de l` Institut National Polytechnique de Toulouse, 2000.
- [WAN05] Wang X. and Syrmos V., *Optimal Cluster Selection Based on Fisher Class Separability Measure*. American Control Conference, 2005.
- [XIE91] Xie L., Xuanli B., *A Validity Measure for Fuzzy Clustering*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 13 No 8, 1991.
- [YAG79] Yager R.R., *On measures of fuzziness and fuzzy complements*, Int. J.
-

Gen., vol. 5 Syst, 1979.

- [ZAD68] Zadeh L. A., *Probability measures of fuzzy events*, J. Math. Analysis and Applications, vol. 23, pp. 421-427, 1968.

ANNEXE 1. Matrice de contingence-8 descripteurs, Réaction d'estérification

Partition de Référence

EIATS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	0,051	0	0	0	0	0	0,034	0	0	0	0	0	0	0	0	0,04	0	0	0,038	0	0	0	0	0	0	0	0	0	0	0
2	0	0,025	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0,024	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0,026	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0,026	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0,029	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0,027	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0,379	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0,029	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0,029	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0,029	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0,03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0,028	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0,029	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,029	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0,002	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	8E-04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0,002	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0,002	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,005	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	8E-04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,006	0,006	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,002	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	8E-04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0,005	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,005	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,003	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8E-04	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0,004	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,002	0
29	0	0,004	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,003	0	0	0	0
31	0	0	0	0,003	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,002	0	0	0
33	0	0	0	0	0	8E-04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34	0,019	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,003	0	0
35	0	0	0	0	0	0,002	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
36	0	0,002	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,003	0	0	0	0	0	0	0	0	0	0	0	0
38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,008

Partition Automatique





DIAGNOSTIC PAR TECHNIQUES D'APPRENTISSAGE FLOUES: CONCEPTION D'UNE METHODE DE VALIDATION ET D'OPTIMISATION DES PARTITIONS.

RESUME

Ce travail se situe dans le domaine du diagnostic des processus défini comme étant l'identification de ses états fonctionnels. Dans le cas où l'obtention d'un modèle précis du processus est délicate ou impossible, la connaissance sur le système peut être extraite à partir des signaux obtenus lors d'un fonctionnement normal ou anormal en incluant des mécanismes d'apprentissage. Cette connaissance s'organise sous l'aspect d'une partition de l'espace des données sous forme de classes (représentant les états du système). Parmi les techniques d'apprentissage, celles incluant de la logique floue ont l'avantage d'exprimer les appartenances d'un individu à plusieurs classes, ceci permet de mieux connaître la situation réelle du système et prévoir des changements vers des états de défaillance. Nonobstant leurs performances adéquates, leur forte dépendance aux paramètres d'initialisation est une difficulté pour l'apprentissage. Cette thèse se situe dans l'amélioration de ces techniques, en particulier notre objectif est l'élaboration d'une méthode permettant de valider et d'adapter automatiquement la partition de l'espace de données obtenue par une technique de classification floue. Elle permet de trouver automatiquement une partition optimale en termes de compacité et de séparation des classes, à partir uniquement de la matrice des degrés d'appartenance obtenue par une classification antérieure. Cette méthode est donc une aide importante donnée à l'expert du processus pour établir les états fonctionnels dans l'implémentation d'une technique de surveillance d'un procédé complexe. Son application est illustrée sur des exemples académiques et sur le diagnostic de 3 procédés chimiques.

MOTS CLES : Supervision, diagnostic, classification floue, qualité des partitions, optimisation des partitions floues.

DIAGNOSIS BY TRAINING FUZZY TECHNIQUES: DESIGN OF A PARTION VALIDATION AND OPTIMIZATION METHOD

ABSTRACT

This work is in the field of the process diagnosis defined as the identification of process functional states. If obtaining a precise model of the process is delicate or impossible, the system knowledge can be extracted from the signals obtained during a normal or abnormal operation by including mechanisms of training. This knowledge is organized through a data space partition into clusters (representing the states of the system). Among the training techniques, those including fuzzy logic have the advantage of expressing the memberships of an individual to several classes, this makes possible to better know the real situation of the system and to envisage changes to failure states. Notwithstanding their adequate performances, their strong dependence on the initialization parameters is a difficulty for the training. This thesis proposes the improvement of these techniques, specifically our objective is the development of a method to validate and adapt automatically the partition of data space obtained by a fuzzy classification technique. This makes possible to find automatically an optimal partition in terms of clusters compactness and separation from only the membership matrix obtained by an initial classification. This method is thus an important help given to the process expert to establish the functional states in the implementation of a monitoring technique of a complex process. Its application is illustrated on academic examples and on the diagnosis of 3 chemical processes.

KEYWORDS: Supervision, diagnosis, fuzzy classification, partition quality, fuzzy partition optimization.

