



HAL
open science

Reconnaissance des formes évolutives par combinaison, coopération et sélection de classifieurs

Veyis Gunes

► **To cite this version:**

Veyis Gunes. Reconnaissance des formes évolutives par combinaison, coopération et sélection de classifieurs. Apprentissage [cs.LG]. Université de La Rochelle, 2001. Français. NNT : . tel-00631621

HAL Id: tel-00631621

<https://theses.hal.science/tel-00631621>

Submitted on 12 Oct 2011

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.

Reconnaissance des formes évolutives par combinaison, coopération et sélection de classifieurs

THÈSE

présentée le 15 Octobre 2001

pour l'obtention du

Doctorat de l'Université de La Rochelle
(spécialité informatique)

par

Veyis GUNES

Composition du jury

| | | |
|-----------------------------|----------------|---|
| <i>Rapporteurs :</i> | J.G. POSTAIRE | Professeur des Universités U.S.T.L. Lille |
| | M. ROMBAUT | Professeur des Universités I.U.T. de Troyes |
| <i>Examineur :</i> | J. MITERAN | Maître de Conférences Université de Bourgogne |
| <i>Directeur de thèse :</i> | P. ESTRAILLIER | Professeur des Universités Université de La Rochelle |
| <i>Co-directeurs :</i> | P. LOONIS | Maître de Conférences Université de La Rochelle |
| | M. MENARD | Maître de Conférences Université de La Rochelle |

Reconnaissance des formes évolutives par combinaison, coopération et sélection de classifieurs

THESE

présentée le 15 Octobre 2001 pour l'obtention du

Doctorat de l'Université de La Rochelle
(spécialité informatique)

par

Veyis GUNES

Composition du jury

| | | |
|----------------------|----------------|---|
| Rapporteurs : | J.G. POSTAIRE | Professeur des Universités U.S.T.L. Lille |
| | M. ROMBAUT | Professeur des Universités I.U.T. de Troyes |
| Examineur : | J. MITERAN | Maître de Conférences Université de Bourgogne |
| Directeur de thèse : | P. ESTRAILLIER | Professeur des Universités Université de La Rochelle |
| Co-directeurs : | P. LOONIS | Maître de Conférences Université de La Rochelle |
| | M. MENARD | Maître de Conférences Université de La Rochelle |

Avant-propos

Je remercie tout d'abord M. Jack-Gérard Postaire, Professeur à l'Université des Sciences et Technologies de Lille, et Mlle Michèle Rombaut, Professeur à l'I.U.T. de Troyes, pour avoir soigneusement rapporté ce mémoire et pour leur participation au jury. L'intérêt qu'ils ont porté à ce travail, leurs nombreuses remarques et suggestions m'ont beaucoup touché.

J'ai été très sensible à l'honneur que m'a fait M. Johel Mitéran, Maître de Conférences à l'Université de Bourgogne, en participant au jury en tant qu'examinateur.

J'adresse mes plus vifs remerciements à M. Pierre Loonis et à M. Michel Ménard, Maîtres de Conférences à l'Université de la Rochelle, pour leur investissement dans mes travaux de recherches ainsi que pour leurs précieux conseils.

Je remercie sincèrement M. Pascal Estrailier, professeur à l'Université de La Rochelle et directeur du Laboratoire L3i, pour la direction de ma thèse, bien sûr, mais aussi pour ses qualités humaines et relationnelles.

Enfin, que l'ensemble des membres du laboratoire L3i soient remerciés. Je pense en particulier à Bertrand Vachon, Rémy Mullot, Pierre Courtellemont, Jean-Christophe Burie, Christophe Demko, Thierry Bouwmans, El-Hadi Zahzah, Carl Frelicot et tous les autres membres, anciens ou actuels, permanents ou doctorants.

*A la mémoire de mon père,
à l'Humanité,
à la Diversité.*

Table des matières

| | |
|------------------------------|-----------|
| Notations principales | 11 |
| Table des figures | 13 |
| Liste des tableaux | 17 |
| Introduction | 19 |

Chapitre 1

Combinaison, coopération et sélection de classifieurs

| | | |
|-------|---|----|
| 1.1 | Introduction | 25 |
| 1.2 | La RdF et le classifieur | 26 |
| 1.2.1 | Les étapes de la RdF | 26 |
| 1.2.2 | Le classifieur | 27 |
| 1.2.3 | Taxonomies de classifieurs | 28 |
| 1.3 | Les systèmes de classifieurs | 30 |
| 1.3.1 | Objectifs et intérêts | 30 |
| 1.3.2 | Définition | 32 |
| 1.3.3 | Architecture | 34 |
| 1.3.4 | Taxonomie des systèmes de classifieurs | 36 |
| 1.4 | Performances des classifieurs et des systèmes de classifieurs | 37 |
| 1.4.1 | Les différentes approches | 37 |
| 1.4.2 | La matrice de confusion | 38 |
| 1.5 | Combinaison de classifieurs | 38 |
| 1.5.1 | Combinaison usuelles | 41 |
| 1.5.2 | Combinaison par vote | 42 |
| 1.5.3 | Combinaison par la Théorie Bayésienne de la décision | 42 |

| | | |
|-------|--|----|
| 1.5.4 | Combinaison par la Théorie de l'Evidence | 42 |
| 1.5.5 | Autres cadres théoriques de la combinaison | 45 |
| 1.6 | Coopération de classifieurs | 45 |
| 1.7 | Sélection de classifieurs | 46 |
| 1.8 | Système de classifieurs hybride | 48 |
| 1.9 | Conclusion | 51 |

Chapitre 2

Aspects statiques et dynamiques de la RdF

| | | |
|-------|--|----|
| 2.1 | Introduction | 55 |
| 2.2 | Objets dynamiques | 57 |
| 2.3 | Quelques modèles de classes dynamiques | 59 |
| 2.4 | Quelques expérimentations | 60 |
| 2.4.1 | Classement Bayésien | 61 |
| 2.4.2 | Génération de distributions Gaussiennes | 62 |
| 2.4.3 | Protocole expérimental | 64 |
| 2.4.4 | Dégradations dues aux déplacements des centres de classes | 65 |
| 2.4.5 | Dégradations dues aux variations des densités de classes | 65 |
| 2.5 | Reconnaissance de classes dynamiques | 67 |
| 2.5.1 | Approche 1 : Rendre statique un problème dynamique | 69 |
| 2.5.2 | Approche 2 : Rendre dynamique le système de reconnaissance | 69 |
| 2.6 | Choix d'une approche pour les évolutions périodiques | 70 |
| 2.6.1 | Critères de choix | 70 |
| 2.6.2 | Algorithme | 72 |
| 2.7 | Conclusion | 73 |

Chapitre 3

Coopération de classifieurs

| | | |
|-------|---|----|
| 3.1 | Introduction | 77 |
| 3.2 | Une méthode de coopération de classifieurs | 78 |
| 3.3 | Détermination des ensembles d'apprentissage | 79 |
| 3.3.1 | Coalescence floue | 79 |
| 3.3.2 | Coalescence avec rejet d'ambiguïté | 81 |
| 3.3.3 | Le concept d'ambiguïté : frontières dans les problèmes de coalescence | 82 |
| 3.3.4 | Distance à une CCA | 85 |

| | | |
|-------|--|-----|
| 3.3.5 | Distance à la classe de rejet de distance | 85 |
| 3.3.6 | L'algorithme $c+2$ moyennes floues | 86 |
| 3.4 | Une fusion multi-classifieurs parallèle et adaptée | 86 |
| 3.4.1 | La procédure de coalescence : Module de sélection | 87 |
| 3.4.2 | Les classifieurs | 88 |
| 3.4.3 | Module de fusion | 88 |
| 3.4.4 | Algorithmes | 89 |
| 3.5 | Application | 91 |
| 3.5.1 | Données synthétiques | 91 |
| 3.5.2 | Résultats | 91 |
| 3.6 | Influences des paramètres | 94 |
| 3.6.1 | Influence du taux de rejet de distance | 94 |
| 3.6.2 | Influence du taux de rejet d'ambiguïté | 98 |
| 3.6.3 | Ambiguïté élevée | 98 |
| 3.6.4 | Frontières de décisions | 98 |
| 3.6.5 | Autres types de combinaisons | 98 |
| 3.7 | Conclusion | 103 |

Chapitre 4

Une planification temporelle de systèmes de classifieurs

| | | |
|-------|---|-----|
| 4.1 | Introduction | 107 |
| 4.2 | Problématique | 107 |
| 4.3 | Représentation de connaissances dynamiques par des réseaux de Petri flous | 109 |
| 4.3.1 | RPF pour le raisonnement | 110 |
| 4.3.2 | RPF pour connaissances dynamiques | 111 |
| 4.3.3 | Exemple de RPF | 112 |
| 4.4 | Synthèse d'un Réseau de Petri Flou | 113 |
| 4.4.1 | Aspect dynamique de l'apprentissage | 113 |
| 4.4.2 | Algorithme de construction du système de RdF | 115 |
| 4.5 | Aspect statique de l'apprentissage | 117 |
| 4.6 | Fonctionnement du système de RdF | 117 |
| 4.7 | Validation | 117 |
| 4.7.1 | Données synthétiques | 117 |
| 4.7.2 | Apprentissage | 119 |
| 4.7.3 | Classement | 124 |

| | | |
|---|---|------------|
| 4.7.4 | Variation du nombre de positions dans une période | 124 |
| 4.7.5 | Variation du seuil de confusion tolérée | 124 |
| 4.7.6 | Variation de l'écart-type des distributions | 128 |
| 4.7.7 | Autre type de trajectoire | 128 |
| 4.7.8 | Autres positions initiales | 132 |
| 4.8 | Conclusion | 137 |
| Conclusions et perspectives | | 139 |
| Bibliographie | | 143 |
| <hr/> | | |
| Annexe A Résultats de l'algorithme 3C avec les données ELENA | | 151 |
| A.1 | Les données ELENA utilisées | 151 |
| A.2 | Influence de l'ambiguïté | 152 |
| A.3 | Résultats pour α optimal | 153 |
| A.4 | Compromis entre précision et généralisation | 153 |
| A.5 | Résultats sur un ensemble de test | 155 |
| Annexe B Robustesse de l'algorithme RCD-RPF | | 159 |
| B.1 | Robustesse aux chevauchements des classes | 159 |
| B.2 | Robustesse aux évolutions des trajectoires | 159 |
| Index | | 163 |

Notations principales

| | |
|------------|--|
| x_k | Vecteur de Forme k |
| v_i | Centre de classe i |
| X | Ensemble de vecteurs de Formes |
| X^{lr} | Ensemble d'apprentissage |
| X^t | Ensemble de test |
| Classe | Ensemble d'individus ayant, selon un expert ou selon un classifieur, une même interprétation |
| l | Nombre de classes réelles (d'après l'expert) ou d'interprétation (d'après un algorithme de classement) |
| ϕ_i | Classe i , réelle (d'après l'expert) ou d'interprétation (d'après un algorithme de classement) |
| Φ | Ensemble des classes (réelles) d'interprétation |
| c | Nombre de classes obtenues par un algorithme de classification (singletons) |
| ω_i | Classe de coalescence i |
| Ω | Ensemble des classes d'interprétation = Cadre de discernement = Univers du discours |
| CC | Classe de coalescence |
| CCA | Classe de coalescence de rejet d'ambiguïté (en Anglais : cluster) |
| RPF | Réseau de Petri Flou |
| P_i | Place i d'un RPF |
| T_i | Transition i d'un RPF |
| e_i | Classifieur i |
| SC | Système de classifieurs |

Table des figures

| | | |
|-----|---|----|
| 1.1 | Les différentes étapes de la construction d'un système de classement. | 27 |
| 1.2 | Exemple de <i>Système de Classifieurs</i> (selon \mathcal{R}_1) : les classifieurs e_1 et e_3 fournissent des informations au classifieur e_2 (car $r_{12} = 1$ et $r_{32} = 1$), les décisions des classifieurs e_2 ($r_{22} = 1$) et e_4 ($r_{44} = 1$) sont combinées pour obtenir un vecteur de décision final. | 34 |
| 1.3 | Système de classifieurs effectuant les différents opérations de combinaison, de coopération et de sélection de classifieurs. | 35 |
| 1.4 | Les différents types d'associations de classifieurs. Si un système opère avec plusieurs types d'associations, le système est dit hybride. | 36 |
| 1.5 | Système de classifieurs effectuant la combinaison de classifieurs pour un problème à l classes et d caractéristiques disponibles. | 39 |
| 1.6 | Système de classifieurs effectuant la coopération de classifieurs. | 46 |
| 1.7 | Système de classifieurs effectuant la sélection de classifieurs. | 48 |
| 1.8 | Exemples de travaux concernant les systèmes de classifieurs. A notre connaissance, certaines solutions n'ont pas encore été traitées (cases vides). | 50 |
| 2.1 | Les différentes situations dans le cadre de la RdF. | 57 |
| 2.2 | Passage d'une problématique de RdF à une autre. | 58 |
| 2.3 | Distributions à l'origine ($Y = 30$), trajectoires et distributions finales des 3 classes. La durée entre deux pas successifs est $\Delta_t = 12$ | 66 |
| 2.4 | Evolution des taux reconnaissance pour les 3 classes d'individus. Les taux concernant les classes 2 et 3 finissent par s'annuler car elles sont confondues avec une autre classe. | 66 |
| 2.5 | Distributions à l'origine (zones plus sombres) et distributions finales des 3 classes. | 68 |
| 2.6 | Evolution des taux reconnaissance pour les 3 classes d'individus. | 68 |
| 2.7 | Les différentes étapes d'un système de classement, dans le cas où les classes évoluent périodiquement. | 69 |
| 2.8 | Exemple de problème à deux classes : pas de croisement de trajectoires, la seconde classe est bimodale. | 70 |
| 2.9 | Exemple de problème à deux classes : il y a croisement entre les trajectoires des deux classes. | 71 |
| 3.1 | Partition des ensembles d'individus d'apprentissage, $X^{lr} \subset X$, lorsque le rejet d'ambiguïté est exercé. $\{\omega_1\}, \{\omega_2\}, \{\omega_3\}$: CCs ; $\omega_{12} = \{\omega_1, \omega_2\}, \omega_{23} = \{\omega_2, \omega_3\}$: CCAs. | 79 |

| | | |
|------|---|-----|
| 3.2 | Un classifieur est associé à chaque CC. L'ensemble d'apprentissage, S_{ω_i} du classifieur e_i , est défini sur ω_i et ses CCAs associées (par exemple, la CCA ω_{12} est associée à ω_1 et à ω_2 | 80 |
| 3.3 | Principe de la coopération. | 80 |
| 3.4 | La figure (a) illustre la partition de l'espace de représentation en trois CCs. La figure (b) illustre la partition de l'espace de représentation lorsque le concept d'ambiguïté est appliqué. Nous associons une CCA à chacun des sous-ensembles $A \in 2^\Omega \setminus \{\emptyset, (\{\omega_i\})_{i \in [1,3]}\}$ | 82 |
| 3.5 | La fonction d'appartenance maximale dans le plan factoriel principal des données IRIS, parmi les trois CCs, avec $m = 1.2$ et $\alpha = 0.8$. Distance de Mahalanobis. . . | 87 |
| 3.6 | La fonction d'appartenance maximale dans le plan factoriel principal des données IRIS, parmi les quatre CCAs, avec $m = 1.2$ et $\alpha = 0.8$. Norme de Mahalanobis. . | 88 |
| 3.7 | La prise de décision est le résultat de la fusion des sorties des classifieurs les mieux adaptés. | 89 |
| 3.8 | Ensemble d'apprentissage : $X^{lr} \subset X$ | 92 |
| 3.9 | Ensemble de test : $X^t \subset X$ | 93 |
| 3.10 | Résultats obtenus avec notre méthode de coopération (3C) pour $c = 2$. Les individus entourés (6 parmi 210) sont ceux qui sont mal classés. | 93 |
| 3.11 | Masses affectées à la classe 2, avec rejet de distance ($T = 5$), d'où la limitation dans un cercle dont le rayon dépend du taux de rejet de distance choisi. | 95 |
| 3.12 | Masses affectées à la classe 2, sans rejet de distance ($T \rightarrow \infty$). On voit apparaître les frontières entre les deux classes (iso-masses). | 96 |
| 3.13 | Masses affectées à la classe 1, avec ou sans rejet de distance. | 96 |
| 3.14 | Influence de T sur les frontières de décisions de la classe 2 pour $T = 3.5$ (et $\alpha = 0.1$). 97 | |
| 3.15 | Influence de T sur les frontières de décisions de la classe 2 pour $T = 7$ (et $\alpha = 0.1$). 97 | |
| 3.16 | Iso-masses de la classe 2 pour $\alpha = 0.048$ ($T = 5$). | 99 |
| 3.17 | Iso-masses de la classe 1 pour $\alpha = 0.048$ | 99 |
| 3.18 | Iso-masses de la classe 2 pour $\alpha = 0.03$ ($T = 5$). | 100 |
| 3.19 | Iso-masses de la classe 1 pour $\alpha = 0.03$ | 100 |
| 3.20 | Les régions obtenues pour $\alpha = 0.048$ et $T = 5$ | 101 |
| 3.21 | Les régions obtenues pour $\alpha = 0.02$ et $T = 5$ | 101 |
| 4.1 | Représentation d'une règle de production floue par un RPF. | 112 |
| 4.2 | Représentation de connaissance dynamique par un RPF. | 113 |
| 4.3 | Construction des fonctions d'appartenance des transitions $\mu_{T_n}(t)$ et de celles des jetons $\mu_{P_n}(t)$ | 116 |
| 4.4 | Fonctionnement du système de RdF proposé et principe de classement d'un individu inconnu x : chaque place (P_i) est associée à un état du système de RdF et chaque état du système est représenté par un classifieur ou un Système de Classifieurs (SC). L'individu x est classé par le SC dont l'état du système est adapté à l'instant considéré. Plusieurs SCs (les mieux adaptés) peuvent également contribuer à classer l'individu à travers une phase de fusion et de décision. | 118 |
| 4.5 | Exemple de problème à deux classes, dont les trajectoires sont elliptiques. La position 1 est superposée avec la position 10. Pour les autres positions, seules les centres de classes sont représentées. | 119 |
| 4.6 | Configuration des deux classes dans la position 196. | 120 |

| | | |
|------|---|-----|
| 4.7 | Evolution du taux de reconnaissance en tenant compte des changements d'états, lors de l'apprentissage. Les 4 changements d'états sont définis suites aux 4 chutes visibles sur la courbe. | 121 |
| 4.8 | Résultat de l'apprentissage des différentes positions des centres de classes (pour les différents états indiqués sur la figure). | 122 |
| 4.9 | Places (états) successives et transitions (événements) obtenues à la fin du cycle 3. | 122 |
| 4.10 | Fonctions d'appartenance des transitions $\mu_{T_n}(t)$ (pour les transitions T_1 à T_4). On peut constater que les fonctions générées sont non symétriques et de supports différents (donc adaptées à l'application envisagée). | 123 |
| 4.11 | Fonctions d'appartenance des jetons $\mu_{P_n}(t)$ d'initialisation des places P_n . On peut constater que les fonctions générées sont non symétriques et de supports différents (donc adaptées à l'application envisagée). | 123 |
| 4.12 | Fonctions d'appartenance des jetons $\mu_{P_n}(t)$ obtenues pour les différentes places P_n , résultants de l'opération t-norme entre la fonction de sa transition d'entrée et celle de la place d'entrée de cette transition. | 125 |
| 4.13 | Evolution du taux de reconnaissance en classement (avec $X^t = X^{lr}$ et une même évolution dans le temps). Les chutes sont minimales car le système de RdF anticipe l'évolution des classes. | 125 |
| 4.14 | Evolution du taux de reconnaissance en classement avec fusion (avec $X^t = X^{lr}$). Les résultats sont améliorés par la fusion des sorties des classifieurs pour les deux états les plus valides. | 126 |
| 4.15 | Evolution du taux de reconnaissance en classement avec bruit ($\sigma_{centres} = 1$, $X^t \neq X^{lr}$) et sans fusion. Peu de dégradation due au bruit : le système de RdF est robuste. | 126 |
| 4.16 | Evolution du taux de reconnaissance en classement avec bruit ($\sigma_{centres} = 1$, $X^t \neq X^{lr}$) et avec fusion. Le résultat est également amélioré. | 127 |
| 4.17 | Influence du nombre de positions pour $\sigma = 8$ (avec $cf_i = 0.1$). Plus le chevauchement peut être important et plus le nombre d'états générés sera élevé. Dans ce cas, le nombre de positions influe peu. | 128 |
| 4.18 | Les deux classes dans les positions 0 et 10 avec leurs trajectoires pour $\sigma = 8$ (et $n_p = 256$ positions). | 129 |
| 4.19 | Influence du seuil de confusion tolérée (pour $\sigma = 8$ et $n_p = 256$ positions). | 129 |
| 4.20 | Influence de l'écart-type des distributions des deux classes (pour $cf_i = 0.1$ et $n_p = 256$ positions). | 130 |
| 4.21 | Les deux classes dans les positions 0 et 16 avec leurs trajectoires (pour $\sigma = 2$ et $n_p = 256$ positions). | 130 |
| 4.22 | Positions des centres de classes pour les différents états. | 131 |
| 4.23 | Initialisation 1 : les deux classes dans les positions 0 et 16 avec leurs trajectoires (pour $\sigma = 5$ et $n_p = 256$ positions). | 132 |
| 4.24 | Initialisation 1 : positions des centres de classes pour les différents états (pour $\sigma = 5$ et $n_p = 256$ positions). | 133 |
| 4.25 | Initialisation 1 : marquages obtenus au cours du temps (pour $\sigma = 5$ et $n_p = 256$ positions). | 133 |
| 4.26 | Initialisation 1 : taux de reconnaissance en apprentissage (pour $\sigma = 5$ et $n_p = 256$ positions). Le taux de confusion toléré sur une classe est de $cf_i = 10\%$ | 134 |
| 4.27 | Initialisation 1 : taux de reconnaissance en classement (pour $\sigma = 5$ et $n_p = 256$ positions). Le taux de reconnaissance global reste supérieur ou égal à 96% | 134 |

| | | |
|------|---|-----|
| 4.28 | Initialisation 2 : les deux classes dans les positions 0 et 16 avec leurs trajectoires (pour $\sigma = 5$ et $n_p = 256$ positions). | 135 |
| 4.29 | Initialisation 2 : positions des centres de classes pour les différents états ($\sigma = 5$ et $n_p = 256$ positions). | 135 |
| 4.30 | Initialisation 2 : taux de reconnaissance en apprentissage ($\sigma = 5$ et $n_p = 256$ positions). Le taux de confusion toléré sur une classe est de $cf_i = 10\%$. | 136 |
| 4.31 | Initialisation 2 : taux de reconnaissance en classement ($\sigma = 5$ et $n_p = 256$ positions). Le taux de reconnaissance global reste supérieur ou égal à 96% . | 136 |
| A.1 | Les deux classes des données <i>clouds</i> . | 152 |
| A.2 | Degrés d'appartenance maximales parmi les appartenances aux 3 ensembles d'apprentissage. | 153 |
| A.3 | Iso-masses de la classe ϕ_1 pour $\alpha = 0.7$ ($T = 5$). | 154 |
| A.4 | Régions/frontières obtenues pour $\alpha = 0.7$ ($T = 5$). Les régions claires correspondent à la classe 1. | 154 |
| A.5 | Iso-masses de la classe ϕ_1 pour $\alpha = 0.38$ ($T = 5$). | 155 |
| A.6 | Régions/frontières obtenues pour $\alpha = 0.38$ ($T = 5$). Les régions claires correspondent à la classe 1. | 156 |
| A.7 | Régions/frontières obtenues avec un seul classifieur Bayésien. Les régions claires correspondent à la classe 1. On constate que le premier mode de la classe 1 est complètement ignoré. | 156 |
| B.1 | Les deux classes dans la position initiale et dans la position 156 (avec la trajectoire de chaque classe). | 160 |
| B.2 | Evolution du taux de reconnaissance en apprentissage lorsqu'il y a chevauchement des classes. | 160 |
| B.3 | Evolution du taux de reconnaissance en classement lorsqu'il y a chevauchement des classes. | 161 |
| B.4 | Les trajectoires de classes lors de l'apprentissage et lors du classement (+5 et -5 sur la caractéristique 1 de la classe 1 et 2 respectivement). Les deux classes sont représentées dans la position initiale en apprentissage et en classement. | 161 |
| B.5 | Evolution du taux de reconnaissance en classement lorsque les trajectoires évoluent. | 162 |

Liste des tableaux

| | | |
|-----|--|-----|
| 1.1 | Exemples de classifieurs et de travaux concernant les classifieurs, en fonction des types de sorties et des types d'apprentissage. | 30 |
| 1.2 | Comparatif de quelques travaux récents. | 49 |
| 3.1 | Exemple de Système d'Information pour le concept "appartenant à ω_1 ". | 84 |
| 3.2 | Partition nette de X^{lr} obtenue par l'algorithme Fc+2M (pourcentages) pour $c = 2$: ϕ_1 et ϕ_2 sont les classes réelles ; ω_1 et ω_2 sont les CCs ; ω_{12} est la CCA. | 92 |
| 3.3 | Partition nette de X^{lr} obtenue par l'algorithme Fc+2M pour $c = 3$: Il y a trois CCs et 4 CCAs possibles (2 d'entre elles, ω_{12} et ω_{13} , existent réellement). | 94 |
| 3.4 | Algorithme 3C ; Matrice de confusion pour $c = 2$ | 94 |
| 3.5 | Comparaison avec le classifieur Bayésien : le mauvais résultat du classifieur Bayésien vient du fait que la deuxième classe est bimodale | 94 |
| 3.6 | Influences de T et de α sur le taux de reconnaissance (en %) de l'ensemble de test et sur n_a | 95 |
| 4.1 | Centres des classes obtenus pour les différents états (4) du système (après le cycle 3). | 121 |
| 4.2 | Taux global de confusion engendrant des changements d'états (le seuil de confusion tolérée sur une classe est de 10 %). | 121 |
| 4.3 | Nombre d'états dans une période pour différents σ (seuil de confusion cf_i toléré : 10 %). | 127 |
| 4.4 | Centres des classes obtenus pour les différents états (4) du système (après le cycle 3). | 131 |
| A.1 | Paramètres des 3 distributions de la classe ϕ_1 | 151 |
| A.2 | Influence de α sur le taux de reconnaissance de l'ensemble d'apprentissage (pour $c = 3$). | 152 |
| A.3 | Matrice de confusion sur un ensemble de test (pour $c = 3$), avec un taux de reconnaissance global de 85.75 %. | 157 |
| A.4 | Comparaison avec le classifieur Bayésien : le mauvais résultat du classifieur Bayésien vient du fait que la première classe comporte 3 modes. | 157 |

Introduction

Dans le but de comprendre le fonctionnement de la vision et de la reconnaissance, un certain nombre d'équipes étudient les relations possibles entre la perception visuelle et l'activité neuronale sous-jacente dans le cerveau. On peut citer, plus particulièrement, l'équipe du Lauréat Nobel Francis Crick et de Christof Koch [65]. En reconnaissance artificielle, l'objectif est, selon Simon [96], de "recueillir les données d'un capteur, c'est-à-dire une représentation (le signifiant¹), et en obtenir une ou des interprétations (les signifiés) par l'exécution d'algorithmes". Une autre définition, qui fait allusion à l'aspect temporel, est la suivante : à partir d'exemples de Formes et de décisions correctes concernant ces Formes, apprendre à décider pour une Forme à venir. Les disciplines scientifiques qui traitent ce type de problème sont essentiellement l'Intelligence Artificielle (IA) et la Reconnaissance des Formes (RdF). L'IA vise à reproduire les facultés humaines les plus élevées tandis que la RdF se limite à la simulation des capacités humaines de perception, visuelles ou auditives.

Dans le cadre de la RdF, nous nous intéressons plus particulièrement à la prise de ces décisions² lorsque les Formes subissent des dérives au cours du temps. Ce type de classe, dont le modèle³ évolue au cours du temps est nommée par la suite classe évolutive ou bien encore classe dynamique.

Dans le contexte de la perception, l'incertitude et la prédiction sont intimement liées. En effet, si nous sommes sûrs qu'un événement est possible, nous sommes prêts à admettre que les chances d'apparition de cet événement dans un futur proche sont élevées. Ainsi, la notion de prédiction est liée à la notion d'ignorance qui est une caractéristique des systèmes à apprentissage. Le macro-système qu'est la société peut avoir diverses attitudes vis-à-vis de ces deux notions. Actuellement, l'attitude défensive est dominante. L'ignorance, en particulier, est qualifiée de "fléau" social qu'il faut éliminer, même si elle se régénère par la naissance de nouveaux "ignorants" que sont les nouveau-nés et par l'innovation scientifique, technologique et culturelle dont une grande partie de la population reste ignorante. Le problème n'est pas d'éliminer l'ignorance, mais plutôt de l'accepter afin de mieux en tenir compte et de lui donner la place qui lui revient dans la société, au lieu de faire abstraction de son existence.

¹Les termes cités entre parenthèses sont de l'auteur de la citation.

²A chaque décision correspond une classe de Formes.

³Exemple : son centre et sa matrice de variance-covariances.

Dans le domaine de la vision humaine, la mise à profit de l'ignorance est effective ; par exemple, lors de l'identification d'un accusé par un témoin, le suspect est prié de se placer parmi un groupe d'individus, au lieu d'être présenté seul au témoin. Un avis plus sûr, ou plus complet, est censé être donné par le témoin dans une situation où son jugement est rendu plus difficile. Dans le cas contraire, lorsque le témoin a en face de lui un seul "suspect", il lui est plus facile de donner un avis. Mais dans ce cas, son avis sera de type binaire : "oui" ou "non". Paradoxalement, plus sa réponse est précise, plus il est permis de douter de son opinion. Ceci est dû au fait que l'avis obtenu est incomplet et fait abstraction, notamment, de son ignorance (de ses doutes).

Dès lors, pour diminuer l'influence de l'ignorance dans les processus de décision, deux approches sont possibles : améliorer l'information et/ou l'éducation du "signifieur"⁴ (dans l'exemple ci-dessus, le témoin) afin de diminuer la part de doute dans son avis, ou bien consulter plusieurs signifieurs, chacun étant spécialisé sur un aspect du problème de reconnaissance et décider en conséquence. Dans le contexte de la reconnaissance, nous sommes limités par notre propre capacité à informer et/ou à éduquer. L'information et/ou l'éducation de manière continue, adaptées à l'instant considéré, constitue une autre limitation. Le second cas présuppose la recherche de stratégies de décision : une combinaison entre les avis et/ou une coopération entre les signifieurs et/ou une sélection des avis des signifieurs. Compte tenu des limitations liées à la première approche, nous avons choisi, par cette étude, de contribuer à la seconde. Nous proposons de rechercher les meilleures stratégies d'associations de ces signifieurs, en fonction de leur capacité à s'adapter à la reconnaissance de classes évolutives.

Dans les deux disciplines (RdF et IA), deux phases se distinguent clairement. L'une consiste à effectuer l'apprentissage (par analogie avec la vision humaine, le témoin supposé doit réellement avoir vu le suspect !) et l'autre à reconnaître, c'est-à-dire à classer en plusieurs catégories ou classes (dans l'exemple considéré : c'est le suspect, c'est peut-être le suspect, ce n'est pas le suspect, etc.). Le cadre de la Reconnaissance d'Objets Naturels fait appel à :

- la RdF : les informations propagées dans le processus de décision sont le plus souvent analogiques, ce qui permet de traduire la grande variété des Formes des objets naturels évolutifs ;
- l'IA : la généralisation s'effectue par accumulation des informations analogiques et apprentissage de modèles de comportement.

Lorsque plusieurs signifieurs ou classifieurs sont amenés à concourir à une même tâche de reconnaissance, plusieurs stratégies de décisions - impliquant ces classifieurs de différentes manières - sont possibles :

1. une première stratégie consiste à décider à la suite de différents avis ; il s'agit de la combinaison de classifieurs ;

⁴Le signifieur fournit les signifiés pour un signifiant donné.

-
2. une deuxième stratégie consiste à utiliser un ou plusieurs avis pour mieux guider d'autres classifieurs, dans leur phase d'apprentissage, et à utiliser un ou plusieurs avis pour améliorer la prise de décision d'autres classifieurs, dans la phase de classement ; il s'agit de la coopération de classifieurs ;
 3. enfin, la troisième et dernière stratégie consiste à privilégier un ou plusieurs classifieurs en fonction de divers critères ou en fonction de la situation ; il s'agit de la sélection de classifieurs.

Dans ce travail, nous montrons qu'il est plus simple de prendre en compte la dynamique d'une classe dans la troisième stratégie. En effet, la sélection des classifieurs peut se faire en fonction de l'évolution du processus : ce que nous appelons, par la suite, sélection dynamique de classifieurs.

L'ignorance peut être gérée à travers les trois stratégies. Différents cadres théoriques gérant l'incertitude, l'imprécision, l'incomplétude, la fiabilité, la subjectivité, l'asynchronisme, la dépendance, la pertinence, ou la complexité peuvent satisfaire cette tâche (Aupriou [5]). Les principaux sont : les théories des probabilités [112], des ensembles flous [114], des possibilités [26], de l'évidence [93] [97], des logiques [2], et des méthodes connexionnistes [66] [102].

En pratique, la prise en compte de l'ignorance pose encore de nombreux problèmes dans la phase de combinaison des avis des classifieurs, en particulier parce que ces derniers gèrent l'ignorance de différentes manières, ou ne la gèrent pas du tout (l'ignorance est alors intégrée dans leurs avis). D'autre part, l'espace de discernement peut ne pas être commun (dans l'exemple, deux témoins peuvent avoir vu le coupable, or on présente à chacun d'eux un autre suspect !). Dans ces cas, la combinaison est souvent difficile, voire impossible, à mettre en œuvre. Exception faite de ces problèmes et à condition que les classifieurs entrant en combinaison soient assez performants (taux de reconnaissance individuels supérieurs à 50 %), la combinaison est aujourd'hui une stratégie éprouvée.

Par rapport à un seul classifieur, la combinaison et la coopération de classifieurs présentent l'avantage de pouvoir spécialiser les classifieurs dans l'espace de représentation ou dans l'espace des caractéristiques. De plus, la coopération permet de spécialiser les classifieurs dans l'espace de discernement de manière explicite, alors que la combinaison peut le faire implicitement. Dans l'une ou l'autre de ces stratégies, malgré le fait que l'ignorance soit également intégrée dans les avis, son importance quantitative est réduite grâce à la spécialisation des classifieurs. Dans la plupart des travaux existants, en particulier ceux qui traitent de la combinaison, les classifieurs retenus ne dépendent pas de la Forme à classer. **Nous proposons, dans le chapitre 3, un système de reconnaissance mettant en œuvre ces deux stratégies, dans lequel un classifieur permet d'activer un ensemble d'autres classifieurs adaptés à la Forme à classer.**

Les classes peuvent avoir une forme complexe lorsque les vecteurs de Formes, acquis à des instants expérimentaux différents, sont issues d'un système non-stationnaire. Dans ce cas, cela veut dire qu'il y a eu intégration dans le temps des vecteurs de Formes des différentes classes. Lorsque le niveau de complexité conduit à des trajectoires (de classes) qui se croisent, les méthodes globales décrites ci-dessus ne suffisent plus. Il devient alors nécessaire de considérer un système de reconnaissance dynamique et une sélection de classifieurs dépendant du temps. **Dans le chapitre 4, nous proposons une méthode de classement utilisant la sélection de classifieurs pour la reconnaissance de classes évolutives. Cette sélection est dynamique.** Cette approche constitue, à notre connaissance, le seul système de reconnaissance dont le comportement est dynamique (en fonction de la Forme à classer et en fonction de l'instant considéré).

Dans ce mémoire, nous commençons par définir les termes de combinaison, de coopération et de sélection de classifieurs et par recenser les travaux déjà effectués dans le cadre de la RdF (chapitre 1). Il s'ensuit que la stratégie de sélection dynamique de classifieurs est la plus appropriée à la reconnaissance de classes évolutives. Par la suite, nous étudions les aspects statiques et dynamiques de la RdF ainsi que les classes dynamiques (chapitre 2). En fonction de critères permettant de juger la complexité de l'évolution de ces classes, deux approches sont dégagées. Les deux chapitres suivants mettent en œuvre chacune d'entre-elles et proposent deux méthodes de RdF :

1. une spécialisation des classifieurs :

Cette étape s'effectue sur des sous-ensembles d'apprentissage de l'espace de représentation, obtenus par une partition qui met en évidence les zones denses d'informations (sur lesquelles les classifieurs seront spécialisés) et des zones d'ambiguïtés, faiblement denses. La redondance des classifieurs est alors exploitée sur ces zones grâce à une méthode de fusion. La méthode proposée est une méthode de sélection adaptative de classifieurs mettant en œuvre la coopération et la combinaison de classifieurs (chapitre 3). Nous montrons que, pour les classes multimodales, la méthode permet d'obtenir des taux de reconnaissance largement supérieurs à ceux d'un seul classifieur ;

2. une séparation des aspects statiques et dynamiques de la RdF :

Deux types d'apprentissage sont effectués : l'aspect statique de l'apprentissage est pris en charge par des classifieurs ou des systèmes de classifieurs, et l'aspect dynamique est pris en charge à travers l'apprentissage de la planification des différents états par un réseau de Petri flou. La méthode proposée effectue une sélection dynamique de classifieurs avec combinaison de classifieurs (chapitre 4). Nous montrons que ce type de sélection permet de reconnaître des classes dynamiques, sans nécessiter la recherche de modèles d'évolution des classes.

Enfin, nous proposons quelques conclusions et perspectives.

Chapitre 1

Combinaison, coopération et sélection de classifieurs

Ce chapitre établit un état de l'art des stratégies de prises de décisions possibles lorsque plusieurs classifieurs interviennent dans un processus de RdF. Cette association de classifieurs est présentée comme un système capable de mettre en œuvre plusieurs types d'opération : la combinaison qui est analogue à une prise de décision démocratique, la coopération analogue à la concertation (avec échanges de points de vue) et la sélection qui est analogue à la consultation d'experts (classifieurs spécialisés). L'aspect temporel de la RdF, c'est-à-dire l'évolution possible des classes à reconnaître, est traité dans le cadre de la sélection de classifieurs.

Sommaire

| | | |
|------------|--|-----------|
| 1.1 | Introduction | 25 |
| 1.2 | La RdF et le classifieur | 26 |
| 1.2.1 | Les étapes de la RdF | 26 |
| 1.2.2 | Le classifieur | 27 |
| 1.2.3 | Taxonomies de classifieurs | 28 |
| 1.3 | Les systèmes de classifieurs | 30 |
| 1.3.1 | Objectifs et intérêts | 30 |
| 1.3.2 | Définition | 32 |
| 1.3.3 | Architecture | 34 |
| 1.3.4 | Taxonomie des systèmes de classifieurs | 36 |
| 1.4 | Performances des classifieurs et des systèmes de classifieurs | 37 |
| 1.4.1 | Les différentes approches | 37 |
| 1.4.2 | La matrice de confusion | 38 |
| 1.5 | Combinaison de classifieurs | 38 |
| 1.5.1 | Combinaison usuelles | 41 |
| 1.5.2 | Combinaison par vote | 42 |
| 1.5.3 | Combinaison par la Théorie Bayésienne de la décision | 42 |
| 1.5.4 | Combinaison par la Théorie de l'Evidence | 42 |
| 1.5.5 | Autres cadres théoriques de la combinaison | 45 |
| 1.6 | Coopération de classifieurs | 45 |
| 1.7 | Sélection de classifieurs | 46 |
| 1.8 | Système de classifieurs hybride | 48 |
| 1.9 | Conclusion | 51 |

1.1 Introduction

La Reconnaissance des Formes (RdF) est la discipline scientifique qui a pour objectif le classement d'*individus* en un certain nombre de classes ou catégories⁵. En fonction de l'application, un *individu* est un ensemble de mesures caractérisant un objet, une image (ou une partie de celle-ci), un signal ou un environnement (cas du diagnostic par RdF). Le terme générique *Forme* est parfois mieux adapté.

Dans le cadre de la RdF pour des objets naturels subissant des évolutions dans le temps, nous avons à traiter des classes de forme complexe (complexité due en particulier aux évolutions de ces objets). Un exemple de problème réel est posé lors de la reconnaissance des différentes espèces de poissons vivants dont les caractéristiques⁶ (par exemple, la couleur) subissent des évolutions saisonnières. Dans ce chapitre, nous recensons les approches existantes qui traitent cette problématique d'un point de vue multi-décisionnel (i.e., faisant intervenir les décisions de plusieurs classifieurs). Pour ces classifieurs individuels, nous justifierons, dans la section suivante, le choix des méthodes statistiques parmi les techniques existantes.

La RdF comporte généralement deux étapes. La première étape consiste à effectuer un apprentissage tandis que la deuxième consiste à effectuer un classement (et/ou une prédiction). Lorsque l'apprentissage est effectué de manière automatique, nous parlerons de méthode de RdF non-supervisée. Lorsqu'il est guidé par un expert, nous parlerons de méthode de RdF supervisée. D'un point de vue formel, ce type de problème peut s'exprimer comme suit :

- on considère n individus, chacun décrit par p caractéristiques définissant ainsi un vecteur de Forme⁷ x_k , appartenant à un sous-ensemble de \mathbb{R}^p où $k \in [1, n]$, n est le nombre de Formes ;
- chaque individu appartient à une classe ω_j (ou catégorie lorsqu'il s'agit d'une méthode de RdF non-supervisée) du cadre de discernement

$$\Omega = \{\omega_1, \dots, \omega_i, \dots, \omega_l\}$$

- le but est de classer un nouvel objet x_k dans l'une des l classes.

Souvent, deux décisions supplémentaires sont introduites ; le rejet d'ambiguïté (hésitation entre plusieurs classes) et le rejet d'individus aberrants (en Anglais : "outliers").

⁵D'où les expressions *classification* et *catégorisation* pour désigner une méthode regroupant la population en l classes ou catégories.

⁶Certains auteurs utilisent le terme "caractère" et d'autres le terme "paramètre". Ce dernier prend, parfois, une autre signification (par exemple, lorsqu'il s'agit de paramètres seuils de certains algorithmes ou de paramètres d'un modèle).

⁷Nous utilisons également le terme "individu" ou encore le terme "objet".

Ainsi, une Forme peut être représentée par un point x_k dans l'espace (de représentation) de dimension p . Dans le cas des méthodes statistiques de la RdF, les coordonnées du point $x_k \in \mathbb{R}^p$ sont définies par les différentes caractéristiques extraites de la Forme. Une phase *d'analyse* a permis de sélectionner les plus pertinentes (i.e., celles qui permettent le mieux de distinguer une classe d'une autre dans l'univers d'interprétation) des caractéristiques observables par quantification (évaluation numérique) et/ou par qualification (évaluation symbolique). Ce travail traite exclusivement les aspects numériques.

1.2 La RdF et le classifieur

1.2.1 Les étapes de la RdF

Les p caractéristiques observées sont utilisées pour construire les vecteurs de Formes $x_k \in \mathbb{R}^p$, tel que :

$$x_k = (x_{k1}, x_{k2}, \dots, x_{kp})^t$$

Si les p caractéristiques concernent des individus (au sens général) non-manufacturés tels que, par exemple, des visages humains, des fruits ou des poissons, on dira alors que les individus en question sont des objets naturels.

La reconnaissance est liée aux classes d'individus à discriminer dans l'espace de représentation à p -dimensions. La plupart des processus de reconnaissance sur des objets naturels sont considérés complexes : un seul classifieur, gérant toutes les classes et toutes les caractéristiques, est alors souvent mal adapté ; l'ensemble d'apprentissage présente une structure de données trop complexe pour une modélisation correcte des différentes classes.

Theodoridis et al. [101] proposent de résumer les différentes étapes de la construction d'un système de classement par celles représentées sur la figure 1.1. Nous nous intéresserons plus particulièrement aux deux dernières étapes. Les trois étapes qui précèdent sont largement traitées dans la littérature et détaillées notamment par ces auteurs. D'une manière générale, la conception d'un système de classement nécessite les phases suivantes :

- mesures par un capteur ou un ensemble de capteurs,
- extraction de caractéristiques à partir de ces mesures,
- sélection des caractéristiques les plus pertinentes,
- choix de la méthode de classement,

- construction du classifieur,
- évaluation de l'apprentissage, souvent par le classement de l'ensemble d'apprentissage,
- évaluation des performances du système, généralement par le taux d'erreur en classement.

Les techniques d'extraction des caractéristiques sont parfois qualifiées de techniques de discrimination à but descriptif [90]. Le choix des caractéristiques extraites est plus lié au problème à résoudre (l'application spécifique) que le choix de la méthode de classement. La construction du classifieur dépend de l'approche et de la technique choisie comme méthode de classement. Pour un réseau de neurones, il s'agit de déterminer la matrice de pondération, tandis que pour un classifieur statistique telle que le Bayésien, il s'agit de déterminer les matrices de variance-covariances et les moyennes des différentes classes. Il est évident qu'en fonction des performances obtenues, on peut modifier le comportement en modifiant les paramètres des phases précédentes ou en choisissant d'autres caractéristiques, afin d'optimiser le système de manière plus ou moins automatique.

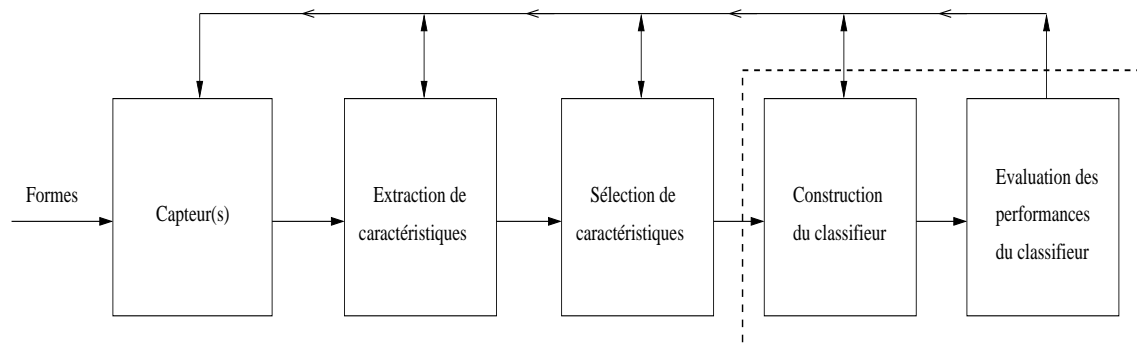


FIG. 1.1 – Les différentes étapes de la construction d'un système de classement.

1.2.2 Le classifieur

Soient :

- p , le nombre de caractéristiques des Formes,
- $X \subset \mathbb{R}^p$, un ensemble de vecteurs de Formes,
- Ω , le cadre de discernement de taille l (nombre de classes d'interprétation),

- 2^Ω , l'ensemble de puissance (i.e., tous les sous-ensembles possibles de Ω),
- $x \in X$, un individu à classer,
- et ϕ , le vecteur d'interprétation issu du classifieur e appelé vecteur de décision de e et de taille l .

Afin de faire abstraction de l'architecture interne d'un classifieur, nous proposons de définir un classifieur par une fonction e , de la manière suivante :

$$\begin{aligned} e &: X \longrightarrow 2^\Omega \\ x &\longmapsto \phi = e(x) \end{aligned}$$

Ainsi, le classifieur est vu comme une "boite noire" qui reçoit, en entrée, un ensemble de caractéristiques et qui fournit, en sortie, le ou les appartenances de l'individu (présenté en entrée) aux différentes classes.

1.2.3 Taxonomies de classifieurs

On peut les classer suivant différents critères. Ci-dessous, nous présentons les trois taxonomies les plus couramment citées et employées : selon le type de méthodes utilisées, selon le type des sorties de classifieurs et selon l'autonomie des classifieurs, i.e., de leur dépendance vis-à-vis d'un expert. D'autres critères peuvent être considérés. Ils conduisent à d'autres taxonomies de classifieurs. La dichotomie en méthodes paramétriques et non-paramétriques se fonde sur le type de représentation des classes par un classifieur. Ainsi, les méthodes qui tentent de représenter les densités des classes par des distributions de probabilités (modèles mathématiques) sont qualifiées de méthodes paramétriques (exemple : classifieur Bayésien). Les autres (exemples : k-ppv, noyaux de Parzen) sont qualifiées de méthodes non-paramétriques.

Taxonomie selon les techniques ou méthodes

Les quatre approches les plus connues et utilisées sont les suivantes :

1. Mise en correspondance et en coïncidence : La Forme à reconnaître (typiquement, une Forme 2D) est comparée aux différents prototypes ou gabarits représentant les différentes classes, tout en permettant des translations, des rotations et des changements d'échelles (Jain et al. [61]). Ainsi, on travaille directement dans l'espace de représentation, sans passer par une phase d'extraction de caractéristiques. La mesure de similarité entre la Forme à reconnaître et le prototype est souvent fondée sur une corrélation.

2. Méthodes syntaxiques et structurales : Une analogie formelle est faite entre la structure des Formes et la syntaxe d'une langue. Ces méthodes gèrent les *Formes* sous forme de "mots" et de "phrases" respectant les grammaires des différentes classes. A travers cette approche, une perspective hiérarchique est privilégiée ; une Forme peut être divisée en sous-Formes, comme une phrase peut être divisée en mots.
3. Méthodes connexionnistes (appelées "réseaux de neurones", par abus de langage) : Un neurone est doté de p entrées $x_i, i \in [1, p]$, chacune affectée d'un poids a_i et d'une sortie S . S est fonction de la somme des $a_i x_i$. La communauté des réseaux de neurones travaille essentiellement sur l'assemblage de telles entités et sur l'apprentissage des poids a_i (exemple : Perceptron Multi Couches).
4. Méthodes statistiques : Cette approche préconise la représentation des Formes par p caractéristiques, sur lesquelles des densités, propres à chaque classe, sont définies. Ensuite, les outils de la théorie de la décision statistique permettent d'effectuer un classement. Les méthodes paramétriques émettent l'hypothèse que ces densités appartiennent à une famille donnée de distributions (exemple : classement Bayésien). Au contraire, les méthodes non-paramétriques ne posent pratiquement aucune hypothèse restrictive sur la forme des distributions (exemple : C moyennes).

Les deux premières approches nécessitent des moyens de calculs importants, ce qui pose des problèmes pour la reconnaissance d'objets naturels évolutifs. Quant à la troisième approche, elle peut être vue comme une alternative à la quatrième. En effet, certains auteurs considèrent que les méthodes connexionnistes constituent un sous-ensemble des méthodes statistiques (par exemple, Holmström et al. [56]). Nous concentrons nos efforts sur les méthodes statistiques où les classes sont représentées par des moments du premier et/ou du second ordre. Les phases d'apprentissage et de test seront étudiées et évaluées à l'aide de critères de performances fondés sur la matrice de confusion.

Pour chacune de ces approches, de nombreux travaux ont conduit à l'élaboration d'une importante quantité de classifieurs. La Table 1.1 présente quelques-uns des ces classifieurs et quelques travaux concernant les classifieurs. Les termes "abstrait", "classement" et "distance" utilisés dans ce tableau sont expliqués ci-dessous, à travers deux taxonomies. Un état de l'art concernant les méthodes statistiques a été récemment proposé par Jain et al. [61].

Taxonomie selon les types de sorties des classifieurs

Selon Xu et al. [112], trois modèles sont possibles. Cette taxonomie (ou une variante de celle-ci) a également été adoptée par Ho [52] et par Loonis [72] :

1. Niveau abstrait ; le classifieur propose une classe ou plusieurs classes "également possibles".

| Niveaux : | Type d'apprentissage | |
|------------|------------------------|--------------------------|
| | Supervisé | Non-supervisé |
| Abstrait | Arbre de décision [54] | C moyennes [50] |
| Classement | k-ppv [24] | "clustering" k-ppv [113] |
| Distance | Bayésien [61] | C moyennes floues [9] |

TAB. 1.1 – Exemples de classifieurs et de travaux concernant les classifieurs, en fonction des types de sorties et des types d'apprentissage.

2. Niveau classement ; le classifieur applique une relation d'ordre sur les classes possibles.
3. Niveau distance ; le classifieur attribue à chaque classe une mesure de similarité entre celle-ci et l'individu à classer. Pour les méthodes statistiques paramétriques, il s'agit d'estimations des lois de probabilités *a posteriori*.

Les types de sorties les plus riches en informations sont ceux du niveau distance. En effet, à partir de ce type de sorties, on peut établir un ordre de préférence du classifieur (on obtient alors une sortie de niveau classement), puis de retenir que la première ou les premières classes proposées (auquel cas, on obtient une sortie du niveau abstrait). De ce fait, le niveau abstrait est le niveau le plus générique.

Taxonomie selon les types d'apprentissage

1. Apprentissage supervisé (exemple : k-ppv). Un expert indique la classe réelle pour chaque individu présenté au système de RdF lors de l'apprentissage.
2. Apprentissage non supervisé (exemple : C moyennes floues). Les méthodes de catégorisation et de coalescence effectuent ce type d'apprentissage.

1.3 Les systèmes de classifieurs

1.3.1 Objectifs et intérêts

Les premières allusions aux *systèmes de classifieurs* ou *systèmes multi-classifieurs*⁸ (ou bien encore *ensemble* de classifieurs⁹, terme utilisé surtout par la communauté des scientifiques traitant les réseaux neuro-mimétiques connexionnistes), citées par Ho [52], remontent à Nilsson[81] et à Haralick[49] pour le calcul d'erreurs concernant un ensemble

⁸Autres termes employés dans la littérature ; combinaison multi-experts, comité de classifieurs, fusion de classifieurs et autres.

⁹Le terme "ensemble" est repris dans la littérature anglophone.

de classifieurs Bayesiens. Cependant, c'est à partir des années 90 que les systèmes de classifieurs ont été mis en œuvre concrètement, notamment dans le cadre de la reconnaissance de mots (Ho, 1992 [52]) et de l'écriture manuscrite (Xu et al., 1992 [112]; Suen et al., 1990 [99], 1992 [98]). Il est important de préciser que le terme *système classifieur* est utilisé par la communauté des algorithmes génétiques (voir notamment la thèse de Giani [38]) dans le sens d'un système expert dont le moteur de recherche est fondé sur un algorithme génétique pour générer des règles de comportement, en interaction avec son environnement.

Dans la RdF, l'objectif est d'obtenir un système de classifieurs (i.e., une association de classifieurs) hautement performant fondé sur les décisions d'un ensemble de classifieurs (Xu et al., 1992 [112]; plus récemment, Giacinto et Roli, 1999 [35]). Une autre motivation pour ce type de système réside dans la complexité des formes des classes traitées; lorsque les classes sont multimodales ou lorsqu'elles se chevauchent, il devient difficile de les modéliser, parce que leur représentation (par exemple, par des centres, des distributions de probabilités ou des ensembles flous) devient compliquée. Certes, il existe des méthodes où interviennent des modèles de mélange (voir notamment les travaux de McLaclan et Basford [76]) aptes à modéliser de telles classes, mais dans les cas où la taille de l'ensemble d'apprentissage est limitée, cette modélisation conduit à une sur-adaptation aux données disponibles qui ne représentent pas forcément bien les classes réelles. En associant différents classifieurs (algorithmes différents et/ou constructions différentes), on espère obtenir de meilleures performances. Enfin, dans le cadre de la reconnaissance d'objets évolutifs, il est utile de déterminer les stratégies d'association les mieux adaptées à l'aspect temporel des classes dynamiques.

Les intérêts de l'utilisation de systèmes de classifieurs sont les suivants :

- De la même manière qu'une caractéristique supplémentaire permet à un classifieur de mieux décider (à condition qu'elle soit complémentaire aux autres caractéristiques), un avis supplémentaire provenant d'un autre classifieur permet à un système classifieurs de mieux décider (à condition que cet avis soit complémentaire aux autres avis).
- Ce type d'associations favorise une approche modulaire et répartie d'un système de RdF.
- Lorsque pour un problème de RdF donné, certaines caractéristiques sont réelles et d'autres sont discrètes ou symboliques, il est souvent inévitable d'utiliser deux classifieurs différents. Grâce à un espace de discernement commun, les avis des deux classifieurs peuvent être facilement combinés.
- Dans certains cas, la complexité algorithmique (en phase d'apprentissage et/ou en phase de classement) peut être réduite si le traitement des données est réparti sur plusieurs niveaux.
- Enfin, pour les classes évolutives, il peut être nécessaire de spécialiser plusieurs clas-

sifieurs sur différents intervalles temporels, et d'employer ces classifieurs en fonction du temps et/ou en fonction de l'individu à classer.

Actuellement, le type d'opération le plus couramment utilisé pour associer les différents classifieurs d'un système est le type "combinaison" des sorties de classifieurs (Giacinto et al. [35]). Récemment, certains chercheurs ont étudié un autre type d'opération, appelé "sélection adaptative de classifieurs" (Giacinto et al. [33]) ou "sélection dynamique de classifieurs" (Ho et al. [53]; Woods et al. [111]). Cependant, il est à noter que la très grande majorité des travaux actuels concernent la combinaison seulement.

D'autres auteurs ont proposé d'utiliser les sorties de classifieurs individuels en tant que caractéristiques d'entrée d'un autre classifieur (de "décision"), effectuant l'apprentissage sur les décisions de ces classifieurs (Breukelen et al. [108]; Achermann et Bunke [1]). Cet état de l'art concerne, en plus de la combinaison de classifieurs, les travaux existants (à notre connaissance) dans le cadre d'autres stratégies d'associations de classifieurs.

Enfin, très récemment, une méthode de conception automatique de "systèmes multi-classifieurs" a été proposée (Giacinto et al. [36]). Sa proposition consiste à choisir un ensemble de classifieurs de méthodes et/ou de paramètres différents, à sélectionner (avec un ensemble de validation différent de l'ensemble de test) le sous-ensemble des classifieurs dont les performances sont élevées. Ensuite, les classifieurs candidats qui font double emploi (ceux dont les erreurs sont corrélées avec les erreurs d'un ou plusieurs autres classifieurs) sont éliminés. Enfin, en phase de classement, les sorties des classifieurs retenus sont combinées par la règle de vote majoritaire.

Lors de la mise en œuvre d'un système de classifieurs, les phases nécessaires à la conception du système de reconnaissance restent les mêmes que pour un seul classifieur (cf. section précédente), à la différence près qu'il s'agit, désormais, de construire une association de classifieurs et d'évaluer les performances de cet ensemble. Ci-dessous, nous proposons une formalisation des systèmes de classifieurs.

1.3.2 Définition

Soit $E = \{e_i\}, i \in [1, S]$, l'ensemble des classifieurs, où S est le nombre de classifieurs intervenant dans le système de classifieurs et $\Omega = \{\omega_i\}_{i \in [1, l]}$, les différentes classes d'interprétation. Un système de classifieurs doit permettre d'obtenir des performances supérieures à celles d'un seul classifieur. Ces performances peuvent être déterminées selon certains critères. Généralement, le taux de reconnaissance (cf. section suivante) sur un ensemble de test constitue un bon critère. Nous proposons de formaliser un système de classifieurs (SC) à l'aide de la définition 1 et d'une fonction \mathcal{R} .

Définition 1 *Un système de classifieurs est un ensemble de classifieurs associés mettant en œuvre la combinaison et/ou la coopération et/ou la sélection de classifieurs. Il peut être défini formellement par un ensemble de classifieurs et un ensemble de relations entre ces classifieurs.*

$$\begin{aligned} SC &: X \longrightarrow 2^\Omega \\ x &\longmapsto \phi = SC(x) \end{aligned} \quad (1.1)$$

où $x \in X$ est un individu et ϕ le vecteur de décision sur l'ensemble de puissance 2^Ω .

Une fonction \mathcal{R} peut représenter les relations entre les classifieurs, les termes r_{ij} , $i, j \in [1, S]$ servant à qualifier les types de relations entre deux classifieurs :

$$\begin{aligned} \mathcal{R} &: E \times E \longmapsto \{0, 1\}, \\ & [e_1 \quad \cdots \quad e_j \quad \cdots \quad e_S] \\ \mathcal{R} &= \begin{bmatrix} r_{11} & \cdots & r_{1j} & \cdots & r_{1S} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ r_{i1} & \cdots & r_{ij} & \cdots & r_{iS} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ r_{S1} & \cdots & r_{Sj} & \cdots & r_{SS} \end{bmatrix} \begin{bmatrix} e_1 \\ \vdots \\ e_i \\ \vdots \\ e_S \end{bmatrix} \end{aligned}$$

Les relations r_{ij} peuvent prendre deux valeurs :

1. $r_{ij} = 0$: Pour $i \neq j$, le classifieur e_i ne fournit pas d'informations au classifieur e_j . Si $i = j$, le classifieur e_i n'est pas sélectionné.
2. $r_{ij} = 1$: Le classifieur e_i fournit des informations au classifieur e_j (exemple : il lui fournit son vecteur de décision). La sortie de ce dernier est combinée avec celles des autres classifieurs. Si $i = j$ (e_i coopère avec lui-même), la sortie du classifieur e_i est également à combiner.

Les relations dépendent des types de classifieurs impliqués. Chaque classifieur peut se comporter d'une manière différente des autres classifieurs et chaque classifieur peut fonctionner avec un espace des caractéristiques différent de ceux des autres classifieurs. Cependant, les classifieurs doivent contribuer (éventuellement, à différents niveaux) à une tâche commune qui consiste à classifier un individu. Les décisions des classifieurs individuels doivent ensuite être combinées selon l'architecture retenue afin d'obtenir une décision finale qui sera alors la sortie du système.

Si la fonction \mathcal{R} est constante, alors la sélection de classifieurs est statique. Dans le cas contraire, la sélection est adaptative ou dynamique (nous ferons la distinction entre ces deux types de sélection dans la section consacrée à la sélection de classifieurs). Dans l'exemple suivant, \mathcal{R}_1 représente un système de classifieurs multi-niveaux (équivalent à la figure 1.2) :

$$\mathcal{R}_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1.2)$$

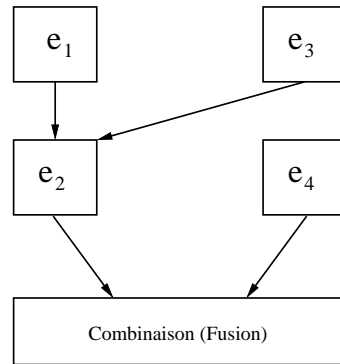


FIG. 1.2 – Exemple de *Système de Classifieurs* (selon \mathcal{R}_1) : les classifieurs e_1 et e_3 fournissent des informations au classifieur e_2 (car $r_{12} = 1$ et $r_{32} = 1$), les décisions des classifieurs e_2 ($r_{22} = 1$) et e_4 ($r_{44} = 1$) sont combinées pour obtenir un vecteur de décision final.

1.3.3 Architecture

La figure 1.3 illustre les interactions possibles entre les différents classifieurs d'un système mettant en œuvre les différentes opérations de combinaison, de coopération et de sélection. Les liaisons en pointillés, entre les différents "modules" représentent des informations diverses. Ces informations peuvent être communiquées à la fois lors de l'apprentissage et lors du classement. En ce sens, le module de sélection est un module intelligent qui retient un certain nombre d'informations issues de la phase d'apprentissage et les utilise dans la phase de classement. Un classifieur peut transmettre ses décisions à un ou plusieurs autres classifieurs.

Exemples d'informations échangées :

- Le classifieur e_1 communique à e_3 son ensemble d'apprentissage (phase d'apprentissage du système),
- Le classifieur e_1 informe le module de sélection qu'il faut prendre ou non en considération la sortie du classifieur e_3 (phase de classement),
- Le module de sélection "demande une réponse" à e_1 et e_2 , concernant un individu inconnu présenté en entrée.

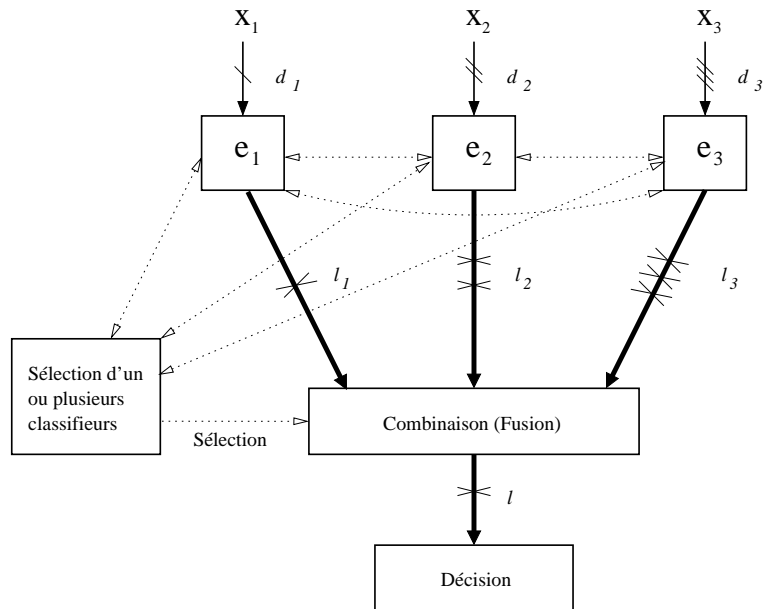


FIG. 1.3 – Système de classifieurs effectuant les différentes opérations de combinaison, de coopération et de sélection de classifieurs.

Les liaisons en traits pleins représentent les caractéristiques (celles qui sont barrées de traits) ou les décisions (barrées de croix). Dans cette configuration, les vecteurs de caractéristiques X_1 , X_2 et X_3 concernent un même individu à classer, mais ils peuvent être différents les uns des autres. Si on dispose de d caractéristiques, on a $d_1 \leq d$, $d_2 \leq d$ et $d_3 \leq d$. Les classifieurs e_1 , e_2 et e_3 n'ont pas nécessairement un espace de discernement commun. La dimension de Ω étant l , on a $l_1 \leq l$, $l_2 \leq l$ et $l_3 \leq l$. Lorsque un seul classifieur est sélectionné, il n'y a pas d'opération de combinaison. Par contre, si on veut retenir les résultats de plusieurs classifieurs afin de les combiner, il sera alors préférable que les classifieurs sélectionnés aient un espace de discernement commun. Lorsque cette dernière contrainte n'est pas respectée, c'est, en général, à travers la théorie des croyances que la combinaison est traitée.

Exemple : Dans le cas où la sélection préconise le choix des résultats des classifieurs e_1 et e_2 , il est préférable que les l classes soient les mêmes d'un classifieur à un autre ;

$$\begin{cases} l_1 = l_2 = l \\ \text{et } \Omega_1 = \Omega_2 = \Omega \end{cases} \quad (1.3)$$

Le module de sélection peut être fondé sur un algorithme de catégorisation dont le résultat peut être utilisé en classement pour sélectionner des classifieurs. (c'est le cas pour le système 3C, proposé au chapitre 3). D'autre part, un classifieur e_i peut lui-même être constitué de plusieurs classifieurs (i.e., d'un système de classifieurs).

1.3.4 Taxonomie des systèmes de classifieurs

Récemment, Giacinto et al. [35] ont insisté sur la possibilité de faire intervenir différents types d'opérations dans les systèmes de classifieurs. En plus de l'opération "combinaison" des sorties de classifieurs, des chercheurs ont récemment introduit d'autres types d'opérations. Gosselin [39] utilise les termes de *coopération* et *combinaison* dans le même sens. Franke et Mandler [32] font également cette confusion tout en proposant deux types d'assemblages différents ; "les sorties combinées de classifieurs coopérants servant de caractéristiques à un classifieur de plus haut niveau" ou "combinaison des sorties de classifieurs s'interprétant comme des votes de spécialistes automatiques possédant chacun un certain point de vue sur l'individu en entrée". Nous appelons ces deux types d'associations *coopération* de classifieurs et *combinaison* ou *fusion* de classifieurs, respectivement. Nous proposons de distinguer les systèmes de classifieurs en quatre catégories (figure 1.4), selon les types d'opérations entre les classifieurs :

- combinaison de classifieurs,
- coopération de classifieurs,
- sélection de classifieurs,
- systèmes hybrides ou mixtes.

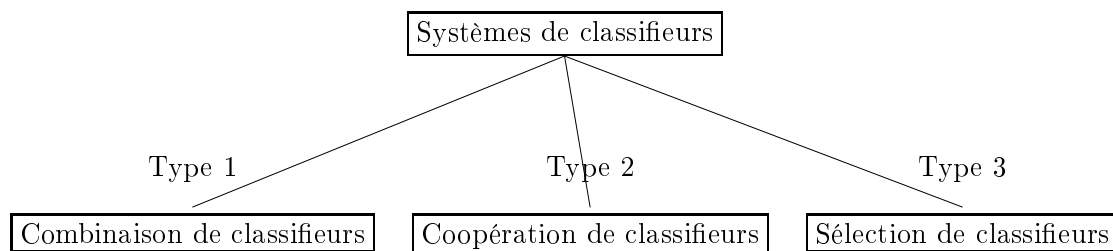


FIG. 1.4 – Les différents types d'associations de classifieurs. Si un système opère avec plusieurs types d'associations, le système est dit hybride.

Les deux premiers types d'opérations (ou associations) apparaissent également comme des préoccupations majeures dans la communauté des réseaux de neurones. Selon Sharkey [95], cette communauté utilise principalement deux termes ; *combinaison d'un ensemble* (de classifieurs) et *combinaison de modules* (de classement). Ainsi, le terme *combinaison* est ici utilisé dans le sens d'association. Selon cet auteur, "*dans un ensemble, les réseaux sont redondants dans le sens où ils fournissent tous une solution pour une même tâche (de classement). Par contre, l'approche modulaire considère qu'une tâche est décomposée en plusieurs sous-tâches et qu'une solution complète de la tâche nécessite la contribution de plusieurs modules (même si des entrées individuelles peuvent être traitées par un seul de ces modules)*". Il est clair que l'approche *ensemble* est une expression de la combinaison de classifieurs tandis que l'approche *modulaire* est une expression de la coopération

de classifieurs. Le contenu des dernières parenthèses de cette citation fait implicitement allusion à la sélection (adaptative, voir la définition de ce terme plus loin) de classifieurs. L'auteur précise également que la notion de sélection des membres d'un *ensemble* a été abordée dans un certain nombre de publications et propose de s'en servir explicitement. Sa proposition est de déterminer les performances des différents *ensembles* de classifieurs (parmi un plus grand nombre, fixe, de classifieurs) et ceci en les testant (méthode "test and select") sur un ensemble de validation, différent à la fois de l'ensemble d'apprentissage et de l'ensemble de test. Mais les inconvénients de cette approche sont que, d'une part, la phase de validation/constitution d'un ensemble nécessite un effort de sélection/mise au point important et d'autre part, que l'*ensemble* retenu est fixe et ne peut changer en fonction d'un individu à classer.

Certains auteurs, en s'intéressant aussi à la complexité des algorithmes, ont proposé d'effectuer des combinaisons conditionnelles. Ainsi, Gosselin [39] propose de classer les classifieurs selon leur performances et de traiter un individu inconnu par le premier classifieur. Il propose d'accepter sa décision, si l'individu n'est pas rejeté. Dans le cas contraire, la décision sera prise suite à la combinaison du premier classifieur avec la sortie du deuxième classifieur. Le même raisonnement peut s'appliquer, jusqu'à ce que l'individu soit classé ou que les sorties de tous les classifieurs soient combinées. Cette combinaison conditionnelle permet de réduire les temps d'exécution efficacement. L'inconvénient est la nécessité de fixer de multiples seuils de rejets associés aux différentes décisions.

Le tableau de la figure 1.8, en fin de ce chapitre, présente de façon synthétique les travaux existants. Ces travaux sont présentés selon la taxonomie ci-dessus ainsi qu'en fonction des types de spécialisations des classifieurs. Dans la section qui suit, nous verrons que ces spécialisations sont d'une importance primordiale dans le cadre de la combinaison de classifieurs.

1.4 Performances des classifieurs et des systèmes de classifieurs

1.4.1 Les différentes approches

Le calcul de la matrice de confusion (voir plus loin) est la solution la plus connue et la plus employée. Cependant, d'autres mesures, telles que celles issues de la théorie de l'information existent. Récemment, Loonis [73] a proposé l'Outil d'Analyse Informationnelle (OAI) fondé sur la théorie de l'information, qui évalue la qualité d'une méthode de RdF classe par classe.

L'ensemble des individus disponibles est souvent divisé en plusieurs sous-ensembles. Le plus souvent, cet ensemble est divisé en deux, la moitié servant à l'apprentissage (ensemble d'apprentissage), l'autre moitié (ensemble de test) servant à tester le classifieur ou le système. Sharkey [95] subdivise l'ensemble en trois sous-ensembles (pas forcément de même importance numérique), l'ensemble supplémentaire servant à ordonner les clas-

sifieurs en fonctions de leurs performances (validation des classifieurs). Twomey et Smith [104] utilisent une technique de validation croisée : si on dispose de n individus, on se sert de $(n - 1)$ individus pour l'apprentissage, puis on utilise l'individu qui reste pour la validation, et ceci dans les n voies possibles.

1.4.2 La matrice de confusion

Pour définir cette matrice, définissons d'abord le taux de reconnaissance et le taux de confusion. Soit l , le nombre de classes réelles. Soient n_{ii} , l'effectif des individus appartenant à une classe i (d'après un expert) et affectés à une même classe i (par un classifieur), et n_{ij} , l'effectif des individus appartenant à une classe i et affectés à une classe j (i et $j \in [1, l]$). Le taux de reconnaissance est le taux des effectifs correctement classés.

Définition 2 *Le taux de reconnaissance global τ_g est le rapport de la somme des éléments n_{ii} sur l'effectif total de l'ensemble d'apprentissage.*

$$\tau_g = \frac{\sum_{i=1}^l n_{ii}}{\sum_{i=1}^l \sum_{j=1}^l n_{ij}} \quad (1.4)$$

Définition 3 *Le taux de confusion est la mesure duale du taux de reconnaissance :*

$$cf_g = 1 - \tau_g \quad (1.5)$$

La matrice de confusion dont les termes interviennent dans l'équation 1.4 est définie par la matrice \mathcal{M} tel que ;

$$\mathcal{M} = \begin{bmatrix} n_{11} & \cdots & n_{1j} & \cdots & n_{1l} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ n_{i1} & \cdots & n_{ij} & \cdots & n_{il} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ n_{l1} & \cdots & n_{lj} & \cdots & n_{ll} \end{bmatrix} \quad (1.6)$$

Dans cette définition, les termes de la matrice sont exprimés en nombre d'individus, mais souvent, ils sont aussi exprimés en pourcentages de la classe réelle i . Calculée en classant l'ensemble d'apprentissage, cette matrice peut servir dans le cadre de la combinaison de classifieurs. En effet, Xu et al. [112] et plus récemment, Parker [82], ont proposé d'utiliser cette matrice pour la détermination des probabilités *a priori* servant à la combinaison.

1.5 Combinaison de classifieurs

La combinaison de classifieurs¹⁰ se place dans le cadre plus général de combinaison multi-sources. Les problèmes rencontrés sont ceux que l'on retrouve aussi dans les domaines suivants :

¹⁰Dans la littérature, on parle aussi de fusions de classifieurs et de fusion de données.

- combinaison de capteurs,
- combinaison d’avis d’experts,
- et combinaison de bases de données (de leurs réponses).

Une application particulière de la combinaison de capteurs est la fusion de données en traitements d’images, mis en œuvre, par exemple, par Bloch et Maître [11], et par Mathévet et al. [75]. Les informations fournies par ces sources (capteurs, experts humains, bases de données, classifieurs) peuvent être de nature et de type très variés. En effet, elles peuvent constituer un signal, un son ou une image, et être de type numérique et/ou symbolique. Malheureusement, les sources sont, souvent, entachées d’imprécision et/ou d’incertitude et aussi d’incomplétude et/ou d’ambiguïté. La figure 1.5 illustre les interactions possibles entre les différents modules d’un système de classifieurs mettant en œuvre la combinaison. Dans cette configuration, les vecteurs de caractéristiques choisis pour les différents classifieurs peuvent être différents les uns des autres. Le plus souvent, les classifieurs e_1 , e_2 et e_3 ont un espace de discernement commun, i.e., les l classes ou hypothèses en sorties des différents classifieurs ont les mêmes significations d’un classifieur à un autre.

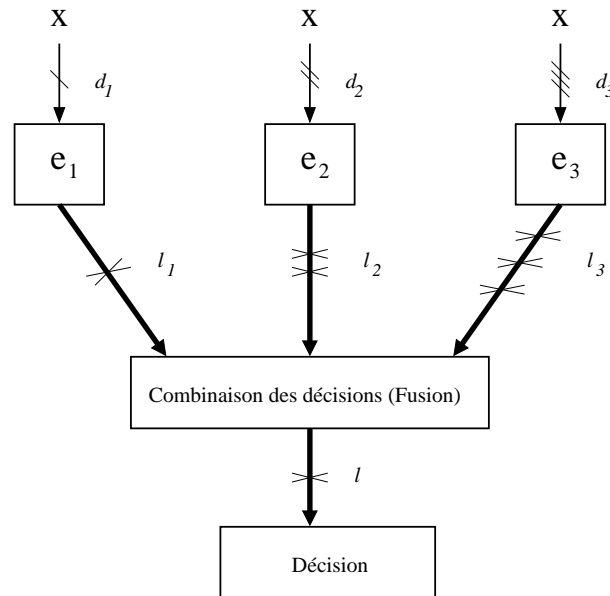


FIG. 1.5 – Système de classifieurs effectuant la combinaison de classifieurs pour un problème à l classes et d caractéristiques disponibles.

Les résultats théoriques et expérimentaux relevés dans la littérature montrent clairement que la combinaison de classifieurs ne peut être efficace que si les classifieurs individuels sont "performants" et "divers" (Sharkey [94]; Tumer et Ghosh [103]), c’est-à-dire, s’ils présentent des taux d’erreur faibles, i.e., des taux de reconnaissance largement supérieurs à 50 %, et s’ils font des erreurs différentes. Cette dernière condition signifie que

les décisions ou les sorties des classifieurs sont décorrélés (chaque classifieur se spécialise sur un certain nombre de classes). Elle peut être satisfaite en utilisant des espaces de caractéristiques différents, des ensembles d'apprentissage différents ou bien des classifieurs de construction intrinsèquement différents (paramétrages différents ou types de classifieurs différents). Une approche mixte, combinant ces différents moyens est également possible. Cependant, fondamentalement, trois types de combinaisons de classifieurs sont possibles. Ces trois types proviennent de la spécialisation des classifieurs suivant des ;

- caractéristiques différentes,
- méthodes différentes,
- ou des ensembles d'apprentissage différents.

Les trois techniques correspondent à trois types de spécialisations différentes des classifieurs. En effet, en utilisant différentes caractéristiques, on spécialise les classifieurs dans l'espace des caractéristiques. En utilisant différentes méthodes (et/ou paramétrages), on spécialise les classifieurs dans l'espace des décisions. En effet, privilégier une méthode de classement par rapport à une autre ou bien un paramétrage ou un autre, revient à privilégier une décision par rapport à une autre ou certaines décisions par rapport à d'autres. Enfin, en utilisant différents ensembles d'apprentissage, on spécialise les classifieurs dans l'espace de représentation. La majorité des travaux déjà réalisés concernent les spécialisations du premier type. Pourtant, il y a autant d'intérêt à exploiter les deux autres types.

Les techniques de *Bagging* et *Boosting*, utilisées notamment par Quinlan [87] et par Ho [55], spécialisent implicitement les différents classifieurs produits dans l'espace de représentation. La technique de *Bagging* redéfinit l'ensemble d'apprentissage pour chaque classifieur, par sélection, indépendante et aléatoire (distribution de probabilité uniforme), des individus d'un ensemble d'apprentissage disponible, alors que celle de *Boosting*, affecte aux vecteurs de l'ensemble d'apprentissage différents poids pour les différents classifieurs. Dans la dernière technique, des poids sont également affectés aux classifieurs, lors de la phase de combinaison des décisions.

Kuncheva et al. [68] ont montré que l'amélioration est aussi possible dans le cas où les classifieurs sont dépendants (i.e., dont les décisions sont corrélées), mais à condition que les erreurs commises par les classifieurs concernent des individus très différents (les auteurs appellent ceci la "dépendance négative").

Dans ce type de système de classifieurs, tous les classifieurs sont appliqués en parallèle et leurs sorties sont combinées pour atteindre un consensus ; ces approches ont pour cadre la théorie du vote (Lam et Suen [69] ; Xu et al. [112]), le consensus unanime (Xu et al. [112] ; Ho et al. [53]), les méthodes de vote utilisant des règles de décision heuristiques (Kimura et Shridar [63]), la théorie Bayésienne (Xu et al. [112]), les théories de l'incertain (Shafer [93] ; Xu et al. [112]), et la théorie des possibilités (Dubois et Prade [26]). Le choix du type de combinaison dépend, en particulier, du type des sorties des classifieurs.

Pour les sorties de classifieurs de type 3 (cf. page 29) et dans le cadre de la théorie Bayésienne de la décision, Kittler et al. [64] ont récemment montré qu'en émettant certaines hypothèses et en utilisant différentes approximations, on peut déduire la règle disjonctive(*somme*) de la règle conjonctive(*produit*). Ensuite, ils déduisent l'ensemble des autres règles (minimum, maximum, médiane et vote majoritaire) à partir de ces deux règles de base, mettant ainsi en évidence que ceux-ci sont des cas particuliers de ces deux règles fondamentales. Les différences et les similarités entre ces deux règles, en théorie et en pratique, ont été étudiées notamment par Tax et al. [100]. Pour les sorties de type 1 et 2 (cf. page 29), les deux règles fondamentales prennent respectivement les formes d'intersections et d'unions. Du fait que les sorties de type 3 (niveau "distance") soient les plus riches en information, nous nous intéressons plus particulièrement à la fusion pour ce type de sorties.

1.5.1 Combinaison usuelles

Pour généraliser, définissons $v_{ik}(x)$ comme étant une valeur numérique calculée par le classifieur k pour la classe i lors du classement d'un individu x . Ces valeurs peuvent être par exemple des probabilités ou des degrés d'appartenance. Soient l le nombre de classes possibles et K le nombre de classifieurs dont les sorties entrent en combinaison. Soient $l + 1$ l'indice représentant la classe de rejet et $0 < T \leq 1$, un seuil prédéfini de rejet. Les règles de décision les plus couramment employées sont :

- La règle Maximum :

$$SC(x) = \begin{cases} j, & \text{si } v_{jk}(x) = \max_{i=1}^l (\max_{k=1}^K v_{ik}(x)) \geq T \\ l + 1, & \text{sinon .} \end{cases} \quad (1.7)$$

- La règle Minimum :

$$SC(x) = \begin{cases} j, & \text{si } v_{jk}(x) = \max_{i=1}^l (\min_{k=1}^K v_{ik}(x)) \geq T \\ l + 1, & \text{sinon .} \end{cases} \quad (1.8)$$

- La règle Somme :

$$SC(x) = \begin{cases} j, & \text{si } \sum_{k=1}^K v_{jk}(x) = \max_{i=1}^l (\sum_{k=1}^K v_{ik}(x)) \geq T \\ l + 1, & \text{sinon .} \end{cases} \quad (1.9)$$

- La règle Moyenne :

$$SC(x) = \begin{cases} j, & \text{si } \frac{1}{K} \sum_{k=1}^K v_{jk}(x) = \max_{i=1}^l (\frac{1}{K} \sum_{k=1}^K v_{ik}(x)) \geq T \\ l + 1, & \text{sinon .} \end{cases} \quad (1.10)$$

- La règle Médiane :

$$SC(x) = \begin{cases} j, & \text{si } v_{jk}(x) = \max_{i=1}^l (\text{mediane}_{k=1}^K v_{ik}(x)) \geq T \\ l + 1, & \text{sinon .} \end{cases} \quad (1.11)$$

Dans les sections suivantes, nous allons présenter la règle de vote majoritaire et la règle *produit* dans le cadre de la théorie Bayésienne de la décision puis dans le cadre de la théorie de l'Evidence.

1.5.2 Combinaison par vote

Supposons que chaque classifieur $e_i \in E$ effectue un classement "net" affectant chaque vecteur de caractéristiques (ou individu) à l'une des l classes. Une méthode simple de combinaison des sorties de ces classifieurs est d'interpréter chaque sortie en tant que "vote" pour l'une des l classes. La classe qui a obtenu un nombre de votes supérieur à un seuil préfixé est retenue en tant que décision finale. Si ce seuil est choisi comme étant la moitié du nombre de classifieurs votants, la règle est appelée combinaison à vote majoritaire.

L'avantage de ce type de combinaison est qu'il peut être utilisé pour tout type de classifieur, quelque soit le type de sorties de ces classifieurs (cf. section 1.2.3). En effet, lorsque ces sorties sont du niveau "classement" ou du niveau "distance", on peut toujours les transformer en niveau "abstrait" (avec perte de richesse d'informations). La réciproque n'est pas vraie, car le niveau le plus riche en informations est celui de "distance".

1.5.3 Combinaison par la Théorie Bayésienne de la décision

Soient S le nombre de classifieurs et C_i ($i \in [1, l]$) une classe. Les probabilités *a posteriori*, pour les différentes classes C_i , sont déterminées de la manière suivante :

$$P(C_i/m_1, m_2, \dots, m_S) = \frac{P(C_i) \prod_{z=1}^S p(m_z/C_i)}{\sum_{j=1}^l P(C_j) \prod_{z=1}^S p(m_z/C_j)} \quad (1.12)$$

où m_z représente le vecteur de forme utilisé par le classifieur e_z ($z \in \{1, \dots, S\}$), les $P(C_i)$ et les $P(C_j)$ sont les probabilités *a priori* d'apparition des l classes et les $p(m_z/C_i)$ sont les probabilités conditionnelles. S'il s'agit de combiner des classifieurs Bayésiens, ces probabilités sont définies dans le chapitre 2, page 61.

La détermination de ces probabilités *a priori* n'est pas aisée. Dans certains travaux , elles sont considérées équiprobables et dans d'autres, elles sont déterminées en fonction des effectifs des classes (i.e., rapport de l'effectif d'une classe sur le nombre total d'individus pour toutes les classes).

Ensuite, on peut affecter un individu x à une classe suivant la règle de décision de Bayes ;

$$SC(x) = \begin{cases} j, & \text{si } P(C_j/m_1, m_2, \dots, m_S) = \max_{i \in [1, l]} (P(C_i/m_1, m_2, \dots, m_S)) \geq T \\ l + 1, & \text{sinon .} \end{cases} \quad (1.13)$$

où $0 < T \leq 1$ est un seuil prédéfini de rejet.

1.5.4 Combinaison par la Théorie de l'Evidence

Dans certains cas, l'utilisation de cette règle de Bayes peut être inadaptée à la combinaison de décisions de plusieurs classifieurs. En effet, la règle de Bayes nécessite que les mesures faites se comportent comme des probabilités. Dans le cas des décisions d'experts,

cette condition est souvent impossible à respecter (Bahler et Navarro [6]), car un expert peut très bien vouloir accorder un certain crédit à deux classes (hypothèse mixte), ce que la théorie des probabilités ne permet pas d'exprimer simplement. La théorie de l'Evidence, appelée aussi théorie des Croyances (Shafer [93]) est un des outils de raisonnement les plus connus pour la représentation de l'incertitude et de l'imprécision dans les systèmes à base de connaissances. Par rapport aux approches statistiques et floues, la théorie de l'Evidence présente l'avantage d'inclure une modélisation à la fois de l'incertitude et l'imprécision lors de la combinaison de classifieurs et lors de l'affectation d'un individu à une classe. Des masses sont affectées à tous les éléments de 2^Ω , plutôt qu'aux éléments de Ω seulement (Ω étant l'espace de décision). La fusion Bayésienne se retrouve comme cas particulier de la théorie de l'Evidence. Les exemples de cas où la théorie de l'Evidence peut être mise à profit sont :

- lorsque la décision d'un classifieur est ambiguë, la théorie prend en compte cela en affectant des masses à la réunion de plusieurs classes,
- lorsque les classifieurs ont des fiabilités différentes ; il est possible de les prendre en compte en affaiblissant ou renforçant leurs masses par une pondération,
- lorsque la connaissance sur la fiabilité des classifieurs dépend des classes, ceci peut être pris en compte en modifiant les masses en fonction de cette connaissance.

Soient Ω , un ensemble fini de l hypothèses mutuellement exclusives $\omega_i, i \in \{1, \dots, l\}$ et 2^Ω , l'ensemble de puissance de Ω . Soit \mathcal{A} , un sous-ensemble de Ω et donc élément de 2^Ω ($\mathcal{A} \subseteq \Omega, \mathcal{A} \in 2^\Omega$). Le concept fondamental représentant l'incertitude est une distribution de croyances par une fonction appelée fonction d'allocation de masse, définie comme une application m , de l'ensemble de puissance 2^Ω dans l'intervalle $[0,1]$, vérifiant :

$$\begin{aligned} m(\mathcal{A}) &\geq 0 \text{ ssi } \mathcal{A} \in 2^\Omega \\ \text{et } \sum_{\mathcal{A} \subseteq \Omega} m(\mathcal{A}) &= 1. \end{aligned} \tag{1.14}$$

Tout sous-ensemble de Ω tel que $m(\mathcal{A}) > 0$ est appelé élément focal de m . La quantité $m(\mathcal{A})$ peut être interprétée comme une mesure de la croyance affectée exactement à \mathcal{A} et à aucun de ses sous-ensembles, en l'absence de tout autre information. Dans cette théorie, l'incertitude liée à un événement $\mathcal{A} \in 2^\Omega$ est mesurée à l'aide des fonctions de Croyance (*Bel*) et de Plausibilité (*Pl*).

$Bel(\mathcal{A})$ est interprétée comme une croyance totale affectée à \mathcal{A} :

$$\begin{aligned} Bel &: 2^\Omega \longrightarrow [0, 1] \\ \mathcal{A} &\longmapsto Bel(\mathcal{A}) = \sum_{\mathcal{B} \subseteq \mathcal{A}} m(\mathcal{B}). \end{aligned} \tag{1.15}$$

$Pl(\mathcal{A})$ est interprétée comme une croyance affectée à \mathcal{A} . En fait $1 - Pl(\mathcal{A})$ est une mesure de doute en \mathcal{A} . Elle est définie de la manière suivante :

$$\begin{aligned} Pl &: 2^\Omega \longrightarrow [0, 1] \\ \mathcal{A} &\longmapsto Pl(\mathcal{A}) = \sum_{\mathcal{B} \cap \mathcal{A} = \emptyset} m(\mathcal{B}) = 1 - Bel(\overline{\mathcal{A}}). \end{aligned} \quad (1.16)$$

L'opération la plus importante concernant les distributions de croyances est la somme (ou règle) orthogonale de Dempster [23]. Deux corps d'évidence (distributions de Croyances) m_1 et m_2 sont dits combinables si il y a au moins deux sous-ensembles non-disjoints \mathcal{B} et \mathcal{C} tel que $m_1(\mathcal{B}) > 0$ et $m_2(\mathcal{C}) > 0$. La somme orthogonale de m_1 et m_2 , notée $m_{1,2} = m_1 \oplus m_2$ est alors définie par :

$$m_{1,2}(\mathcal{A}) = \frac{\sum_{\mathcal{B} \cap \mathcal{C} = \mathcal{A}} m_1(\mathcal{B})m_2(\mathcal{C})}{1 - \sum_{\mathcal{B} \cap \mathcal{C} = \emptyset} m_1(\mathcal{B})m_2(\mathcal{C})} \quad (1.17)$$

Pour S sources, du fait que la règle de Dempster est associative et commutative, la combinaison se fait par combinaison successives (dans un ordre quelconque) et permet d'obtenir le corps d'évidence final m qui est noté :

$$m = \bigoplus_{i=1}^S m_i. \quad (1.18)$$

Finalement, une règle de décision doit être choisie et appliquée afin de retenir l'hypothèse la plus vraisemblable. La fonction de décision la plus couramment employée est celle de maximum de Crédibilité (ou de Croyance). Lorsque les hypothèses composées ne sont pas traitées et que $m(\theta) = 0$ (on se retrouve alors dans un cas similaire à la fusion Bayésienne), elle devient :

$$SC(x_k) = \begin{cases} j, & \text{si } Bel(C_j) = \max_{i \in [0,l]} (Bel(C_i)) \geq T \\ l + 1, & \text{sinon.} \end{cases} \quad (1.19)$$

où $0 < T \leq 1$ est un seuil prédéfini de rejet .

Des études plus détaillées, sur le plan théorique, sont celles de Smets [97] et de Bloch [10]. Sur le plan applicatif, une étude appliquée à l'interprétation automatique d'images satellites a été faite par Zahzah [115]. Franke et Mandler [32] notent que la combinaison de probabilités par la règle de Dempster se réduit à une combinaison par la théorie Bayésienne de la décision. La règle de Dempster peut alors être vue comme une généralisation.

Dans le cas des combinaisons de classifieurs Bayésiens (cf. chapitre suivant), les probabilités *a posteriori* ne peuvent être combinées par la règle de Dempster que si les différentes classes sont équiprobables (ou proches de l'équiprobabilité), c'est-à-dire,

$$P(C_1) = P(C_2) = \dots = P(C_l) = 1/l. \quad (1.20)$$

Sinon, les probabilités *a priori* seraient combinées S fois (S étant le nombre de classifieurs concernés par la fusion), chaque classifieur ayant tenu compte des probabilités *a priori* une fois. Il est possible d'utiliser les règles de combinaison proposées par Appriou [4] qui permettent de s'affranchir de cette limitation.

1.5.5 Autres cadres théoriques de la combinaison

D'autres cadres théoriques permettent également d'effectuer des combinaisons, souvent avec des résultats équivalents. Tout comme il est difficile de prouver que tel classifieur est meilleur que tel autre pour tout problème de RdF, il n'est pas aisé non plus de prouver que tel cadre théorique convient mieux que tel autre pour tout problème de RdF. En particulier, la théorie des ensembles flous et la théorie des possibilités constituent des cadres théoriques performants. Verikas et al. [109], entre autres, ont effectué des combinaisons par l'intermédiaire des intégrales floues de Sugeno et de Choquet.

1.6 Coopération de classifieurs

Ces méthodes utilisent les décisions et/ou les résultats d'un classifieur pour mieux guider un ou plusieurs autres classifieurs dans leur prise de décision. Le plus souvent, c'est le vecteur de décision d'un classifieur qui est transmis à l'entrée d'un autre classifieur (notamment pour Huang et Suen [57][58] ; Happel et Bock [48]), car ce type d'informations est celui qui comporte le moins de différences (dans sa forme) d'un type de classifieur à un autre (ce qui conduit à une méthode plus générale). Cet avantage permet de faire coopérer des classifieurs de différents types plus facilement. Cependant, d'autres types d'informations peuvent également être transmis ou reçus (exemple : un classifieur pourrait communiquer les classes rejetées).

La coopération peut engendrer soit une réduction de l'espace de décision (travaux de Cao et al. [13]), soit une réduction de l'espace de représentation (voir l'algorithme 3C, au chapitre 3), soit une réduction de l'espace des caractéristiques (travaux de Jaimes et Chang [60] et de Loonis et Ménard [73]). Ainsi, les classifieurs entrant en coopération peuvent être spécialisés de trois manières différentes :

- dans l'espace des caractéristiques,
- dans l'espace de décision,
- et/ou dans l'espace de représentation.

La figure 1.6 illustre les interactions possibles entre les différents classifieurs d'un système mettant en œuvre la coopération. Dans cette configuration, les vecteurs de caractéristiques x_1 , x_2 et x_3 concernent un même individu à classer, mais ils peuvent être différents les uns des autres. Les classifieurs e_1 , e_2 et e_3 n'ont pas nécessairement un espace de discernement commun (c'est le cas, notamment, des classifieurs hiérarchiques).

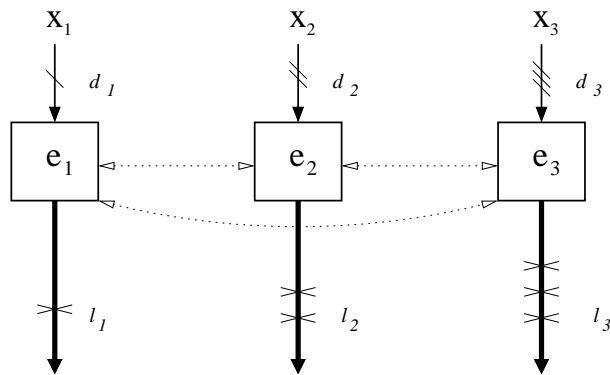


FIG. 1.6 – Système de classifieurs effectuant la coopération de classifieurs.

Un cas particulier de système de classifieurs en coopération est celui des classifieurs multi-niveaux (ou hiérarchiques) mis en œuvre essentiellement par la communauté des réseaux connexionnistes. Ce type de système est un cas particulier de la coopération, puisque à partir de la figure 1.6, lorsqu'on élimine la liaison entre e_2 et e_3 et qu'on rend unidirectionnelles les liaisons entre e_1 et e_2 d'une part et entre e_1 et e_3 d'autre part (dirigés vers la droite), on obtient un système de classifieurs hiérarchiques où e_1 se trouve au niveau supérieur et e_2 et e_3 se trouvent au niveau inférieur. Généralement, ils permettent de réduire l'espace de décision, un classifieur du niveau supérieur désignant un classifieur du niveau inférieur apte à classer l'individu présenté en entrée du système. Le système proposé par Cao et al. [13] est de ce type. Il permet également d'effectuer une sélection adaptative (voir ce terme plus loin, dans la section sélection de classifieurs).

Le lecteur peut trouver dans le tableau de la figure 1.8 les références de quelques travaux en fonction du type de système de classifieurs et en fonction du type de spécialisations des classifieurs impliqués.

1.7 Sélection de classifieurs

Dans la communauté *Apprentissage*, la plupart des études actuelles concernent la comparaison des différents algorithmes (Jaimés et Chang [59]). Le critère de qualification le plus souvent adopté est celui de la performance de ces algorithmes sur des ensembles de données standards. Comme Salzberg l'a précisé [91], puisque les performances des classifieurs dépendent fortement du domaine spécifique dans lequel ils sont appliqués, de telles comparaisons peuvent conduire à des erreurs. C'est la raison pour laquelle les chercheurs de cette communauté préconisent l'évaluation empirique des algorithmes dans leur domaine d'application. Ceci conduit tout naturellement à la sélection statique de classifieurs. Or, dans le cadre des systèmes de classifieurs, le meilleur classifieur (ou le meilleur sous-ensemble de classifieurs) peut être différent suivant l'individu à classer. On parle alors de sélection adaptative et de sélection dynamique.

Selon Giacinto et al. [37], les approches utilisées par la communauté des réseaux connexionnistes peuvent être classées en deux stratégies de conception : la stratégie "directe" et la stratégie "surproduire et choisir" (en se référant à des classifieurs). La première stratégie consiste à générer des réseaux de neurones décorrélés directement (dont les sorties peuvent alors être combinées), tandis que la deuxième consiste à créer un ensemble initial conséquent de réseaux et de choisir le sous-ensembles des réseaux les plus décorrélés. Cette seconde stratégie est en fait une stratégie de sélection *statique* de classifieurs.

Dans le cadre de la sélection adaptative, plusieurs articles présentent des méthodes de sélection d'un classifieur utilisant une estimation des performances locales dans de petites régions de l'espace de représentation entourant un échantillon de test inconnu (Woods et al. [111]). Cependant, c'est uniquement la sortie du meilleur classifieur pour cette donnée qui est utilisée pour la prise de décision finale. Ces algorithmes sont souvent fondés sur des méthodes d'optimisation, comme par exemple, les algorithmes génétiques et les réseaux de neurones.

Certains auteurs considèrent que la sélection adaptative et la sélection dynamique ont la même signification (Ho et al. [53]; Giacinto et Roli [33]). Nous faisons la distinction entre ces deux notions. S'il est vrai que, dans le cas de la sélection adaptative, différents classifieurs ou sous-ensembles de classifieurs sont sélectionnés en fonction de l'individu présenté en entrée du système, plusieurs individus identiques mais issus d'instantants différents conduisent à la sélection du, ou des même(s), classifieurs. L'aspect temporel, c'est-à-dire le fait qu'un individu soit associé à un instant t donné, n'est donc pas pris en compte. Une sélection dynamique doit, au contraire, conduire à des choix différents suivant l'instant considéré. Ceci est nécessaire lorsque les classes sont évolutives (par exemple, lorsque les centres de classes se déplacent au cours du temps). D'où la taxonomie que nous proposons ci-dessous :

Sélection statique : Pour tout individu x_k , à tout instant t , le classifieur ou le sous-ensemble de classifieurs retenu par le système de classifieurs est celui défini dans la phase d'apprentissage. La stratégie "surproduire et choisir" de la communauté des réseaux connexionnistes utilise ce type de sélection. Giacinto et al. [36] ainsi que Sharkey et al. [95] ont proposé ce type de systèmes de classifieurs.

Sélection adaptative : Pour tout individu x_k , à tout instant t , le système de classifieurs retient les sorties d'un ou plusieurs classifieurs adaptés aux valeurs numériques des caractéristiques de l'individu à traiter. L'algorithme 3C que nous proposons au chapitre 3 est un exemple de ce type de sélection. Woods et al. [111] et Cao et al. [13] ont également proposé ce type de systèmes.

Sélection dynamique : Pour un individu x_k donné, le système de classifieurs retient les sorties d'un ou plusieurs classifieurs adaptés à l'individu (i.e., aux valeurs numériques des caractéristiques de l'individu) à traiter, à l'instant considéré. Le comportement du système peut être périodique : pour deux individus x_{t_1} et x_{t_2} pris respectivement aux instants t_1 et $t_2 = T + t_1$, où T est la périodicité du comportement du processus, le système

doit sélectionner, dans les deux cas, le même classifieur ou le même sous-ensemble de classifieurs. A notre connaissance, la méthode que nous proposons au chapitre 4 constitue le seul exemple vérifiant cette définition.

La figure 1.7 illustre les interactions possibles entre les différents classifieurs d'un système mettant en œuvre la sélection. Dans cette configuration, les vecteurs de caractéristiques X_1 , X_2 et X_3 concernent un même individu à classer, mais ils peuvent être différents les uns des autres. Les classifieurs e_1 , e_2 et e_3 n'ont pas nécessairement un espace de discernement commun. Par contre, si on veut retenir les résultats de plusieurs classifieurs, il faut respecter les contraintes de la combinaison (cf. section 1.3.3). Ainsi, on peut associer à cette sélection, une combinaison et/ou une coopération de classifieurs. On a alors une approche mixte ou hybride de systèmes de classifieurs.

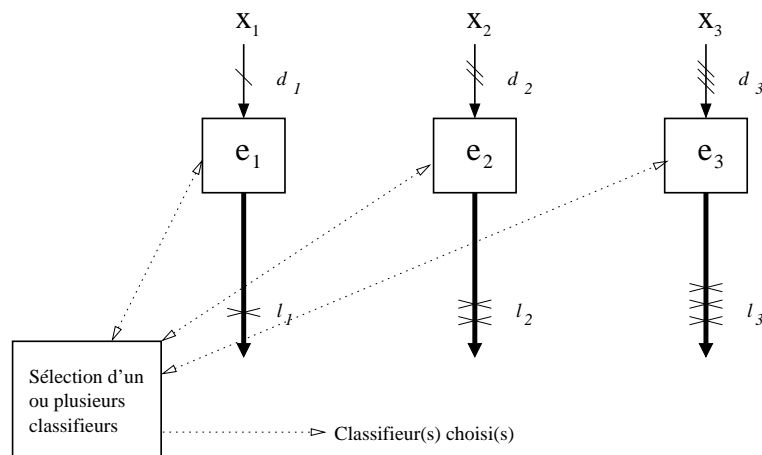


FIG. 1.7 – Système de classifieurs effectuant la sélection de classifieurs.

1.8 Système de classifieurs hybride

Certains systèmes de classifieurs effectuent à la fois des opérations de type combinaison, coopération et/ou sélection (statique, adaptative ou dynamique) de classifieurs. Le schéma de la figure 1.3 (section *Définition d'un système de classifieurs*) résume les différents types d'associations dans un système de classifieurs.

Le système proposé par Jaimes et Chang [60] est de ce type. Il permet la combinaison et la coopération de classifieurs. La coopération se fait entre des classifieurs hiérarchiques, i.e., les décisions des classifieurs de bas niveaux servent d'entrée à des classifieurs de plus haut niveau qui prennent des décisions de plus haut niveau.

Nous proposons, dans le chapitre 3, un système qui effectue les trois types d'opérations (combinaison, coopération et sélection adaptative). Le tableau de la figure 1.8 résume les

travaux existants et toutes les possibilités offertes dans le cadre des systèmes de classifieurs. Les systèmes hybrides y sont facilement identifiables par le fait qu'on les trouve ou non dans plusieurs colonnes (dans ce sens, les trois colonnes de *sélection* n'en font qu'une). Le tableau 1.2 cite les différents types d'opérations effectuées dans les systèmes les plus récemment proposés dans la littérature.

| Auteurs | Types d'opérations effectuées | | | | |
|---------------------|-------------------------------|-------------|-----------|------------|-----------|
| | Combinaison | Coopération | Sélection | | |
| | | | Statique | Adaptative | Dynamique |
| Sharkey et al.[95] | oui | | oui | | |
| Giacinto et al.[36] | oui | | oui | | |
| Giacinto et al.[34] | oui | oui | possible | | |
| Cao et al.[13] | possible | oui | possible | oui | |
| Woods et al.[111] | oui | | possible | oui | |
| Gunes et al.[46] | oui | oui | possible | oui | |
| Gunes et al.[42] | oui | possible | possible | | oui |

TAB. 1.2 – Comparatif de quelques travaux récents.

| Spécialisation des classifieurs | | Combinaison de classifieurs | Coopération de classifieurs | Sélection | | |
|---------------------------------|--|--|--|---|----------------------|------------------|
| ↓ Espaces ↓ | ↓ Méthodes ↓ | | | Statique | Adaptative | Dynamique |
| Caractéristique | Heuristiques et diverses | Van Breukelen et al.[108] Chen et al.[15] Xu et al.[112] Ho [54] Loonis et Ménard [73] | Jaimes et Chang[60] | Sharkey et al.[95] | Gosselin [39] | |
| Décision | Différents principes | Giacinto et al.[36] Ho et al.[52] | Giacinto et Roli[34] | Giacinto et al.[36] | Woods et al.[111] | |
| | Différents paramètres ou initialisations | Ng et Singh [80] Rogova [88] Cho et Kim [17] | Giacinto et Roli[34] Cao et al.[13] | Giacinto et al.[36] Sharkey et al.[95] | Cao et al.[13] | |
| Représentation | Sélection aléatoire | Wolpert [110] Quinlan [87] Ho [55] | | | Giacinto et Roli[33] | |
| | Catégorisation (clustering) | Gunes et al.[46] | Gunes et al.[46] | | Gunes et al.[46] | |
| | Sélection temporelle | Gunes et al.[42] | | | | Gunes et al.[42] |

FIG. 1.8 – Exemples de travaux concernant les systèmes de classifieurs. A notre connaissance, certaines solutions n’ont pas encore été traitées (cases vides).

1.9 Conclusion

On peut remarquer que dans la littérature, la combinaison, la coopération et la sélection (statique ou adaptative) ont déjà été utilisées dans les *Systèmes de Classifieurs* pour reconnaître des classes statiques. Lorsque les classes ont des formes complexes, ces opérations permettent, généralement, d'obtenir de meilleurs taux de reconnaissance, par rapport à un seul classifieur. La sélection adaptative a parfois été confondue avec la sélection dynamique de classifieurs. Ceci est dû au fait que l'aspect temporel a été négligé, alors qu'il devrait être pris en compte explicitement.

D'autre part, il apparaît clairement que pour reconnaître des classes dynamiques, il est nécessaire de faire intervenir, en plus de ces opérations (utilisés dans le cas statique), la sélection dynamique de classifieurs. Par conséquent, il faut non seulement combiner, faire coopérer et sélectionner les classifieurs en fonction de la situation (de l'individu à classer, des performances des classifieurs individuels, etc.) mais aussi en fonction du temps. La reconnaissance statique n'est souvent plus suffisante et il convient de développer des *Systèmes de Classifieurs* capables d'anticiper l'évolution des classes. Ainsi, c'est la capacité d'un système à effectuer la sélection dynamique de classifieurs qui est la plus utile pour la reconnaissance de classes dynamiques. La conception de tels systèmes reste un problème largement ouvert, en particulier pour ceux mettant en œuvre des opérations telles que la coopération et la sélection.

A travers le recensement des travaux effectués dans le domaine des classifieurs et des *Systèmes de Classifieurs*, nous souhaitons mettre en évidence deux aspects différents de la RdF ; les aspects statique et dynamique. Dans le cas de la RdF statique, le temps n'est pas pris en compte ou alors il ne l'est qu'implicitement. Dans le cas de la RdF dynamique, il devient absolument nécessaire de prendre en compte l'aspect temporel. Nous étudions et approfondissons cette problématique dans le chapitre suivant.

Chapitre 2

Aspects statiques et dynamiques de la RdF

A travers ce chapitre, nous étudions l'aspect temporel des systèmes de RdF. Après avoir spécifié les aspects statiques et dynamiques des objets à reconnaître et des classes constituées par ces objets, nous expérimentons quelques cas où les performances des classificateurs se dégradent. Ces dégradations sont dues à la dynamique des classes. Ensuite nous proposons deux approches possibles pour le traitement des classes dynamiques.

Sommaire

| | | |
|------------|---|-----------|
| 2.1 | Introduction | 55 |
| 2.2 | Objets dynamiques | 57 |
| 2.3 | Quelques modèles de classes dynamiques | 59 |
| 2.4 | Quelques expérimentations | 60 |
| 2.4.1 | Classement Bayésien | 61 |
| 2.4.2 | Génération de distributions Gaussiennes | 62 |
| 2.4.3 | Protocole expérimental | 64 |
| 2.4.4 | Dégradations dues aux déplacements des centres de classes | 65 |
| 2.4.5 | Dégradations dues aux variations des densités de classes | 65 |
| 2.5 | Reconnaissance de classes dynamiques | 67 |
| 2.5.1 | Approche 1 : Rendre statique un problème dynamique | 69 |
| 2.5.2 | Approche 2 : Rendre dynamique le système de reconnaissance | 69 |
| 2.6 | Choix d'une approche pour les évolutions périodiques | 70 |
| 2.6.1 | Critères de choix | 70 |
| 2.6.2 | Algorithme | 72 |
| 2.7 | Conclusion | 73 |

2.1 Introduction

Le terme "reconnaissance dynamique" peut sous-entendre plusieurs approches qu'il faut analyser. Le processus met-il en jeu des objets dynamiques ou des classes dynamiques ¹¹ ? Doit-on parler de reconnaissance de classes dynamiques ou de reconnaissance dynamique de classes ? Nous tentons, dans ce chapitre, de répondre à ces questions.

Les objets (au sens physique) sont dynamiques lorsqu'ils (ou leurs points représentatifs) sont en mouvement par rapport à un référentiel donné. C'est le cas d'un objet réel (un solide non déformable) en mouvement, par exemple, un missile. Cette dynamique n'implique pas nécessairement une évolution (variation) de leurs caractéristiques dans le temps. L'individu reste statique dans son espace de représentation. Les classes constituées par de tels objets sont statiques (stationnaires). Dans l'exemple des missiles, leurs caractéristiques visuelles (externes) ne dépendent que du modèle (ou classe) de missile.

Lorsque les caractéristiques d'objets appartenant à une classe subissent des évolutions, cette classe est appelée classe évolutive ou classe dynamique. Un exemple d'objets (naturels) dynamiques subissant des évolutions est donné par les poissons. Les caractéristiques des différentes espèces de poissons varient tout au long de l'année. Ainsi, ce sont les observations d'un même poisson (individu au sens de la RdF) qui varient périodiquement dans l'espace de représentation. On a alors à traiter des classes dynamiques, d'où notre préférence pour l'utilisation de l'expression *reconnaissance de classes dynamiques*. Nous sommes bien conscients qu'à ces variations peuvent s'ajouter des variations dues au vieillissement de ces individus. Pour tenir compte de ces variations, d'autres travaux doivent être entrepris. Par conséquent, nous distinguons deux types d'évolutions : les évolutions périodiques (cas des poissons) et les évolutions apériodiques (exemple : vieillissement des visages).

Ci-dessous, nous distinguerons deux types de trajectoires. La trajectoire d'un objet dans un espace physique, différent de l'espace de représentation et la trajectoire d'une classe dans l'espace de représentation des individus. En effet, la trajectoire d'un objet peut être décrite par un ensemble de mesures caractérisant celui-ci. Ces mesures peuvent être, par exemple, les différents coefficients d'un polynôme décrivant la trajectoire de l'objet. La dimension de l'espace de représentation des individus (égale au nombre de ces coefficients) est ainsi différent de l'espace physique dans lequel se déplace l'objet. Quant à la trajectoire d'une classe, elle se décrit dans l'espace de représentation des individus. Pour définir ce type de classes, nous réduisons, à tout instant, une classe d'individus à son prototype. Dans cette étude, nous prendrons le centre de la classe.

Dans ce chapitre, il est nécessaire de distinguer les notions d'*Objets* et de *Formes* (ou individus). En effet, deux aspects dynamiques différents entrent en jeu pour caractériser un objet : la dynamique de l'*Objet* dans l'espace physique et la dynamique de la *Forme*

¹¹On parle aussi de Classes Non Stationnaires (CNS), par opposition aux classes stationnaires (classes statiques).

associée à l'*Objet*, dans l'espace de représentation. Dans ce travail, nous émettons l'hypothèse que lorsque les *Formes* sont dynamiques, les classes constituées de ces *Formes* le sont également (le centre et/ou la matrice de dispersion de la classe évoluent). Ainsi, lorsque nous parlerons de classes dynamiques, nous sous-entendrons des *Formes* dynamiques. La dynamique de l'*Objet* ne fait pas partie de la problématique de la RdF, mais plutôt de la phase d'extraction des caractéristiques. Cependant, la dynamique de l'*Objet* et la dynamique de la classe peuvent servir de caractéristiques, respectivement, d'un *Objet* et d'une classe dynamique. Les différents cas d'*Objets* et de classes dynamiques sont les suivants :

- **Objets Statiques, Classes Statiques (OSCS) :**

Cette problématique est largement traitée dans la littérature. Un état de l'art récent est celui de Jain et al. [61]. Exemple d'application : tri de produits manufacturés.

- **Objets Dynamiques, Classes Statiques (ODCS) :**

Les caractéristiques des trajectoires d'objets sont prises en compte de la même manière que tout autre caractéristique (Alder [3]). L'étude se ramène alors à celle des OSCSs. Exemple d'application : identification de missiles.

- **Objets Statiques, Classes Dynamiques (OSCD) :**

On distingue les deux types de classes évolutives (ou dynamiques), celles qui subissent des évolutions apériodiques et celles dont les évolutions sont périodiques. Exemples d'applications : reconnaissance de visages (évolution apériodique due au vieillissement), tri automatique de poissons morts, par espèce (évolution périodique, en faisant abstraction du vieillissement).

- **Objets Dynamiques, Classes Dynamiques (ODCD) :**

Les caractéristiques des trajectoires d'objets peuvent devenir des caractéristiques d'objets. Elles sont traitées de manière à caractériser les objets à reconnaître, i.e., à les transformer en caractéristiques de ces objets. La problématique devient alors identique à celle des OSCDs. Exemple d'application : tri automatique de poissons par espèce dans des passes à poissons.

Ces différentes problématiques sont résumées dans la figure 2.1 et les passages d'une problématique à une autre sont montrés dans la figure 2.2. Dans la suite, nous proposerons des algorithmes permettant de traiter les OSCSs (au chapitre 3) et les OSCDs (au chapitre 4). Les principes et méthodes appliqués aux ODCSs peuvent être également appliqués aux OSCDs lorsque l'on transforme les classes dynamiques en classes statiques. Il y a alors intégration du temps dans la construction des classes, i.e., les nouvelles classes sont obtenues en intégrant les classes respectives en fonction du temps. Mais ceci conduit, parfois, à des cas difficiles à résoudre, en particulier lorsque les trajectoires des classes s'intersectent (dans l'espace de représentation, en particulier de faible dimension).

Lorsqu'il est possible de modéliser la trajectoire d'une classe dynamique par la trajectoire du prototype (ou centre) de cette classe, on définit la trajectoire d'une classe par trois composantes :

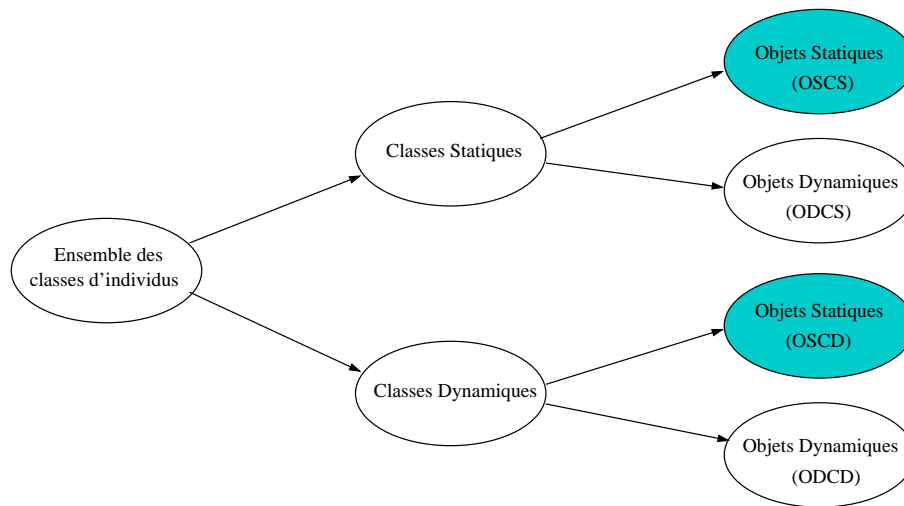


FIG. 2.1 – Les différentes situations dans le cadre de la RdF.

- la tendance, représentée par une fonction déterministe (du temps), variant lentement dans le temps,
- le facteur saisonnier (périodique) que l'on peut traduire par une fonction périodique du temps, de période donnée,
- le mouvement résiduel qui correspond aux fluctuations ponctuelles non prises en compte par les composantes précédentes.

Cette évolution temporelle, caractéristique de la classe dynamique, peut être représentée par un ensemble d'observations de la classe, effectuées à une fréquence donnée et repérées dans l'ordre croissant du temps. Le temps joue alors un rôle fondamental dans le processus de reconnaissance à mettre en œuvre.

Cependant, il est souvent difficile voire impossible de disposer de toutes ces informations permettant de définir les trois composantes définies ci-dessus. Nous proposons alors une approche dans laquelle le système de reconnaissance s'adapte à l'évolution des classes (cf. chapitre 4). Dans d'autres cas, il est possible de faire abstraction de la variable temporelle. C'est ce que nous décrivons dans la section suivante, où les objets sont dynamiques, mais les classes sont statiques.

2.2 Objets dynamiques

Un exemple où il est bénéfique d'utiliser les trajectoires d'objets pour discriminer des *Formes* est celui de la reconnaissance de missiles et d'avions. Ces trajectoires sont également des informations utiles et peuvent servir de caractéristiques permettant de distinguer ces deux classes. Ceci est d'autant plus judicieux que les systèmes radars ne permettent pas d'obtenir d'autres caractéristiques. Cette méthode, mise en œuvre notamment par Alder [3], peut aussi être utilisée dans toute application où les trajectoires d'objets dans

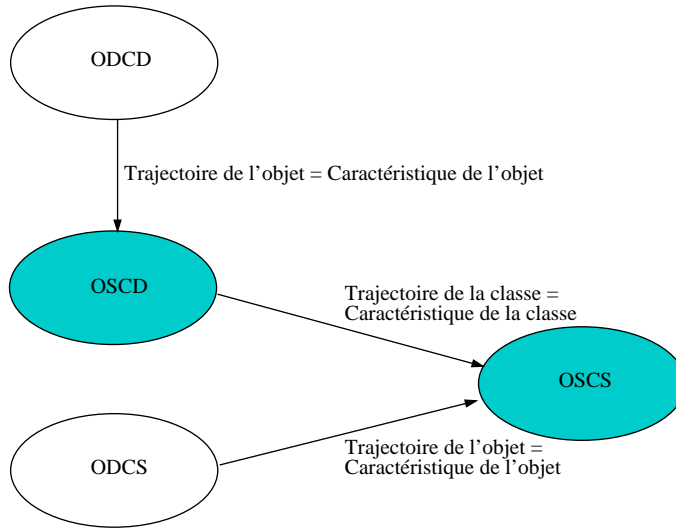


FIG. 2.2 – Passage d’une problématique de RdF à une autre.

un repère permettent l’identification de ces objets. Ces caractéristiques n’évoluent pas en fonction du temps. Ce sont les objets qui sont dynamiques. Dousson et Ghallab [25] ont proposé une méthode similaire, fondée sur une modélisation par des réseaux de Petri, pour le cas où le temps est discret et où une *Forme* peut être décrite par un ensemble d’événements successifs.

Pour illustrer notre propos, prenons un exemple d’objets en mouvement circulaire. Supposons que l’on soit amené à reconnaître un ensemble d’objets X constitué de l classes. Chaque objet peut être décrit par ses caractéristiques classiques (exemples : poids, longueur, largeur, texture .. etc) mais aussi par des caractéristiques de sa trajectoire. Dans le cas choisi, la trajectoire d’un objet peut être décrite par une application f telle que :

$$\begin{aligned}
 f &: \mathbb{R} \longrightarrow \mathbb{R}^2 \\
 t &\longmapsto \begin{pmatrix} a \cos (b t) \\ a \sin (b t) \end{pmatrix}
 \end{aligned}
 \tag{2.1}$$

où a et b sont deux constantes qui caractérisent la trajectoire de l’objet, a correspond au rayon de la trajectoire et b correspond à la vitesse angulaire de l’objet.

Si chaque classe d’objets est caractérisée par une vitesse angulaire (uniforme) et un rayon de rotation propres, la famille d’applications $G = \{g_i\}_{i \in [1, l]}$ telle que :

$$\begin{aligned}
 g_i &: \mathbb{R} \longrightarrow \mathbb{R}^2 \\
 t &\longmapsto \begin{pmatrix} a_i \cos (b_i t) \\ a_i \sin (b_i t) \end{pmatrix}
 \end{aligned}
 \tag{2.2}$$

permet de décrire les l classes d'objets. Les couples (a_i, b_i) représentent alors les centres des classes. Ils correspondent au rayon de rotation et à la vitesse angulaire de la classe i . Si ces caractéristiques n'évoluent pas elles-mêmes dans le temps alors on obtient des classes statiques. Lorsque les trajectoires ne sont pas connues, par exemple dans le cas de la reconnaissance de personnes en mouvement, il arrive qu'un individu soit occulté par un autre. Dans ce cas, l'analyse colorimétrique des images successives peut permettre le suivi automatique de personnes (Postaire et Macaire [85]).

La trajectoire d'un objet peut être aussi bien continue (exemple ci-dessus) que discrète. Dans le deuxième cas, il s'agit de traiter une séquence de points représentant l'objet. Un recensement des travaux concernant les "Classifieurs dynamiques" utilisés en reconnaissance automatique de la parole est fait par Haton [51]. Ce recensement est volontairement limité aux modèles neuronaux et hybrides (combinant l'approche connexionniste avec d'autres approches).

L'inconvénient principal de cette "transformation" d'*Objets* dynamiques en *Formes* statiques est que la reconnaissance nécessite une acquisition préalable de la trajectoire de l'objet à reconnaître sur un certain interval de temps dépendant de l'application, en vue de sa modélisation. Ceci implique que, indépendamment du temps de traitement dans la chaîne conduisant à la reconnaissance, celle-ci devient, en général, impossible à effectuer dans des conditions de temps réel. Il est alors pertinent de mettre en œuvre des méthodes de classement-prédiction (ou de classement-prédiction-vérification).

L'avantage principal en est que, si une partie de la trajectoire de l'objet (en général, le début de la trajectoire) peut suffire à la reconnaissance de l'objet, on peut prédire le reste de la trajectoire de l'objet, désormais identifié. La prédiction de la trajectoire de l'objet est très utilisée, notamment dans les applications militaires. Un autre avantage important de la méthode est qu'on peut identifier un objet sans "le voir". La connaissance de son comportement dynamique (i.e., sa trajectoire) est souvent plus discriminante que des caractéristiques physiques (chromatiques, morphologiques, etc.).

2.3 Quelques modèles de classes dynamiques

Dans les méthodes markoviennes, la Forme est considérée comme un signal continu observable à différents instants, ces observations constituant des "états d'observation". Pour chaque état, un modèle stochastique possède une variable aléatoire $D(t) = d_i$, $t \in [1, T]$ (et $i \in [1, n]$, n étant le nombre d'états possibles, considéré invariant dans le temps), qui prend ses valeurs dans l'ensemble des observations que l'on fait dans le temps (sur le phénomène analysé). Ainsi, une évolution du système (l'ensemble des classes) peut être représentée par une suite de transitions d'états à partir d'un état initial $D() = d_1$. La loi d'évolution du système consiste à obtenir les chaînes de transitions $d_1, d_2, \dots, d_m \quad \forall m \leq T$ avec la probabilité $P(d_1, d_2, \dots, d_T)$. Cette probabilité est donnée par ;

$$P(d_1, d_2, \dots, d_T) = P(d_1) \times P(d_2/d_1) \times P(d_3/d_2, d_1) \times \dots \times P(d_T/d_{T-1}, \dots, d_1) \quad (2.3)$$

Ce processus stochastique vérifie la propriété de Markov, si pour tout t :

$$P(D(t) = d/D(1) = d_1, D(2) = d_2, \dots, D(t-1) = d_i) = P(D(t) = d/D(t-1) = d_i) \quad (2.4)$$

En remplaçant $D(t)$ par D_t et en utilisant la propriété de Markov, il vient :

$$P(D_1, \dots, D_T) = P(D_1) \times P(D_2/D_1) \times P(D_3/D_2) \times \dots \times P(D_T/D_{T-1}) \quad (2.5)$$

L'évolution dépend entièrement de la probabilité initiale et des probabilités des transitions successives. Si la suite de variables aléatoires $D = (D_t)_{t \in [1, T]}$ vérifie la propriété de Markov, elle est appelée chaîne de Markov. Un réseau ou graphe markovien est un graphe représentant une chaîne de Markov. On peut associer aux arcs des probabilités de transitions (indépendantes du temps également). Ces probabilités peuvent se mettre sous la forme d'une matrice appelée matrice de transition ou matrice de passage ou encore matrice stochastique de transition. Un graphe markovien muni de sa matrice stochastique de transition donne un graphe Bayésien muni de ses probabilités conditionnelles (Fabiani [29]). Lorsque les modèles nécessaires comportent beaucoup d'états, ces méthodes deviennent, malheureusement, complexes à gérer (recherche du chemin le plus probable correspondant à la chaîne d'entrée, dans le graphe d'états).

Boutleux [12] a proposé de définir par un couple de classes les états extrêmes de l'évolution d'une classe. Il appelle ces couples de classes des *classes extrémités*. La trajectoire de cette évolution est constituée de segments de droites (nombre de segments fixe). Des degrés d'appartenance sont définis pour les deux classes extrémités. Ces degrés d'appartenance sont déterminés en fonction de la position spatiale de la classe en évolution, sans notion de dépendance du temps. La méthode prend en compte l'évolution possible du nombre de classes au niveau de la règle de décision (ce qui correspondrait à un nouveau mode de fonctionnement). L'inconvénient de cette approche est que seules les classes évoluant "d'une classe extrémité à une autre" sont traitées.

Les travaux de Cron [20] s'appuient sur des ensembles flous dépendants du temps et concernent deux types "d'états dynamiques"¹². Le premier type est la représentation d'un état dont les caractéristiques dépendent du temps. Le second concerne les changements d'états, i.e., l'évolution d'un état vers un autre. Son approche consiste à représenter les classes dynamiques par des ensembles flous dépendants du temps en apprenant l'évolution des classes grâce à une méthode de régression floue.

2.4 Quelques expérimentations

Nous nous proposons, dans cette section, d'étudier l'influence des évolutions des classes sur les performances des classifieurs. Une classe dynamique est susceptible de subir trois

¹²Les classes dynamiques sont nommées ainsi par la communauté diagnostic par RdF. Elles sont aussi appelées *classes évolution* et *classes non stationnaires* (dans le cas des évolutions apériodiques).

types d'évolutions ;

- sur la densité des individus de la classe,
- sur la forme de la classe,
- sur la position du centre de la classe (classe statistique).

Lorsque les positions des centres de classes sont suffisamment éloignées les unes des autres, les deux premiers types de variations influent peu sur les performances d'un classifieur ou d'un système de classifieurs (phase de classement). Par contre, si les positions des centres évoluent (même si les positions sont très éloignées, à l'instant initial), les performances peuvent sensiblement varier et chuter. La variation de la position est celle qui est la plus susceptible de nécessiter une modification des frontières de décisions entre classes. Les travaux de Cron [20] et de Boutleux [12] concernent ce type d'évolutions, dans le cadre du diagnostic par RdF : classes qui évoluent vers d'autres classes (qui sont, dans ce cadre, des "états dynamiques").

L'idée est de simuler, dans un cadre statistique, l'évolution de classes dynamiques par des modèles simples afin d'évaluer l'influence de l'évolution des classes sur le taux de reconnaissance. Nous présentons deux cas où ce taux se dégrade : lors de la variation des positions des centres et lors de la variation des densités de classes.

Dans cette étude, la détermination de la performance est effectuée par le calcul des taux de reconnaissance, car leur interprétation par un expert est simple. La matrice de confusion fournit directement les taux de reconnaissance associés à chaque classe. Nous nous intéressons au taux de reconnaissance global ainsi qu'aux différents taux partiels. Le classifieur utilisé est un classifieur Bayésien.

2.4.1 Classement Bayésien

Propriétés des classes

L'objectif de cette étude est l'observation des performances d'un classifieur et non celui de modéliser le mieux possible les classes réelles. Comme il est d'usage dans la majorité des travaux, nous supposons que la distribution des caractéristiques dans l'espace de représentation \mathbb{R}^2 obéit, pour chaque classe, à une loi normale (Gaussienne). Cette loi est généralement admise pour la plupart des phénomènes naturels et permet de définir une classe statistique. Soit l le nombre de classes. Dans le cas multidimensionnel, la probabilité (conditionnelle) d'avoir un vecteur de Forme $x_k \in \mathbb{R}^2$, sachant qu'elle appartient à la classe $C_i, i \in [1, l]$ s'écrit alors ;

$$p(x_k/C_i) = \frac{1}{2\pi^{\frac{d}{2}} |\Sigma_i|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x_k - \mu_i)^t \Sigma_i^{-1} (x_k - \mu_i)\right] \quad (2.6)$$

où :

- μ_i est le vecteur caractéristique moyen de dimension d , relatif à la classe C_i ,

- Σ_i , la matrice des variance-covariances de dimension $d \times d$, relative à la classe C_i .

Probabilités *a posteriori*

Ces probabilités sont déterminées à l'aide de la formule de Bayes :

$$p(C_i/x_k) = \frac{p(x_k/C_i)P(C_i)}{P(x_k)} \quad (2.7)$$

où :

- $P(C_i)$ est la probabilité d'apparition *a priori* de la classe C_i ,
- $p(C_i/x_k)$ est la probabilité *a posteriori* d'appartenance de x_k à la classe C_i ,
- $P(x_k) = \sum_{i=1}^K p(x_k/C_i)P(C_i)$ est la densité de probabilité du mélange, terme qui sert de normalisation de la distribution des probabilités sur les différentes classes.

Ce calcul est équivalent à la combinaison de deux sources¹³, les probabilités *a priori* et les probabilités conditionnelles aux classes. La première provient généralement de l'expérience d'un expert, ou de l'observation du classement nécessitant le calcul des occurrences des différentes classes divisés par le nombre total des observations. La seconde source provient de la modélisation mathématique des classes d'après les mesures effectuées sur l'ensemble des observations.

Règle de décision Bayésienne

La règle de décision Bayésienne consiste à choisir la classe pour laquelle la probabilité $p(C_i/x_k)$ est la plus élevée. Choisir la probabilité maximale revient à prendre la décision dont le coût est minimal. Soit e_b un classifieur de type Bayésien. Sa décision (avec la possibilité de rejet) est donnée par la fonction :

$$e_b(x_k) = \begin{cases} C_j, & \text{si } P(C_j/x_k) = \max_{i \in [1,l]} (P(C_i/x_k)) \geq T \\ C_{l+1}, & \text{sinon .} \end{cases} \quad (2.8)$$

où $0 < T \leq 1$ est un seuil prédéfini de rejet. La classe étiquetée C_{l+1} correspond alors à la classe de rejet.

2.4.2 Génération de distributions Gaussiennes

Nous générons les différentes classes selon des distributions normales. Pour une représentation plus facile du phénomène simulé, nous nous limitons à la dimension 2. Soit x_1 et x_2 , les deux caractéristiques à générer. La distribution Gaussienne à une dimension, d'écart-type σ et de moyenne μ est de la forme :

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right] \quad (2.9)$$

¹³Dans le chapitre 1 (page 42), nous avons présenté la formule de Bayes adaptée à la combinaison des sorties de plusieurs classifieurs.

Elle est normalisée : $\int_{-\infty}^{+\infty} p(x)dx = 1$.

Les variables aléatoires x_1 et x_2 à obtenir doivent résulter de deux fonctions Gaussiennes indépendantes.

Génération par une méthode de transformation

Pour ce type de méthode, il faut respecter la loi de transformation fondamentale des probabilités qui est, dans le cas multidimensionnel, exprimée par ;

$$p(x_1, x_2, \dots, x_n)dx_1dx_2 \dots dx_n = p(y_1, y_2, \dots, y_n) \left| \frac{\partial(y_1, y_2, \dots, y_n)}{\partial(x_1, x_2, \dots, x_n)} \right| dx_1dx_2 \dots dx_n \quad (2.10)$$

où :

- $\left| \frac{\partial(y)}{\partial(x)} \right|$ est le déterminant Jacobien des y par rapport aux x .
- y_1, y_2, \dots, y_n sont des variables aléatoires avec une distribution de probabilités jointes $p(y_1, y_2, \dots, y_n)$ uniforme.
- et x_1, x_2, \dots, x_n sont chacune des fonctions de tous les y (même nombre de x que de y).

La méthode de transformation de Box-Muller

En utilisant la méthode de transformation de Box-Muller [31] qui respecte la loi de transformation fondamentale des probabilités et pour laquelle :

$$\begin{aligned} x_1 &= \sqrt{-2 \ln y_1} \cos 2\pi y_2, \\ x_2 &= \sqrt{-2 \ln y_1} \sin 2\pi y_2, \end{aligned} \quad (2.11)$$

le déterminant Jacobien devient :

$$\frac{\partial(y_1, y_2)}{\partial(x_1, x_2)} = \left| \begin{array}{cc} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{array} \right| = - \left[\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_1^2}{2}\right) \right] \left[\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_2^2}{2}\right) \right] \quad (2.12)$$

Ceci présente l'avantage d'être mis sous la forme d'une fonction (de 2 variables) à variables séparables. Les deux fonctions représentent chacune une distribution Gaussienne à une variable définie par l'équation 2.9 avec les caractéristiques suivantes :

$$\sigma = 1 \text{ et } \mu = 0$$

Si y_1 et y_2 sont l'ordonnée et l'abscisse d'un point aléatoire (v_1, v_2) à l'intérieur du cercle unité centré sur l'origine, alors la quantité $R = v_1^2 + v_2^2$ est une variable aléatoire uniforme qui peut être utilisée pour y_1 alors que l'angle formé par (v_1, v_2) par rapport à v_1 peut servir d'angle aléatoire $2\pi y_2$. On a alors :

$$\begin{aligned} \cos 2\pi y_2 &= \frac{v_1}{\sqrt{R}}, \\ \sin 2\pi y_2 &= \frac{v_2}{\sqrt{R}}. \end{aligned} \quad (2.13)$$

Cela permet d'éviter l'appel aux fonctions Cosinus et Sinus qui nécessitent un temps de calcul plus élevé. Les deux caractéristiques p_1 et p_2 sont ainsi générées par une même fonction $g(v)$ à laquelle on applique des moyennes et des écarts types différents :

$$\begin{aligned} p_1 &= \mu_1 + \sigma_1 * x_1, \\ p_2 &= \mu_2 + \sigma_2 * x_2. \end{aligned} \tag{2.14}$$

2.4.3 Protocole expérimental

Nous présentons les résultats que nous avons obtenus à travers deux expérimentations. Dans la première, l'expérience consiste à déplacer les centres de classes dans des directions et sens donnés (voir la section suivante). Dans la seconde, deux classes (sur 3) subissent un étalement au cours du temps (les densités de ces classes varient). Dans chaque nouvelle position, les classes sont générées de la même manière que pour la position initiale d'apprentissage, en tenant compte des nouveaux centres et écart-types.

Soient τ_g , le taux de reconnaissance global et τ_j , le taux de reconnaissance (partiel) de la classe j . Pour ces expérimentations, la procédure à respecter est la suivante ;

Procédure *Evolution*

Entrée : Centres et écart-types de départ,
Centres et écart-types d'arrivée,
 n_i : nombre d'intervalles.

Sortie : τ_g pour les n_i positions,
 $\tau_j, j \in [1, l]$ pour les n_i positions.

1 : $i=0$.

2 : **répéter**

3 : Générer une population à classer (pas i).

4 : Classifier.

5 : Calculer la matrice de confusion.

6 : Calculer les mesures de performances.

7 : $i \leftarrow i + 1$

8 : **jusqu'à** ($i > n_i$)

2.4.4 Dégradations dues aux déplacements des centres de classes

Pour les deux expérimentations, lorsqu'il n'y a pas de variations des densités de classes, l'écart-type vaut $\sigma = 10$. Soient 3 nuages de points correspondants à 3 classes représentées par 2 caractéristiques X et Y . Les données sont obtenues par un générateur aléatoire de distribution normale. A l'origine, lors de l'apprentissage, les barycentres des trois classes $\{C_1, C_2, C_3\}$ sont respectivement situées à $(30; 30)$, $(130; 30)$, $(230; 30)$, avec :

- l'ensemble d'apprentissage : 1000 individus par classe,
- l'ensemble de classement (test) : 1000 individus par classe.

Ensuite, ces classes évoluent pour occuper respectivement des positions finales :

$$\{C_1, C_2, C_3\} = \{(80; 630), (80; 630), (230, 630)\}$$

L'expérience consiste à faire varier la position relative des barycentres des classes avec 50 pas (position initiale, position finale et 48 positions intermédiaires uniformément réparties). Dans la figure 2.3, ce sont les classes 1 et 2 qui finissent par se confondre. On constate, dans la figure 2.4, que le taux de reconnaissance global devient faible à partir de la moitié de la trajectoire des classes (indice $Temps = 25$) et qu'il devient complètement aléatoire ($\tau_g = 1/l$) à partir de l'indice $Temps = 35$.

2.4.5 Dégradations dues aux variations des densités de classes

Soient trois nuages de points correspondant à 3 classes représentées par 2 attributs X et Y . Les données sont obtenues par un générateur aléatoire de distribution normale. Les classes sont constituées de ;

- C_1 : définie par $\mathcal{N}((30; 30)(\sigma; \sigma))$
- C_2 : définie par $\mathcal{N}((130; 30)(\sigma; \sigma))$
- C_3 : définie par $\mathcal{N}((230; 30)(10; 10))$

L'expérience consiste à faire varier σ , l'écart-type sur X et Y des classes C_1 et C_2 : $10 < \sigma < 110$ avec 50 pas. La figure 2.5 représente une juxtaposition des 3 classes, entre les distributions à l'origine (indice $Temps = 0$) et les distributions finales (indice $Temps = 50$). Nous utilisons la population suivante ;

- ensemble d'apprentissage : 1000 individus par classe,
- ensemble de classement : 1000 individus par classe.

Les tracés de la figure 2.6 représentent l'évolution de la dégradation des taux de reconnaissance due à la modification structurelle des classes C_1 et C_2 dans l'espace de représentation. Elle montre que les taux de reconnaissance chutent très rapidement. Cette chute est justifiée par le fait que les fonctions discriminantes (induites par l'élaboration des matrices de variance-covariances Σ_i sur l'ensemble d'apprentissage) ne sont plus adaptées

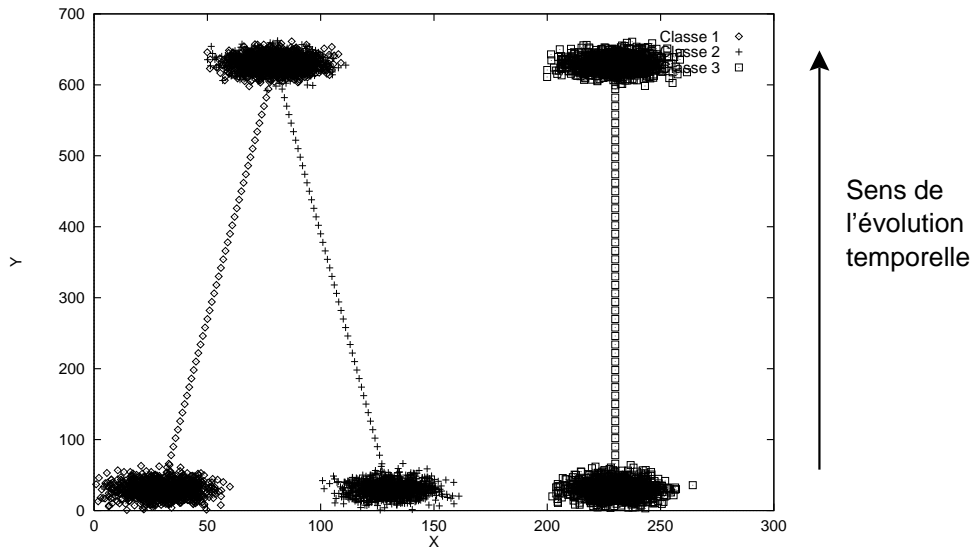


FIG. 2.3 – Distributions à l’origine ($Y = 30$), trajectoires et distributions finales des 3 classes. La durée entre deux pas successifs est $\Delta_t = 12$.

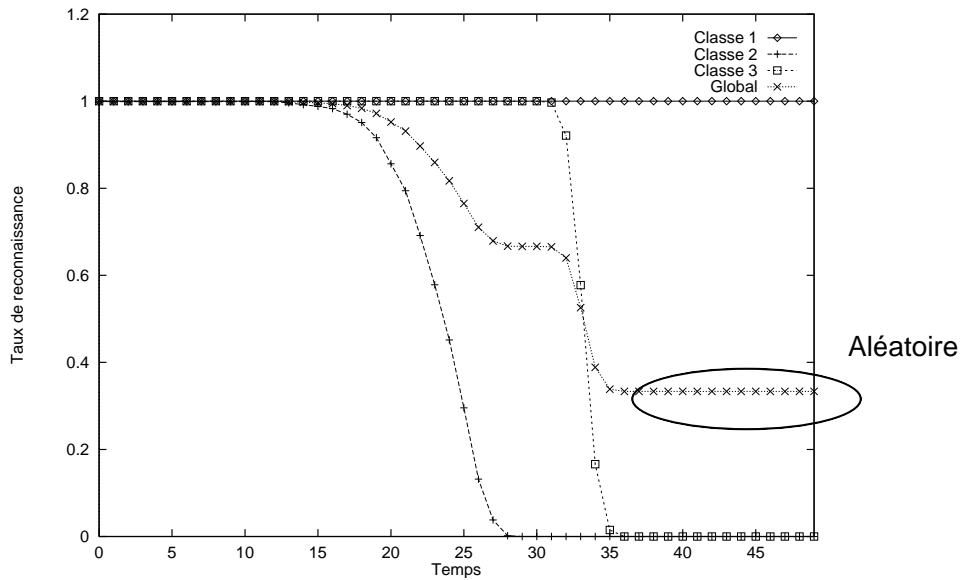


FIG. 2.4 – Evolution des taux reconnaissance pour les 3 classes d’individus. Les taux concernant les classes 2 et 3 finissent par s’annuler car elles sont confondues avec une autre classe.

aux nouveaux ensembles de classement. Dans beaucoup d'applications pratiques, ces taux seraient inacceptables dès l'indice $Temps = 15$ (le taux de reconnaissance global est alors d'environ 85 %).

2.5 Reconnaissance de classes dynamiques

Pour les objets naturels, la RdF pose des problèmes souvent plus complexes que pour les objets manufacturés. En effet, bien souvent, les caractéristiques de ces objets évoluent (i.e., les classes sont dynamiques), ce qui rend difficile la recherche de paramètres invariants par rapport à leur évolution. Les paramètres choisis pour discriminer ces objets subissent une évolution naturelle : évolution saisonnière et vieillissement, le plus souvent. Ce caractère dynamique rend les classifieurs statiques inefficaces. Par ailleurs, il nécessite une modélisation dynamique des classes, à moins d'avoir un système de reconnaissance dynamique qui s'adapte à l'évolution des classes.

Ainsi, on peut distinguer deux types d'évolutions : les évolutions périodiques (ex : variations saisonnières) et apériodiques (ex : vieillissement). Pour le deuxième type, on peut citer comme application, celle de la reconnaissance de visages humains ou bien celle de l'évolution de la fraîcheur des produits de la mer, prenant en compte l'évolution possible liée au vieillissement (travaux de Loonis et Sochacki [74]). Dans de tels cas, l'évolution des caractéristiques est rendue difficilement prédictible à cause de la modélisation complexe des classes dynamiques. Les classifieurs incrémentaux fonctionnent selon le principe d'une mise à jour de leur modèles de représentation des connaissances (exemples : centres de classes, poids dans les réseaux connexionnistes) en fonction de l'affectation de nouveaux individus classés (voir notamment les travaux de Puzenat [86]). Pour cette raison, ils semblent être les mieux adaptés. Mais cette approche est une manière de rendre le système de reconnaissance dynamique, à défaut de modéliser des classes dynamiques. De plus, cette approche ne convient que pour des classifieurs dont les frontières de décisions sont hautement non-linéaires : il faut pouvoir modifier une frontière dans une zone donnée, en influant le moins possible le reste de la frontière. Les méthodes de RdF fondées sur des réseaux connexionnistes semblent être les mieux adaptées à ce type d'approche. La méthode que nous proposons au chapitre 4 est indépendante de cette contrainte de non-linéarité.

Nous nous intéressons plus particulièrement au premier type d'évolutions, les évolutions périodiques. C'est le cas, par exemple, du tri automatique de poissons subissant des évolutions saisonnières. En fonction de la pertinence des paramètres choisis pour ce type d'évolutions, il est possible de distinguer deux catégories. En effet, lorsque ces paramètres sont mal choisis, les trajectoires de ces classes peuvent s'intersecter (croisement entre les trajectoires d'au moins deux classes, en particulier lorsque la dimension de l'espace de représentation est faible). La première catégorie d'évolutions correspond aux cas où il n'y a pas d'intersections entre les trajectoires des classes, tandis que la deuxième catégorie correspond au cas où il y a au moins une intersection. Deux approches sont possibles ; l'une consiste à rendre statique un problème dynamique (Passage du cas OSCD vers

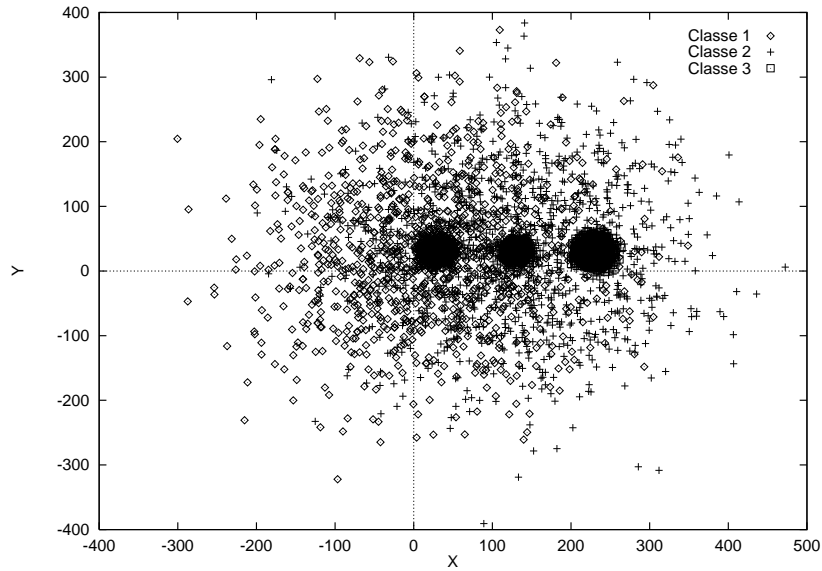


FIG. 2.5 – Distributions à l'origine (zones plus sombres) et distributions finales des 3 classes.

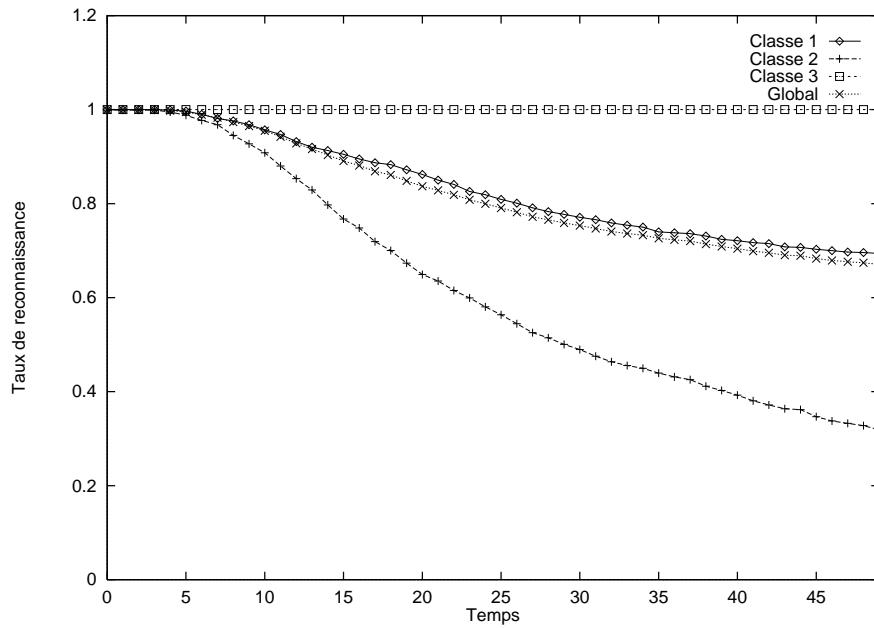


FIG. 2.6 – Evolution des taux reconnaissance pour les 3 classes d'individus.

OSCS) et l'autre consiste à rendre dynamique le système de reconnaissance lui-même. Nous discuterons du choix d'une approche dans la section suivante.

2.5.1 Approche 1 : Rendre statique un problème dynamique

C'est le cas où l'on cherche à représenter une classe dynamique, ou son évolution, par une classe statique. Certaines classes peuvent alors devenir multimodales. Cette approche est pertinente pour les données de la figure 2.8, qui représentent les individus de deux classes, intégrés dans le temps (une période). En effet, dans ce cas, la deuxième classe peut représenter une classe qui évolue périodiquement et dont nous avons des observations **sur au moins une période complète**. Le chapitre 3 présente une nouvelle méthode, adaptée à ce type de problème : l'algorithme 3C. Son application n'est pas toujours possible, par exemple dans les cas où surviennent des croisements de trajectoires ou encore dans les cas où le nombre de modes d'une classe est trop important. Cependant, un grand nombre de problèmes complexes peuvent être résolus par cet algorithme : dans le cas des classes qui se chevauchent partiellement (intersections des distributions de classes) ou encore dans le cas où les frontières entre classes sont hautement non-linéaires.

Les étapes de construction de classifieurs présentées dans le chapitre précédent doivent être modifiées. En effet, il devient nécessaire d'intégrer les caractéristiques dans le temps (sur une ou plusieurs périodes) afin de constituer les formes complètes de classes. C'est-à-dire que de manière implicite ou explicite, il y a insertion d'un opérateur d'intégration entre les fonctions *Sélection de caractéristiques* et *Choix et Construction du système de classifieurs*. Ceci est représenté dans la figure 2.7.

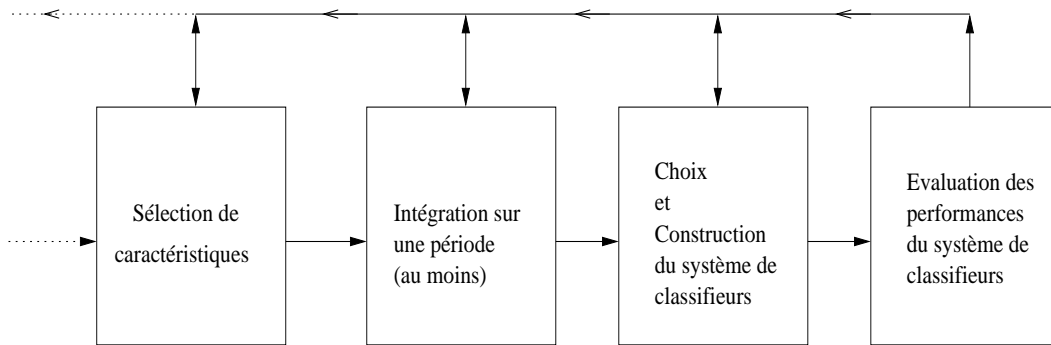


FIG. 2.7 – Les différentes étapes d'un système de classement, dans le cas où les classes évoluent périodiquement.

2.5.2 Approche 2 : Rendre dynamique le système de reconnaissance

C'est le cas où l'approche précédente conduit à obtenir des classes statiques dont les trajectoires se croisent ou dont le nombre de modes est trop important. Cette approche est pertinente pour les données de la figure 2.9. Ces données ont été générées en discrétisant

les trajectoires des deux classes en 48 positions et en générant pour chaque position 20 individus par classe, selon une distribution Gaussienne (avec un écart-type $\sigma = 2$). Dans ce cas, si on tient compte des observations des deux classes sur une période pour constituer l'ensemble d'apprentissage, les deux classes ne peuvent être clairement séparées, même si on utilise un algorithme du type 3C. Le chapitre 4 présente une nouvelle méthode de RdF, adaptée à ce type de problème : RCD-RPF. Cette approche permet au système de reconnaissance d'avoir un comportement dynamique qui s'adapte à l'évolution des classes. De plus, en adoptant une architecture à reconfiguration dynamique telle que celle proposée par Demigny et al. [21], il est possible d'optimiser l'implémentation matérielle de ce type de méthode.

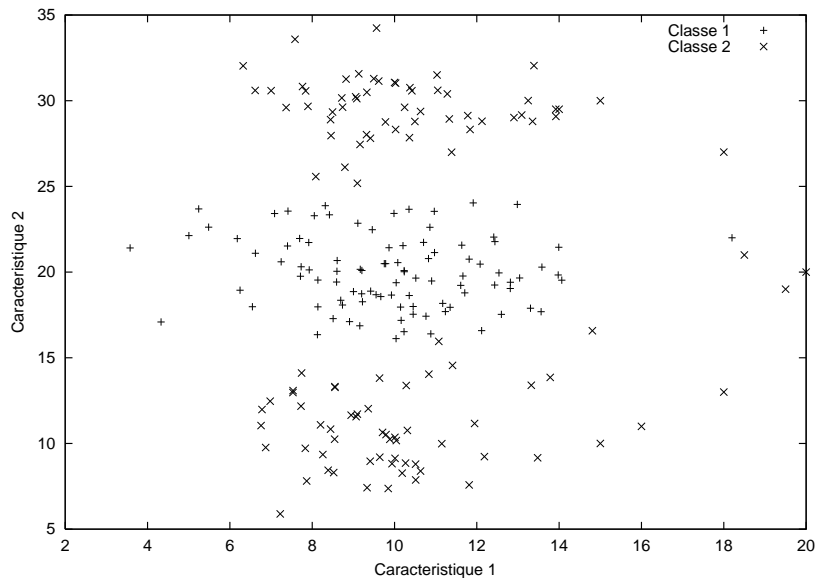


FIG. 2.8 – Exemple de problème à deux classes : pas de croisement de trajectoires, la seconde classe est bimodale.

2.6 Choix d'une approche pour les évolutions périodiques

Dans cette section, nous nous intéressons plus particulièrement aux évolutions de classes dynamiques périodiques¹⁴. Ceci nous amène à effectuer un choix entre les deux approches précédentes, selon des critères que nous devons justifier.

2.6.1 Critères de choix

Pour effectuer un choix entre les deux approches, deux critères sont à prendre en considération :

¹⁴En ce sens, les objets ont des caractéristiques cyclo-stationnaires.

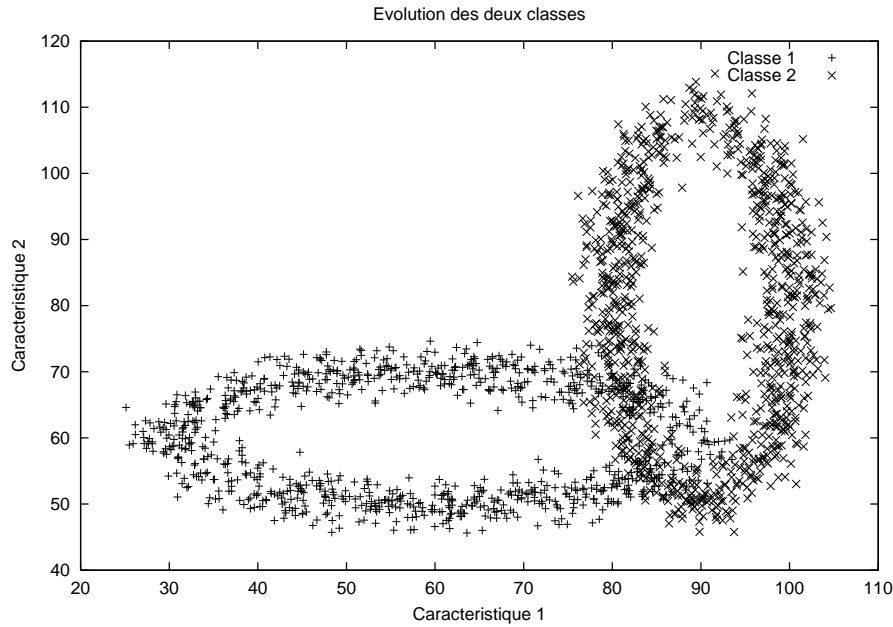


FIG. 2.9 – Exemple de problème à deux classes : il y a croisement entre les trajectoires des deux classes.

- Existe-t-il des croisements entre les trajectoires des différentes classes ¹⁵ ? Deux types de croisements peuvent exister. Si le croisement est spatio-temporel (proximité de deux centres de classes dans l'espace de représentation, à un même instant), il n'y a pas de solution évidente et il devient inutile de vérifier le second critère. Dans le cas d'un croisement uniquement spatial (proximité de deux centres de classes, pris à des instants différents de leur évolution), l'approche 2 permet de résoudre le problème. Dans les deux cas, l'approche 1 ne peut être envisagée, car l'intégration dans le temps génère des croisements dans l'espace de représentation. S'il n'existe aucun croisement, cette dernière approche peut s'appliquer.
- Les classes sont-elles multi-modales ? Ou bien, au contraire, y a-t-il une continuité dans les fréquences d'apparitions des individus le long des trajectoires ? Dans le premier cas, le nombre optimal de catégories n_m (ou classes denses) que l'on peut trouver dans l'ensemble d'apprentissage risque d'être proche du nombre de classes réelles l , car dans beaucoup d'applications pratiques, une classe comporte au plus quelques modes. L'approche 1 peut alors être envisagée. A l'inverse, dans le deuxième cas, le rapport de n_m à l risque d'être élevé (quasi-continuité de fréquences d'individus le long des trajectoires). Dans ce cas, l'approche 2 est mieux adaptée, du fait qu'elle permet d'anticiper sur les individus intermédiaires entre deux modes.

Selon le premier critère, lorsque les trajectoires sont inconnues, le choix d'une approche nécessite l'utilisation d'un algorithme de détermination des croisements éventuels entre

¹⁵En cas de croisements, les classes sont partiellement, voire totalement confondues.

les trajectoires des classes. A notre connaissance, il n'existe pas encore de méthodes pour ce type de problème. Une solution consisterait à vérifier que l'intersection entre les intervalles de variations des différentes classes (selon une même caractéristique) est bien vide et que ceci est vrai pour toutes les caractéristiques. Ceci est une condition suffisante pour qu'il n'y ait aucun type de croisement. Cependant, dans le cas où ces conditions ne sont pas vérifiées, il n'y a pas nécessairement de croisements.

Selon le second critère, le choix d'une approche nécessite l'utilisation d'un algorithme de détermination du nombre optimal de catégories (ou *clusters*) dans l'ensemble d'apprentissage de l'application considérée. La méthode possibiliste de Krishnapuram et Keller [67], ainsi que l'algorithme VGC (catégorisation guidée par la validité) de Bensaid et al. [8] sont particulièrement adaptées à la détermination du nombre optimal de catégories dans un ensemble d'apprentissage.

2.6.2 Algorithme

Soit n_{st} , le nombre de croisements spatio-temporels et n_s , le nombre de croisements spatiaux. Si $n_{st} \neq 0$, il n'y a pas de solution proposé à ce jour. Dans le cas contraire, nous proposons d'effectuer le choix entre les approches 1 et 2 en fonction de n_{st} , n_s de $\frac{n_m}{l}$. Ainsi, l'algorithme permettant d'effectuer le choix entre les deux approches est le suivant :

Procédure *ChoixApproche*

Entrée : n_m, l : entiers ; n_{st}, n_s : booléens

Sortie : *approche* : entier

1 : **si** ($n_{st} \neq 0$) **alors**

2 : *approche* \leftarrow 0

3 : **sinon**

4 : **si** ($n_s \neq 0$) **alors**

5 : *approche* \leftarrow 2

6 : **sinon**

7 : **si** ($\frac{n_m}{l} \gg 1$) **alors**

8 : *approche* \leftarrow 2

9 : **sinon**

10 : *approche* \leftarrow 1

11 : **fin si**

12 : **fin si**

13 : **fin si**

2.7 Conclusion

Dans tout problème de reconnaissance, il faut d'abord identifier ce qui est dynamique et ce qui est statique. On peut ensuite chercher à simplifier le problème en rendant statique soit les caractéristiques (cas des objets dynamiques), soit les classes (cas des classes dynamiques). Cette "simplification" consiste en fait à effectuer un passage d'un espace de représentation de dimension n vers un espace de dimension $n - 1$ (la dimension perdue étant le temps). Les classes peuvent alors prendre des formes complexes nécessitant une approche adaptée. Les *Systèmes de Classifieurs*, alliant les points de vue de plusieurs classifieurs spécialisés dans l'espace de représentation, sont bien adaptés à la reconnaissance pour ce type de classes. Si cette approche échoue, alors les classes (qui sont alors dynamiques) peuvent être reconnues en utilisant un système de reconnaissance dynamique qui spécialise chaque classifieur (ou *Système de Classifieurs*) sur un interval temporel associé. Dans le cas des évolutions périodiques des classes, chaque interval est déterminé lors de la phase d'apprentissage. Cette phase s'effectue sur au moins une période complète.

Le choix entre ces deux approches peut être conduit selon différents critères. Nous proposons d'adopter les suivants :

- existence ou non d'intersections spatiales (dans l'espace de représentation) ou spatio-temporelles entre les trajectoires des classes impliquées,
- la "granularité" de la trajectoire des classes : on définit des modes le long de la trajectoire décrivant l'évolution d'une classe. Le nombre optimal de modes (sur l'ensemble des trajectoires d'évolution), calculé à l'aide d'algorithmes de catégorisation, est-il proche du nombre de classes réelles, ou bien, lui est-il très supérieur ?

Le premier critère se fonde sur une contrainte générale, vraie pour tout problème de RdF. Le second est plus spécifique et permet de choisir entre les deux approches. Le choix de la première approche peut conduire à l'obtention de classes multi-modales. L'algorithme 3C que nous proposons dans le chapitre suivant est alors particulièrement adapté (approche mixte : supervisé et non-supervisé). Dans le cadre de la seconde approche, nous proposons l'algorithme présenté dans le chapitre 4. L'intérêt principal de ces deux approches vient du fait qu'il n'est pas utile d'extraire le mécanisme qui engendre l'évolution des trajectoires et donc de modéliser la structure probabiliste sous-jacente. Elles nécessitent cependant que les conditions d'observations soient identiques pendant toute la durée de l'étude.

Chapitre 3

Coopération de classifieurs

Nous présentons, dans ce chapitre, une approche mettant en coopération une méthode de RdF non-supervisée et une méthode de RdF supervisée. Lors de la phase d'apprentissage du Système de Classifieurs, la méthode non-supervisée coopère avec plusieurs modules de classement en déterminant leurs ensembles d'apprentissage. Elle permet également d'effectuer une sélection (adaptative) des classifieurs les mieux adaptés lors de la phase de classement. Ensuite, les performances et l'influence des paramètres de l'algorithme, appliqué à un ensemble de données synthétiques, sont présentées.

Sommaire

| | | |
|------------|---|------------|
| 3.1 | Introduction | 77 |
| 3.2 | Une méthode de coopération de classifieurs | 78 |
| 3.3 | Détermination des ensembles d'apprentissage | 79 |
| 3.3.1 | Coalescence floue | 79 |
| 3.3.2 | Coalescence avec rejet d'ambiguïté | 81 |
| 3.3.3 | Le concept d'ambiguïté : frontières dans les problèmes de coalescence | 82 |
| 3.3.4 | Distance à une CCA | 85 |
| 3.3.5 | Distance à la classe de rejet de distance | 85 |
| 3.3.6 | L'algorithme c+2 moyennes floues | 86 |
| 3.4 | Une fusion multi-classifieurs parallèle et adaptée | 86 |
| 3.4.1 | La procédure de coalescence : Module de sélection | 87 |
| 3.4.2 | Les classifieurs | 88 |
| 3.4.3 | Module de fusion | 88 |
| 3.4.4 | Algorithmes | 89 |
| 3.5 | Application | 91 |
| 3.5.1 | Données synthétiques | 91 |
| 3.5.2 | Résultats | 91 |
| 3.6 | Influences des paramètres | 94 |
| 3.6.1 | Influence du taux de rejet de distance | 94 |
| 3.6.2 | Influence du taux de rejet d'ambiguïté | 98 |
| 3.6.3 | Ambiguïté élevée | 98 |
| 3.6.4 | Frontières de décisions | 98 |
| 3.6.5 | Autres types de combinaisons | 98 |
| 3.7 | Conclusion | 103 |

3.1 Introduction

Nous avons vu, dans le chapitre précédent, que lorsque les classes évoluent périodiquement, les individus peuvent constituer des classes de forme complexe. En particulier, si l'on veut que l'ensemble d'apprentissage soit bien représentatif des individus à classer, il est nécessaire d'acquérir les individus sur au moins une période complète. On rencontre cette problématique dans le cas du tri automatique de poissons par espèce.

Dans ce chapitre¹⁶, nous proposons une solution originale permettant de combiner les résultats de plusieurs classifieurs. L'objectif est de pouvoir choisir un ou plusieurs classifieurs adaptés à la région de l'espace de représentation dans laquelle se trouve un individu à classer. L'architecture générale du *Système de classifieurs* que nous proposons est fondée sur une méthode RdF non-supervisée permettant de "guider" l'apprentissage d'un ensemble de méthodes de RdF supervisées et de sélectionner les classifieurs les mieux adaptés à un individu à classer. L'objectif étant, bien évidemment, d'atteindre des taux de reconnaissance plus élevés que ceux obtenus grâce à un seul classifieur.

Une phase préliminaire de coalescence permet de diviser l'espace de représentation en c régions de densités élevées. Chaque classifieur apprend sur une de ces régions et sur toutes les régions ambiguës à proximité de celle-ci. En effectuant une itération supplémentaire de l'algorithme de coalescence appliqué à un individu à classer, nous déterminons la région dans laquelle il se trouve. Ainsi, nous sommes capables de sélectionner tous les classifieurs ayant appris sur la région (ambiguë ou non) à laquelle est affecté l'individu par la méthode de RdF non-supervisée. De cette manière, plusieurs classifieurs peuvent être retenus (sélection adaptative) et leurs sorties peuvent être combinées pour une prise de décision finale.

Nous utilisons la spécificité des approches supervisées et non-supervisées pour spécialiser les classifieurs sur un ensemble d'apprentissage. Lorsque le problème de reconnaissance est complexe (exemple : classes multi-modales ou classes qui se chevauchent), deux approches sont possibles. La première consiste à chercher des modèles plus représentatifs de ces classes complexes. Par exemple, dans les travaux de McLachlan et Basford [76] chaque distribution de probabilités est approximée par des modèles de mélange. La deuxième approche réalise une partition de l'espace de représentation, chaque sous-espace obtenu pouvant être modélisé plus simplement. Dans ce but, Ben-Hur et al. [7] ont récemment proposé une méthode qui modélise les classes par un ensemble de vecteurs de support (i.e. un ensemble d'individus qui se trouvent sur les frontières de décisions). Pour tenir compte de classes qui se chevauchent ou qui sont mal définies, nous proposons de gérer les régions de doute / d'ambiguïté. Pratiquement, chaque sous-espace obtenu représente l'ensemble d'apprentissage d'un classifieur. La prise en compte de l'ambiguïté permet de gérer la redondance entre classifieurs voisins (i.e., dont l'intersection entre ensembles d'apprentissage n'est pas vide).

¹⁶Ce chapitre a fait l'objet d'une publication dans une revue internationale [46], dans des congrès internationaux [44] [45] et dans un congrès national [40].

Nous proposons de spécialiser chaque classifieur sur une région donnée de l'espace de représentation. Plus précisément, notre approche consiste à utiliser une méthode de coalescence floue avec rejet d'*ambiguïté* pour le guidage de l'apprentissage supervisé des classifieurs. Les individus ambigus, constituent la frontière floue entre classes et représentent les vecteurs de support dans le sens précisé par Ben-Hur et al. L'ensemble d'apprentissage de chaque classifieur est alors défini sur une classe de coalescence et les classes de rejet d'ambiguïté qui lui sont associées. Dans la phase de classement, le problème est considéré d'un point de vue multi-décisionnel (i.e., plusieurs modules de classement). L'algorithme que nous proposons permet de faire une sélection adaptative de classifieurs, car la prise de décision fait suite à la fusion des sorties des classifieurs les mieux adaptés.

3.2 Une méthode de coopération de classifieurs

Afin de différencier clairement le nombre de classes réelles du nombre de classes de coalescence (CC, en Anglais : cluster) correspondant à l'ensemble des modes des classes réelles, nous les identifions respectivement par l et c . Dans notre approche, une classe de coalescence *ambiguë* et une classe de coalescence de *rejet d'ambiguïté* (CCA) ont la même signification. La méthode est définie de la manière suivante :

- Soit $X = (x_k)_{k \in [1, n]}$ l'ensemble des n individus où

$$x_k = (x_{k1}, x_{k2}, \dots, x_{kp})^t$$

est un individu décrit par p caractéristiques (i.e. $x_k \in \mathbb{R}^p$);

- Soit $\Phi = (\phi_i)_{i \in [1, l]}$ un ensemble de l classes réelles. Soient $X^{lr} \subset X$, un ensemble d'individus d'apprentissage étiquetés selon une famille connue Φ , $X^t \subset X$ (avec $X^t \cap X^{lr} = \emptyset$) un ensemble d'individus inconnus. La reconnaissance est liée aux classes d'individus à discriminer dans l'espace de représentation à p dimensions.
- Soit 2^Ω , l'ensemble de puissance de Ω (par exemple, pour $\Omega = \{\omega_1, \omega_2, \omega_3\}$, on a $2^\Omega = \{\omega_1, \omega_2, \omega_3, \omega_{12}, \omega_{13}, \omega_{23}, \omega_{123}\}$). Par conséquent, pour un problème de coalescence à c CCs, $(2^c - c - 1)$ CCAs peuvent exister (une CCA existe si elle n'est pas vide).

Un algorithme de coalescence génère une nouvelle partition de X^{lr} , composée de CCs (classes de coalescence) et de CCAs (classe de coalescence de *rejet d'ambiguïté*). Cette partition est obtenue en utilisant l'algorithme des Fc+2M (Ménard [78] [77]). Ce dernier est un algorithme des c -moyennes floues généralisées qui exerce le rejet d'ambiguïté lors de la phase de coalescence (cf. section suivante). La figure 3.1 met en évidence les CCs et les CCAs par des cercles ou des ellipses en fonction du maximum des degrés d'appartenance des individus. Dans cet exemple les CCAs non vides sont ω_{12} et ω_{12} .

Un classifieur e_i , noté également $e_i(x_k)$ est alors associé à chaque classe de coalescence ω_i , avec $i \in [1, c]$ (cf. figure 3.1). Puis l'ensemble d'apprentissage S_{ω_i} (mode supervisé) de chaque classifieur e_i est défini sur sa CC et ses CCAs associées (cf. figure 3.1 et 3.2). En ce sens, il y a une coopération entre une méthode de classification et une méthode de classement¹⁷. Chaque classifieur apprend séparément. Lorsqu'un individu à classer est ambiguë (la méthode de coalescence est alors utilisée en classement), l'algorithme de coalescence permet de sélectionner les classifieurs les mieux adaptés (ceux dont l'intersection entre l'ensemble d'apprentissage et la CCA à laquelle est affecté l'individu n'est pas vide). La prise de décision, pour cet individu, est le résultat de la fusion des sorties des classifieurs sélectionnés. La figure 3.3 résume ce qui vient d'être dit.

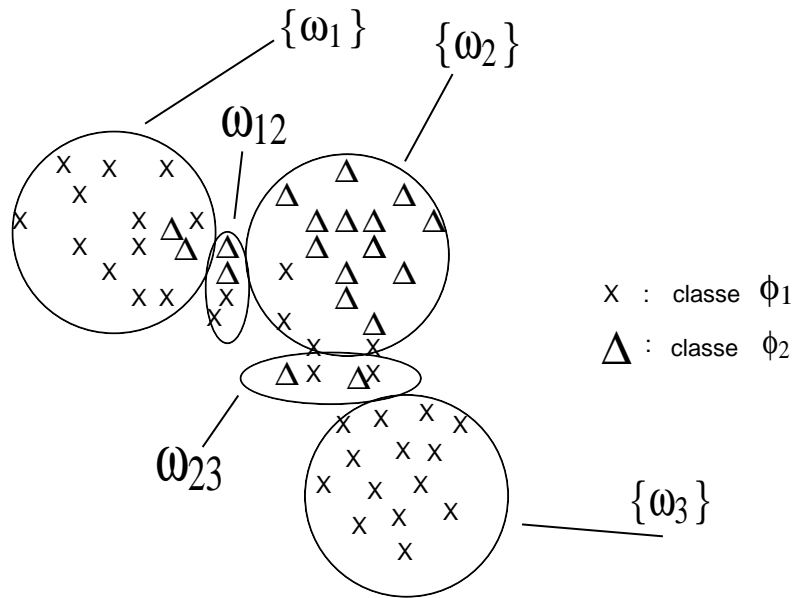


FIG. 3.1 – Partition des ensembles d'individus d'apprentissage, $X^{lr} \subset X$, lorsque le rejet d'ambiguïté est exercé. $\{\omega_1\}$, $\{\omega_2\}$, $\{\omega_3\}$: CCs ; $\omega_{12} = \{\omega_1, \omega_2\}$, $\omega_{23} = \{\omega_2, \omega_3\}$: CCAs.

3.3 Détermination des ensembles d'apprentissage

3.3.1 Coalescence floue

Le problème de la coalescence floue concerne un ensemble de méthodes qui divisent une famille d'individus en classes de coalescence ou CCs, disjointes deux à deux, toutes non vides, et qui reproduisent l'ensemble des données originelles par union. Les individus appartenant à une même CC partagent des propriétés communes qui les distinguent des individus appartenant à d'autres CCs. Dans le domaine de la RdF, les techniques de coalescence floue tentent de trouver une famille $U = (\mu_i)_{i \in [1, c]}$ de fonctions appelées fonctions

¹⁷Selon Postaire [84], une méthode de classification se différencie d'une méthode de classement par le fait que l'on ne dispose pas de prototypes de chacune des classes en présence.

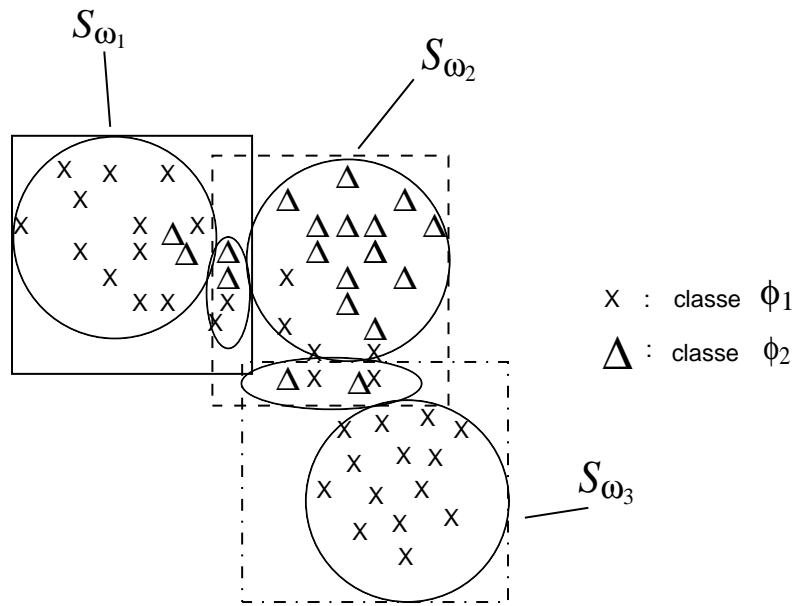


FIG. 3.2 – Un classifieur est associé à chaque CC. L'ensemble d'apprentissage, S_{ω_i} du classifieur e_i , est défini sur ω_i et ses CCAs associées (par exemple, la CCA ω_{12} est associée à ω_1 et à ω_2).

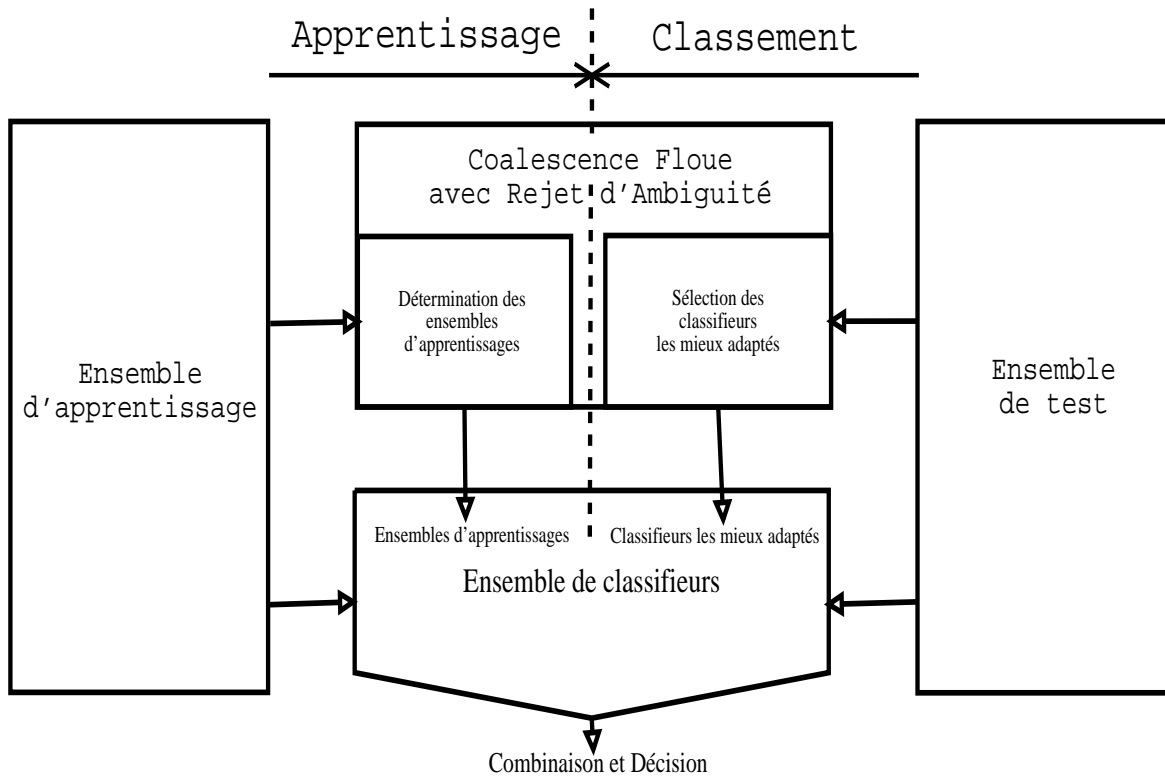


FIG. 3.3 – Principe de la coopération.

d'appartenance telles que :

$$\begin{aligned} \mu_i : X &\rightarrow [0, 1] \\ x_k &\mapsto \mu_i(x_k) \end{aligned}$$

vérifiant $\forall x_k \in X, \sum_{i=1}^c \mu_i(x_k) = 1$. Les algorithmes de coalescence floue génèrent une partition floue de c CCs fournissant une mesure de degré d'appartenance de chaque individu à une CC donnée. Le problème classique des c -moyennes floues (FcM ou fuzzy c -means, en Anglais) consiste à trouver une partition floue de la famille X qui minimise la fonction objective :

$$J_m^0(U, V) = \sum_{i=1}^c \sum_{k=1}^n \mu_{ik}^m d_{ik}^2 \quad (3.1)$$

où $m > 1$ est un exposant de "fuzzification", appelé Facteur de Flou, $\mu_{ik} = \mu_i(x_k)$ et $d_{ik} = \|x_k - v_i\|_G$ est une mesure de distance de x_k au i^{me} prototype de CC, $v_i \in V = [v_1, v_2, \dots, v_c]$, V est une matrice de CCs et G est une norme.

3.3.2 Coalescence avec rejet d'ambiguïté

Pour réduire un taux d'erreur (mauvais classement) excessif dû au bruit (affectant les Formes) et aux autres facteurs d'incertitudes inhérents aux mesures réelles, la coalescence avec rejet d'ambiguïté apporte une solution intéressante. Dans la littérature, les règles permettant de rejeter un individu situé sur la frontière de deux CCs sont fondées sur des valeurs de seuils dans la phase de décision et non dans la phase de coalescence ou d'apprentissage. Pour spécifier cette prise de décision, il est courant de caractériser un individu x_k avec un concept d'ambiguïté.

Le rejet d'ambiguïté a été introduit dans la phase de décision par Chow [18] [19]. Le but est de mesurer la proximité de x_k aux frontières de décision. Dans le cas de la règle de Bayes, le problème est de définir une nouvelle classe de coalescence ω_0 , appelée *classe de rejet*, associée à un coût de rejet constant $Cr = C(0/j) \forall j \in [1, c]$:

$$\begin{cases} e(x_k) = i, & \text{si } P(\omega_i/x_k) = \max_{j \in [1, c]} (P(\omega_j/x_k)) > 1 - Cr \\ e(x_k) = \omega_0 & \text{sinon} \end{cases} \quad (3.2)$$

où $(P(\omega_j/x_k))_{j \in [1, c]}$ sont les probabilités *a posteriori*. Le terme Cr , qui contrôle le taux de rejet, doit être choisi entre 0 et $\frac{c-1}{c}$.

Cette règle partage l'espace de représentation en $(c + 1)$ classes de coalescence. Dans [27], Dubuisson affirme que ce rejet est un rejet d'ambiguïté puisque la région correspondante dans l'espace de représentation est toujours située entre les CCs. Le principal inconvénient de ces définitions et de leur utilisation est que ce rejet consiste à fixer un seuil *a priori*, sans tenir compte de la "forme" des classes d'apprentissage. Une extension de ce rejet simple est le rejet sélectif introduit par Ha [47].

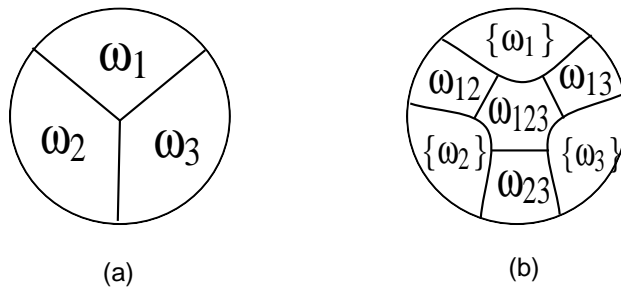


FIG. 3.4 – La figure (a) illustre la partition de l'espace de représentation en trois CCs. La figure (b) illustre la partition de l'espace de représentation lorsque le concept d'ambiguïté est appliqué. Nous associons une CCA à chacun des sous-ensembles $A \in 2^\Omega \setminus \{\emptyset, (\{\omega_i\})_{i \in [1,3]}\}$.

Soit $\Omega = (\omega_i)_{i \in [1,c]}$, un ensemble de CCs. Afin de tenir compte de l'ambiguïté entre CCs, des fonctions d'appartenance sont affectées à tous les éléments de 2^Ω , au lieu des éléments de Ω seulement. Puisqu'il y a 2^c sous-ensembles dans un problème à c classes de coalescence, nous cherchons à réaliser une partition de l'espace des représentations en $(2^c - 1)$ CCs et CCAs (excluant l'ensemble vide). Il y a donc c CCs, $(\{\omega_i\})_{i \in [1,c]}$ et autant de CCAs qu'il y a de combinaisons entre ces CCs, i.e., le nombre de sous-ensembles $A \in 2^\Omega \setminus \{\emptyset, (\{\omega_i\})_{i \in [1,c]}\}$. Dans la figure 3.4, il y a $(2^3 - 1)$ CCs et CCAs correspondants aux sous-ensembles :

$$\begin{aligned} & \{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \\ & \omega_{12} = \{\omega_1, \omega_2\}, \\ & \omega_{13} = \{\omega_1, \omega_3\}, \\ & \omega_{23} = \{\omega_2, \omega_3\} \text{ et} \\ & \omega_{123} = \{\omega_1, \omega_2, \omega_3\}. \end{aligned}$$

Définition 4 Soit $\Omega = \{\omega_1, \dots, \omega_c\}$ l'ensemble des classes de coalescence ω_i . Nous associons une CCA à chacun des sous-ensembles de classes $A \in 2^\Omega \setminus \{\emptyset, (\{\omega_i\})_{i \in [1,c]}\}$. Nous considérons qu'il existe un ensemble de $\sum_{k>1} \binom{c}{k}$ CCAs disjointes dans un problème à c classes de coalescence.

Dans la prochaine section, nous construisons une fonction objective appropriée dans le but de prendre en compte explicitement le rejet d'ambiguïté dans la coalescence floue.

3.3.3 Le concept d'ambiguïté : frontières dans les problèmes de coalescence

Nous faisons la distinction entre deux types de frontières : le premier est net et est défini dans le cadre de la théorie des ensembles approximatifs, le second est obtenu en généralisant ce concept dans le cadre de la théorie des sous-ensembles flous. Le premier utilise

les relations d'indiscernabilité, le second, les fonctions à valeurs réelles. Nous étendons le concept de frontière au concept de frontière floue de CC d'une partition floue en utilisant la théorie des ensembles approximatifs de Pawlak [83] pour préciser le concept d'ambiguïté et définir la fonction objective. Cette théorie modélise l'ambiguïté des connaissances, attribuée à la granularité des connaissances, i.e., à l'indiscernabilité et aux approximations.

Un système d'information est une paire $\mathcal{A} = (U, A)$, où U est un ensemble fini, non vide appelé l'univers et A , un ensemble fini, non vide, d'attributs, i.e. $a : U \rightarrow V_a \forall a \in A$, où V_a est l'ensemble des valeurs de a . Par V nous notons l'ensemble $\cup\{V_a : a \in A\}$. A chaque sous-ensemble d'attributs $B \subseteq A$, une relation appelée B-indiscernable notée $IND_{\mathcal{A}}B$ est associée et est définie par :

$$IND_{\mathcal{A}}B = \{(u, u') \in U^2 : a(u) = a(u') \forall a \in B\} \quad (3.3)$$

Les objets u et u' satisfaisant la relation $IND(B)$ sont indiscernables selon B car ils ont les mêmes valeurs relativement aux attributs de B . On appelle ensemble élémentaire (par rapport à B) un ensemble $[u]_B$ d'objets qui sont tous indiscernables par rapport à B .

Si $\mathcal{A} = (U, A)$ est un système d'information, $B \subseteq A$ est un ensemble d'attributs et $X \subseteq U$ est un ensemble d'objets, alors les ensembles :

$$\underline{B}X = \{u \in U : [u]_B \subseteq X\} \quad (3.4)$$

et

$$\overline{B}X = \{u \in U : [u]_B \cap X \neq \emptyset\} \quad (3.5)$$

sont appelés respectivement la plus basse et la plus haute approximations de X dans \mathcal{A} , et sont notées $\underline{B}X$ et $\overline{B}X$, respectivement. L'ensemble $BN_B(X) = \overline{B}X - \underline{B}X$ est appelé frontière de X . L'ensemble $\underline{B}X$ est l'ensemble de tous les éléments de U qui peuvent être classés de façon certaine comme éléments de X , avec la connaissance représentée par les attributs de B , tandis que l'approximation haute de X , $\overline{B}X$, est l'ensemble de tous les éléments pouvant appartenir à X . L'ensemble $BN_B(X)$ est constitué des éléments que l'on ne peut classer ni dans X , ni dans $-X$, avec la connaissance de B .

Les informations sur les individus sont souvent disponibles sous forme de tables. Supposons que l'on ait les individus x_k , avec $k \in [1, ..7]$, comme indiqués dans la Table 3.1. Soit $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Chaque individu x_k sera d'abord supposé posséder l'étiquette d'une classe de coalescence $I \subseteq \{1, 2, 3\}$ indiquant avec certitude son appartenance à un sous-ensemble de 2^Ω . Ainsi, il y a 4 classes correspondants aux CCAs $\omega_{12}, \omega_{13}, \omega_{23}, \omega_{123}$ et 3 CCs $\{\omega_1\}$, $\{\omega_2\}$ et $\{\omega_3\}$. L'attribut *décision* se trouve dans la troisième colonne. Pour $x_k \in \omega_I$, l'attribut B est "oui" si $1 \in I$ et "non" sinon.

Dans la Table 3.1 les individus x_1, x_4, x_5 et x_7 sont indiscernables selon l'attribut B . L'attribut B génère deux ensembles élémentaires $\{x_1, x_4, x_5, x_7\}$ et $\{x_2, x_3, x_6\}$. x_1 possède

| x_k | B | décision : ω |
|-------|-----|---------------------|
| 1 | oui | $\{\omega_1\}$ |
| 2 | non | $\{\omega_2\}$ |
| 3 | non | $\{\omega_3\}$ |
| 4 | oui | ω_{12} |
| 5 | oui | ω_{13} |
| 6 | non | ω_{23} |
| 7 | oui | ω_{123} |

 TAB. 3.1 – Exemple de Système d'Information pour le concept "appartenant à ω_1 ".

les informations qui nous permettent de le classer, avec certitude comme appartenant à la CC ω_1 . Ainsi, la plus basse approximation de l'ensemble des individus appartenant à la classe de coalescence ω_1 est l'ensemble $\{x_1\} = \{x_k \in \{\omega_1\}\}$ et la plus haute approximation de cet ensemble est l'ensemble :

$$\{x_1, x_4, x_5, x_7\} = \{x_k \in \omega_I \text{ où } \omega_I \in 2^\Omega \setminus \emptyset \text{ avec } 1 \in I\} \quad (3.6)$$

La frontière du concept "appartenant à ω_1 " est l'ensemble :

$$\{x_4, x_5, x_7\} = \{x_k \in \omega_I \text{ où } \omega_I \in 2^\Omega \setminus \{\emptyset, \{\omega_1\}\} \text{ avec } 1 \in I\} \quad (3.7)$$

Cet exemple illustre de manière simple la frontière d'une classe.

Définition 5 Soit $\Omega = (\omega_i)_{i \in [1, c]}$, un ensemble de CCs. Soit $X = (x_k)_{k \in [1, n]}$, l'ensemble des individus où $x_k = (x_{k1}, x_{k2}, \dots, x_{kp})^t$ est un individu décrit par p caractéristiques (i.e. $x_k \in \mathbb{R}^p$). Nous définissons la frontière de la classe de coalescence ω_i avec $i \in [1, c]$ comme : $F_{\omega_i} = \{x_k \in A \text{ où } A \in 2^\Omega \setminus \{\emptyset, \{\omega_i\}\} \text{ avec } \omega_i \in A\}$

Dans une partition floue, chaque individu x_k peut appartenir avec un certain degré aux frontières de chaque CC. Aussi, en nous appuyant sur la définition précédente, nous proposons une mesure de la taille de la frontière floue de la classe de coalescence ω_i :

$$\sum_{k=1}^n \left(\sum_{A \in 2^\Omega \setminus \{\emptyset, \{\omega_i\}\}, \omega_i \in A} \mu_{Ak}^m d_{Ak}^2 \right) = \sum_{k=1}^n \left(\sum_{A \in 2^\Omega \setminus \{\emptyset, (\{\omega_j\})_{j \in [1, c]}\}, \omega_i \in A} \mu_{Ak}^m d_{Ak}^2 \right), \quad (3.8)$$

où d_{Ak} est la distance de l'individu x_k à la CCA A : $d_{Ak} = d(x_k, v_A)$, m est un facteur de flou et μ_{Ak} est le degré d'appartenance de l'individu x_k à la CCA A . Le choix de d_{Ak} sera discuté plus loin. Pour des grandes valeurs de d_{Ak} , si μ_{Ak} prend des valeurs importantes alors la taille de la frontière floue de la i^{eme} classe doit être importante. Du fait qu'une frontière est commune à plusieurs CCs, la taille totale des frontières floues, dans une

partition floue, est définie par :

$$J_m^a(U, V) = \sum_{k=1}^n \left(\sum_{A \in 2^\Omega \setminus \{(\{\omega_j\})_{j \in [1, c]}, \emptyset\}} \mu_{Ak}^m d_{Ak}^2 \right). \quad (3.9)$$

Aussi, nous définissons le critère de coalescence comme :

$$J_m^1 = J_m^0(U, V) + J_m^a(U, V) = \sum_{k=1}^n \sum_{i=1}^c \mu_{ij}^m \|x_k - v_i\|^2 + \sum_{k=1}^n \left(\sum_{A \in 2^\Omega \setminus \{(\{\omega_j\})_{j \in [1, c]}, \emptyset\}} \mu_{Ak}^m d_{Ak}^2 \right). \quad (3.10)$$

3.3.4 Distance à une CCA

Dans le problème de la coalescence floue, les régions ambiguës, situées entre les CCs, peuvent être considérées comme des discontinuités des fonctions d'appartenance définies sur les CCs. Pour détecter ces discontinuités, la distance d'un individu x_k à la CCA $A \in 2^\Omega \setminus \{\emptyset, (\{\omega_j\})_{j \in [1, c]}\}$ est définie par :

$$d_{Ak} = \frac{\alpha (\sum_{\omega_i \in A} d_{ik})^2}{|A|^2 (\prod_{\omega_i \in \Omega} d_{ik})^{\frac{1}{c}}} \quad (3.11)$$

où, le terme α est un nombre positif ajusté de manière à détecter les discontinuités. Pour augmenter le rejet d'ambiguïté il faut diminuer α . d_{ik} est la distance classique utilisée dans l'algorithme des FcM (Bezdek [9]).

3.3.5 Distance à la classe de rejet de distance

La distance entre un individu et une classe de coalescence de rejet de distance est expliquée par Demko et al. [22] et vaut ;

$$d_{\Theta k} = \frac{T(\rho)^2}{(\prod_{i=1}^c d_{ik})^{\frac{1}{c}}} \quad (3.12)$$

où $T(\rho)$ est une valeur seuil permettant d'obtenir un taux global de rejet ρ . Pour simplifier, nous le notons T . Pour augmenter le rejet de distance il faut diminuer T et réciproquement.

3.3.6 L'algorithme $c+2$ moyennes floues

Cet algorithme (Ménard [78] [77]) est déduit des solutions de la minimisation de J_m^1 (à laquelle est ajoutée un troisième terme pour prendre en compte le rejet de distance). Dans ce travail, le rejet de distance est inhibé (sauf pour l'étude des frontières de décisions).

Soit Θ , la classe de coalescence de rejet de distance. D'abord, une initialisation des degrés d'appartenance est effectuée. Ensuite, de manière itérative, les centres de classes sont calculés avec (3.13) et les degrés d'appartenance sont calculés avec (3.14). On arrête d'itérer lorsque la différence entre les centres obtenus dans deux itérations consécutives devient inférieure à une valeur seuil choisie.

$$\forall k \in [1, n], \forall A \in 2^\Omega \setminus \{\emptyset, \Theta\},$$

$$v_A = \frac{\sum_{k=1}^n \mu_{Ak}^m x_k}{\sum_{k=1}^n \mu_{Ak}^m} \quad (3.13)$$

$$\forall k \in [1, n], \forall A \in 2^\Omega \cup \Theta \setminus \emptyset,$$

$$\mu_{Ak} = \frac{1}{\sum_{B \in 2^\Omega \cup \Theta \setminus \emptyset} \left(\frac{d_{Ak}}{d_{Bk}} \right)^{\frac{2}{m-1}}} \quad (3.14)$$

où $\forall A = \{\omega_i\}, i \in [1, c]$, $d_{Ak} = \|x_k - v_A\|_G$ est la distance classique utilisée dans l'algorithme des c -moyennes floues.

Dans notre approche, le principal avantage de cet algorithme est qu'il permet de calculer des centres qui sont plus proches (si on le compare par exemple aux c -moyennes floues) des centres physiques. Ceci est dû au fait que dans le calcul des centres, les individus ambigus sont moins pris en compte. Nous fournissons un exemple pour illustrer la partition de l'espace des caractéristiques en 3 CCs et 4 CCAs. Les données des fleurs IRIS (Fisher [30]), qui ont été souvent utilisées dans l'évaluation des performances des algorithmes de coalescence, sont composées de 150 individus répartis en trois classes physiques représentant différentes sous-espèces des IRIS. Une des trois classes est clairement séparée des deux autres, tandis que les deux dernières admettent un chevauchement important entre elles. La figure 3.5 (respectivement 3.6) montre la forme de la fonction d'appartenance maximale des trois CCs (respectivement, des quatre CCAs) dans le plan factoriel principal, pour les données IRIS. La projection de cette fonction en termes de valeurs d'iso-appartenances semble produire des classes de coalescence ellipsoïdales, ce qui est le résultat espéré (classes de formes ellipsoïdales).

3.4 Une fusion multi-classifieurs parallèle et adaptée

Le système global comporte quatre composants principaux :

- un ensemble de c classifieurs $(e_k)_{k \in [1, c]}$ qui déterminent la probabilité qu'un individu de test particulier appartienne ou non à la classe réelle $\phi_i, i \in [1, l]$,

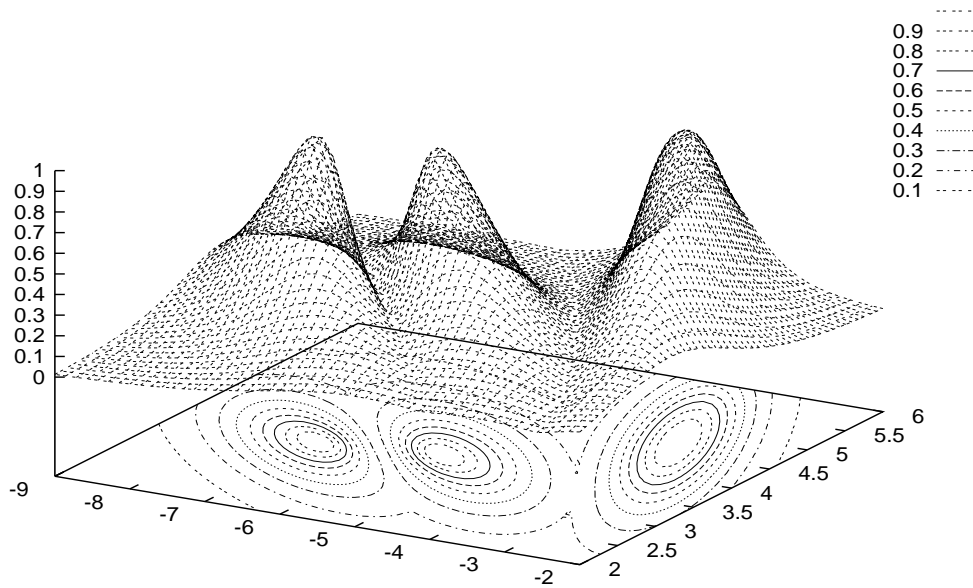


FIG. 3.5 – La fonction d'appartenance maximale dans le plan factoriel principal des données IRIS, parmi les trois CCs, avec $m = 1.2$ et $\alpha = 0.8$. Distance de Mahalanobis.

- un module de sélection \mathcal{S} qui choisit, pour un individu donné, le sous-ensemble de classifieurs (les mieux adaptés à l'individu à classer) à partir de l'ensemble des classifieurs,
- un module de fusion \mathcal{F} qui combine les sorties des sous-ensembles de classifieurs,
- un module de décision \mathcal{D} qui affecte l'individu de test à l'une des l classes.

3.4.1 La procédure de coalescence : Module de sélection

L'idée principale est que la procédure de coalescence peut gérer la sélection des ensembles d'apprentissage pour un mode supervisé. Soit X^{lr} , un ensemble d'individus d'apprentissage étiquetés selon les classes connues. L'algorithme des Fc+2M génère une nouvelle partition \mathcal{P} de X^{lr} , composée de CCs et de CCAs. Le nombre de CCs, c , est généralement choisi plus grand que le nombre de classes l . Ce procédé autorise la coopération entre classifieurs, permettant ainsi la gestion de la décision à points de vue multiples. Ainsi, la procédure de coalescence permet-elle de gérer la sélection des classifieurs les mieux adaptés.

Définition 6 Soit 2^Ω , l'ensemble des classes CCs et CCAs. Nous associons un classifieur e_i à chaque classe de coalescence ω_i (autant de classifieurs que de régions denses en individus). L'ensemble d'apprentissage, S_{ω_i} , pour le classifieur e_i , est défini sur $S_{\omega_i} = \{x_k \in A \text{ où } A \in 2^\Omega \text{ avec } \omega_i \in A\}$. Les classes de décision, pour le classifieur e_i , sont définies

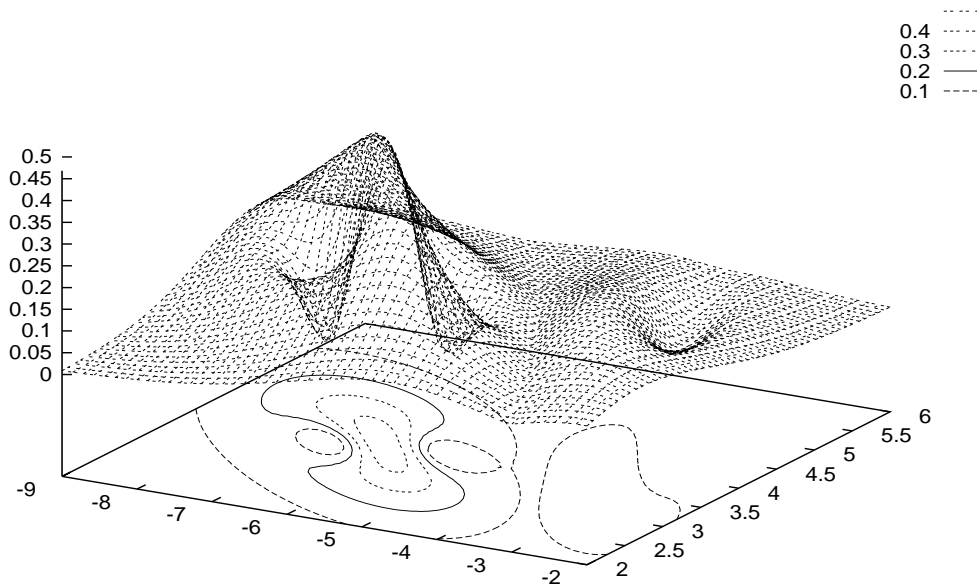


FIG. 3.6 – La fonction d'appartenance maximale dans le plan factoriel principal des données IRIS, parmi les quatre CCAs, avec $m = 1.2$ et $\alpha = 0.8$. Norme de Mahalanobis.

par $\{\phi_j \in \Phi : \phi_j \ni x_k \forall x_k \in S_{\omega_i}\}$.

Définition 7 Soit $(e_i)_{i \in [1, c]}$, l'ensemble des classifieurs. L'ensemble E_k des classifieurs les mieux adaptés, pour un individu x_k , est défini par :

$$E_k = \{e_i : S_{\omega_i} \cap \omega_A \neq \emptyset\} \text{ avec } A \in 2^\Omega \setminus \{\emptyset, (\{\omega_j\})_{j \in [1, c]}\} \quad (3.15)$$

3.4.2 Les classifieurs

Le classifieur Bayésien (voir page 61) est utilisé pour cette étude. L'hypothèse principale est la connaissance des lois de distribution dont les caractéristiques sont déterminées d'après l'ensemble d'apprentissage, X^{lr} . Un classifieur est associé à chaque CC. L'ensemble d'apprentissage S_{ω_i} , pour le classifieur e_i , est défini sur ω_i et ses CCAs associés : $S_{\omega_i} = \{x_k \in A \text{ où } A \in 2^\Omega \text{ avec } \omega_i \in A\}$. Les lois de distribution de probabilités *a priori*, pour le classifieur e_i , sont notées $P_i(x_k/\phi_j)_{k \in [1, l]}$. L'approche Bayésienne conduit typiquement à évaluer la probabilité *a posteriori* $P_i(\phi_j/x_k)$ pour chaque classe ϕ_j à l'aide du théorème de Bayes.

3.4.3 Module de fusion

Soit $X^t \subset X$ (avec $X^t \cap X^{lr} = \emptyset$), l'ensemble des individus inconnus. Lors du classement, un individu inconnu $x_k \in X^t$ est affecté à une CC ou à une CCA. Le classement se

fait de la façon suivante :

- si x_k est situé dans la CC ω_i , les probabilités *a posteriori* $P_i(\phi_j/x_k)$ données par le classifieur e_i (associé à cette CC) sont utilisées pour prendre la décision finale avec la règle du maximum de probabilité *a posteriori*;
- si x_k est situé dans la CCA, $A \in 2^\Omega \setminus \{\emptyset, (\{\omega_j\})_{j \in [1,c]}\}$, la prise de décision est effectuée avec le résultat de la fusion de l'ensemble des sorties des classifieurs les mieux adaptés (i.e., classifieurs ayant appris sur la CCA à laquelle appartient x_k , cf. figure 3.7).

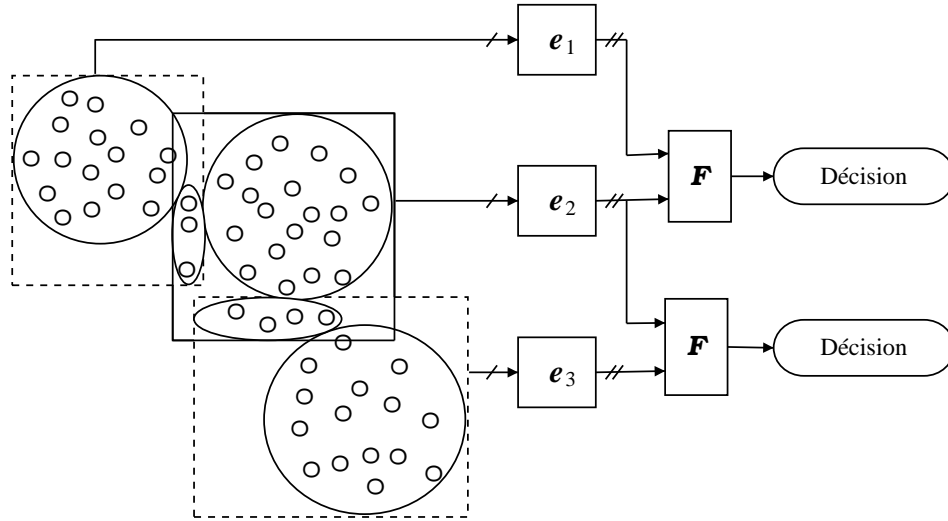


FIG. 3.7 – La prise de décision est le résultat de la fusion des sorties des classifieurs les mieux adaptés.

Plusieurs méthodes de combinaison des entités d'informations issues de plusieurs sources ont été présentées dans le chapitre 1. Le module de fusion que nous proposons est fondé sur une combinaison des sorties de classifieurs. Nous pouvons combiner les probabilités *a posteriori* par la règle de combinaison de Dempster, puisque nous émettons l'hypothèse que les classes sont équiprobables *a priori*. Les probabilités *a posteriori* constituent donc les masses affectées aux différentes classes. Cependant, dans le cas de classes non équiprobables, il est nécessaire d'utiliser d'autres modèles de fusion, par exemple ceux proposés par Appriou [4] (cf. page 102, pour un essai avec le premier de ses modèles).

3.4.4 Algorithmes

Construction du système de classifieurs (apprentissage)

Soit c , le nombre de classifieurs qui correspond aussi au nombre de classes de coalescence (CCs), et X^{lr} , l'ensemble d'apprentissage. Soient e_i , $i \in [1, c]$, un classifieur et e_{cat} un classifieur utilisant les résultats de la coalescence pour classer de nouveaux individus.

Procédure *ConstruireSystème*

Entrée : c : entier, X^l

Sortie : modèles représentant les classes ou fonctions représentant les frontières entre les classes $(e_1, e_2, \dots, e_c, e_{cat})$.

- 1 : Effectuer une coalescence avec rejet d'ambiguïté (construire e_{cat} par F_{c+2M})
- 2 : Identifier les CCAs existantes
- 3 : Constituer les sous-ensembles d'apprentissage S_{ω_i} pour tout e_i , $i \in [1, c]$
- 4 : Construire un classifieur e_i sur chaque S_{ω_i}

Classement d'un individu

Soit x_k , un individu à classer.

Procédure *ClassifierIndividu*

Entrée : $x_k, e_1, e_2, \dots, e_c, e_{cat}$

Sortie : étiquette $i \in [1, l]$

- 1 : Classifier x_k par le classifieur e_{cat}
- 2 : **si** (x_k appartient à une CCA) **alors**
- 3 : Décider avec e_i associé
- 4 : **sinon**
- 5 : Affecter x_k à la CCA (CCA_j)
- 6 : **si** ($CCA_j \neq \emptyset$) **alors**
- 7 : Classifier avec tous les e_i ayant appris sur CCA_j
- 8 : Combiner les résultats
- 9 : Décider
- 10 : **sinon**
- 11 : Rejeter x_k (ou remettre à jour le système)
- 12 : **fin si**
- 13 : **fin si**

Remarque : Dans le cas où le système de classifieurs classe un individu dans une *CCA* inexistante, alors on peut considérer que ce nouvel individu est soit inconnu (rejet), soit membre d'une nouvelle classe. Dans ce dernier cas, il faudra reconstruire le système de classifieurs ou le modifier (son apprentissage, i.e., ses connaissances) pour le remettre à jour. Dans cette étude, puisque nous ne permettons pas l'apparition de nouvelles classes (nous considérons que l'ensemble d'apprentissage est bien représentatif des classes réelles), nous rejetons cet individu. Pratiquement, lorsque la procédure Fc+2M est utilisée en classement (calcul des μ_{Ak}) et que le degré d'appartenance à une *CCA* inexistante (où aucun individu n'a été affecté lors de la phase de coalescence) est la plus grande, alors l'individu est rejeté.

3.5 Application

3.5.1 Données synthétiques

La méthode ci-dessus, que nous nommerons 3C (Coopération Classification-Classement), est testée sur un ensemble de données synthétiques dans \mathbb{R}^2 . L'ensemble d'apprentissage X^{lr} est composé de 210 individus. Avec un écart-type choisi, $\sigma = 2$, nous générons :

- ϕ_1 : classe composée de 100 individus provenant de la distribution Gaussienne ;

$$m_1 = \begin{pmatrix} 10 & 20 \end{pmatrix}^t, \Sigma_1 = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$$

- ϕ_2 : classe multi-modes composée de 110 individus générés de la manière suivante ($\Sigma_{2a} = \Sigma_{2b} = \Sigma_1$) ;

1. Deux populations Gaussiennes générées par :

$$m_{2a} = \begin{pmatrix} 10 & 10 \end{pmatrix}^t 50 \text{ individus},$$

$$m_{2b} = \begin{pmatrix} 10 & 30 \end{pmatrix}^t 50 \text{ individus}.$$

2. et 10 individus reliant ces deux modes entre eux (à droite de ϕ_1).

Cet ensemble est représenté sur la figure 3.8. L'ensemble de test est généré dans les mêmes conditions (cf. figure 3.9). La mesure de distance est faite selon une métrique euclidienne. Notons que, pendant la phase de classification (Fc+2M), le rejet de distance n'est pas prise en compte (les échantillons sont supposés représentatifs des CCs).

3.5.2 Résultats

Les paramètres choisis pour l'algorithme Fc+2M sont : $m = 1.5$, $T = 5$ (suffisant pour qu'aucun individu ne soit rejeté) et $\alpha = 0.1$. La partition obtenue par l'algorithme Fc+2M est représentée dans la Table 3.2 pour $c = 2$ et dans la Table 3.3 pour $c = 3$.

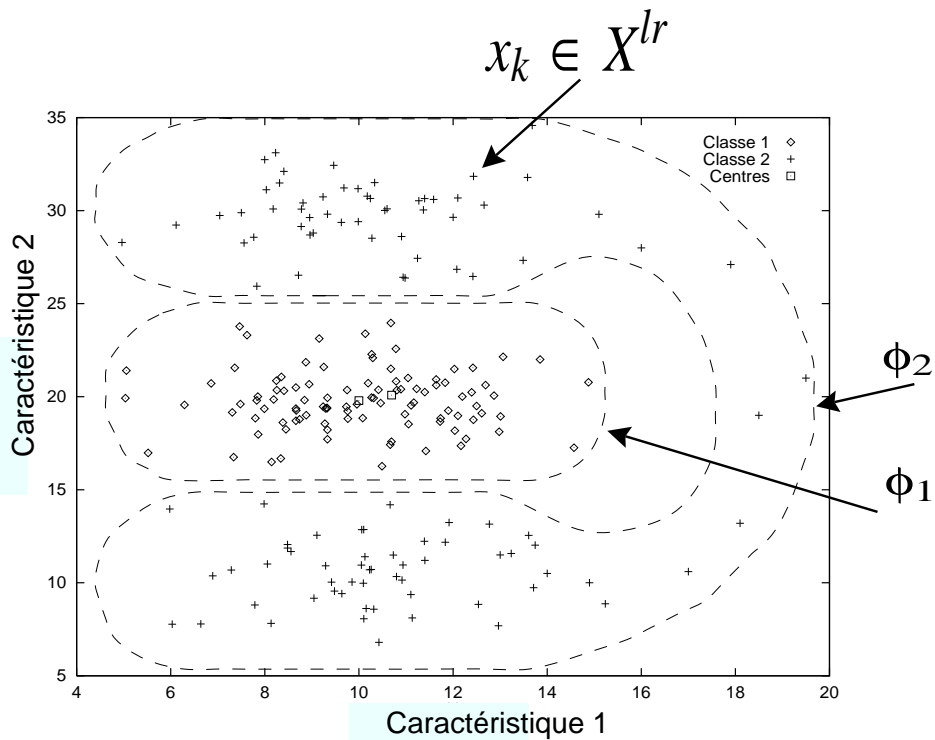


FIG. 3.8 – Ensemble d'apprentissage : $X^{lr} \subset X$.

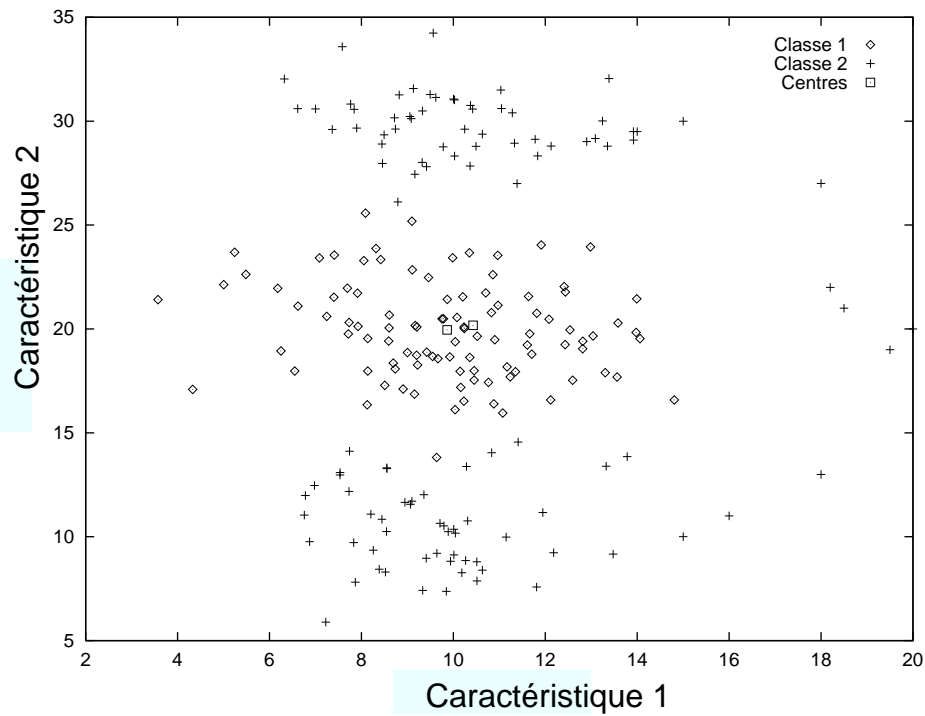
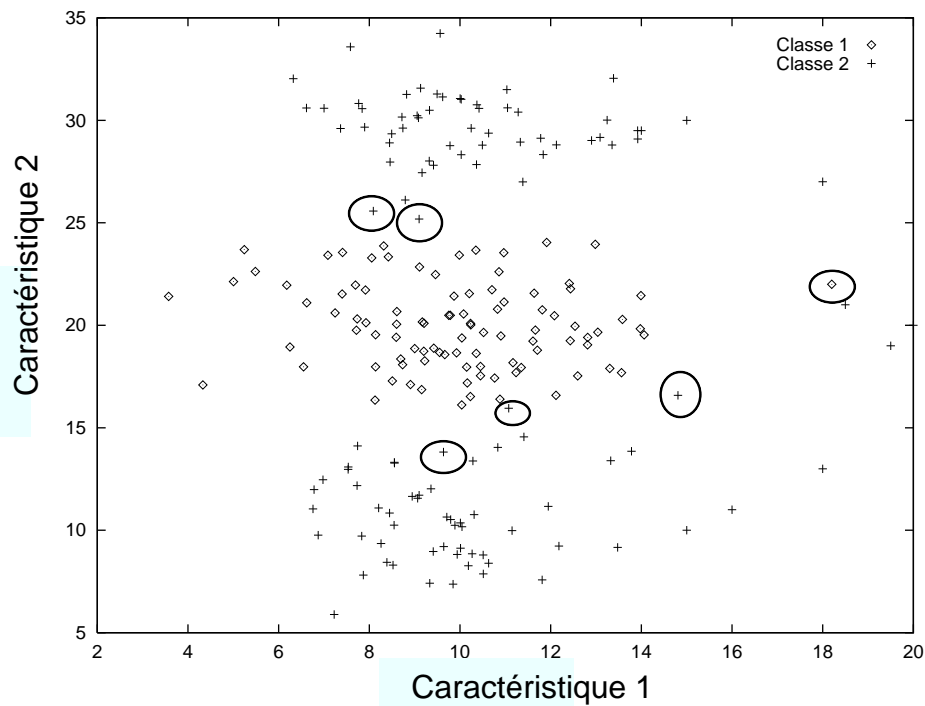
Toutes les valeurs de ces tables sont exprimées en pourcentage. Les résultats après fusions sont montrés dans la Table 3.4 (pour $c = 2$).

Une comparaison avec les résultats de la théorie Bayésienne de la décision appliquée à l'ensemble des données est effectuée (algorithme MAP). La Table 3.5 montre les résultats obtenus avec les trois méthodes (Bayésienne, algorithme 3C avec $c = 2$ et $c = 3$). La meilleure performance est obtenue pour la méthode 3C avec $c=2$ (97.1 % de succès, 88.6 % pour le classifieur Bayésien). Le mauvais résultat du classifieur Bayésien vient du fait que la classe ϕ_2 est bimodale. Les résultats obtenus sont montrés dans la figure 3.10 ($c=2$). Tous les résultats suivants (y compris ceux présentés dans la section suivante) sont obtenus avec $c = 2$.

| \nearrow | ω_1 | ω_2 | ω_{12} |
|------------|------------|------------|---------------|
| ϕ_1 | 48.0 | 15.0 | 37.0 |
| ϕ_2 | 50.91 | 49.09 | 0 |

TAB. 3.2 – Partition nette de X^{lr} obtenue par l'algorithme Fc+2M (pourcentages) pour $c = 2$: ϕ_1 et ϕ_2 sont les classes réelles ; ω_1 et ω_2 sont les CCs ; ω_{12} est la CCA.

La figure 3.11 représente les masses affectées à la classe 2 par l'algorithme 3C. Ces masses sont nulles au-delà d'un certain rayon à partir du centre du nuage de points. Sans exercer de rejet de distance, tout ce qui sort de la région affectée à la classe 1 est alors

FIG. 3.9 – Ensemble de test : $X^t \subset X$.FIG. 3.10 – Résultats obtenus avec notre méthode de coopération (3C) pour $c = 2$. Les individus entourés (6 parmi 210) sont ceux qui sont mal classés.

| \nearrow | ω_1 | ω_2 | ω_3 | ω_{12} | ω_{13} | ω_{23} | ω_{123} |
|------------|------------|------------|------------|---------------|---------------|---------------|----------------|
| ϕ_1 | 67.0 | 0 | 0 | 17.0 | 16.0 | 0 | 0 |
| ϕ_2 | 3.6 | 36.4 | 45.5 | 11.8 | 2.7 | 0 | 0 |

TAB. 3.3 – Partition nette de X^{lr} obtenue par l’algorithme Fc+2M pour $c = 3$: Il y a trois CCs et 4 CCAs possibles (2 d’entre elles, ω_{12} et ω_{13} , existent réellement).

| \nearrow | ϕ_1 | ϕ_2 |
|------------|----------|----------|
| ϕ_1 | 95.0 | 5.0 |
| ϕ_2 | 0.91 | 99.09 |

TAB. 3.4 – Algorithme 3C ; Matrice de confusion pour $c = 2$.

affecté à la classe 2 (cf. figure 3.12). La figure 3.13 représente les masses affectées à la classe 1. Dans la section suivante, nous présentons des résultats complémentaires et étudions l’influence des paramètres α et T . Enfin, en annexe A, nous présentons les résultats obtenus avec les données *clouds* de la base de données du projet européen ELENA.

3.6 Influences des paramètres

Soit n_a , le nombre d’individus classés dans des CCA lors de la phase de coalescence. Pour tous les résultats obtenus ci-après, le facteur de flou vaut $m = 1.5$. La Table 3.6 montre qu’il n’y a pas de variations dans les performances lorsqu’on change la valeur de T , sauf dans le cas où T est faible (rejet de distance important) au point d’influencer les zones d’ambiguïté. Ce cas est présenté dans la section suivante. Par contre, une variation de α provoque une modification des performances (cf. section 3.6.2).

3.6.1 Influence du taux de rejet de distance

Le taux de rejet de distance peut influencer les classes d’ambiguïtés lorsque sa valeur est trop faible (les frontières de décisions du rejet de distance passent alors dans les zones d’ambiguïtés). Ceci est mis en évidence dans la figure 3.14 pour $T = 3.5$. Ainsi, n_a qui a pour valeur 47 (pour $\alpha = 0.1$) est réduit à $n_a = 45$ lorsqu’un taux de rejet de distance important est demandé (cf. la dernière colonne de la Table 3.6). A l’inverse, si le taux de rejet de distance demandé est faible (T plus important), alors T n’affecte plus n_a mais uniquement le nombre d’individus rejetés (cf. figure 3.15).

| | Bayes | 3C (c=2) | 3C (c=3) |
|-----------|-------|----------|----------|
| Succès | 88.6 | 97.1 | 96.7 |
| Confusion | 11.4 | 2.9 | 3.3 |

TAB. 3.5 – Comparaison avec le classifieur Bayésien : le mauvais résultat du classifieur Bayésien vient du fait que la deuxième classe est bimodale

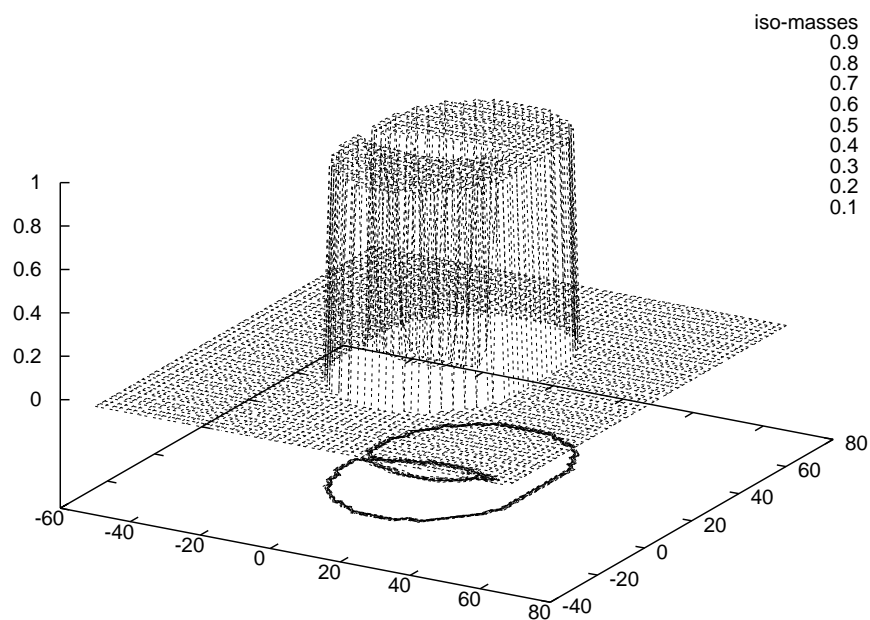


FIG. 3.11 – Masses affectées à la classe 2, avec rejet de distance ($T = 5$), d'où la limitation dans un cercle dont le rayon dépend du taux de rejet de distance choisi.

| α | | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.1 |
|------------|-------|-------|-------|-------|-------|-------|-------|
| $T = 3.5$ | % | 93.81 | 97.14 | 99.05 | 98.09 | 97.14 | 97.14 |
| | n_a | 124 | 110 | 103 | 103 | 98 | 45 |
| $T \geq 5$ | % | 93.81 | 97.14 | 99.05 | 98.09 | 97.14 | 97.14 |
| | n_a | 124 | 110 | 103 | 103 | 98 | 47 |

TAB. 3.6 – Influences de T et de α sur le taux de reconnaissance (en %) de l'ensemble de test et sur n_a .

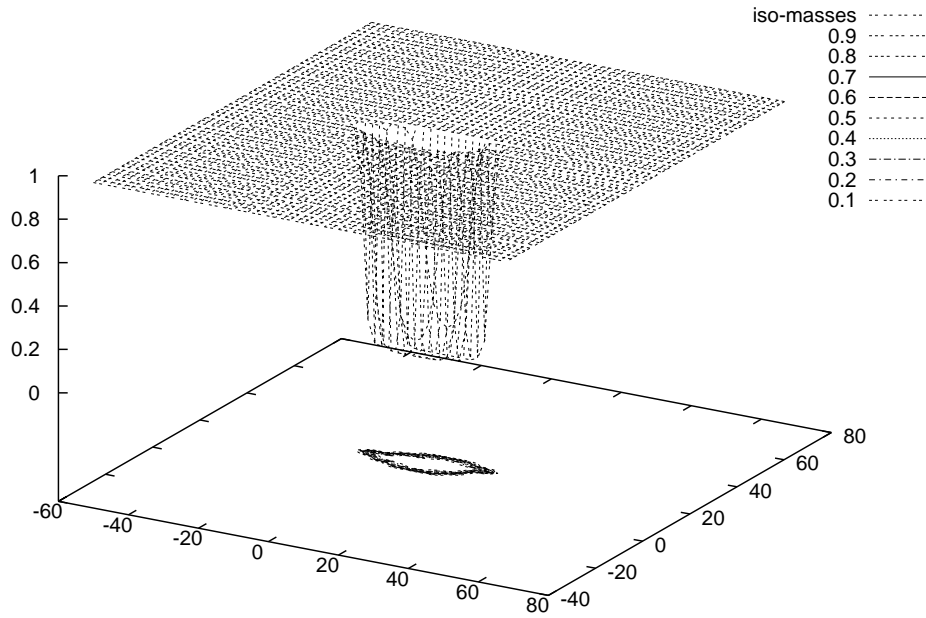


FIG. 3.12 – Masses affectées à la classe 2, sans rejet de distance ($T \rightarrow \infty$). On voit apparaître les frontières entre les deux classes (iso-masses).

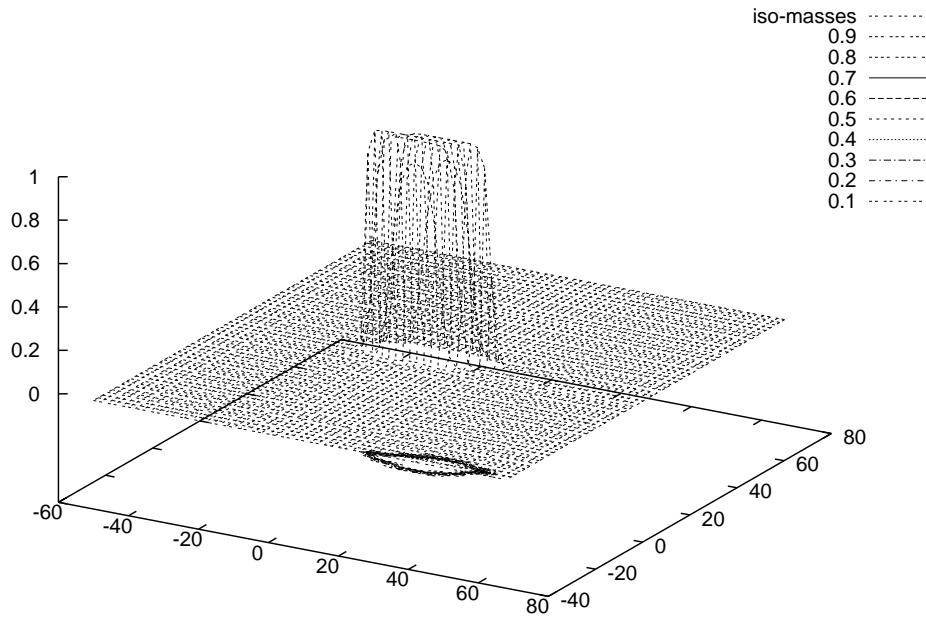


FIG. 3.13 – Masses affectées à la classe 1, avec ou sans rejet de distance.

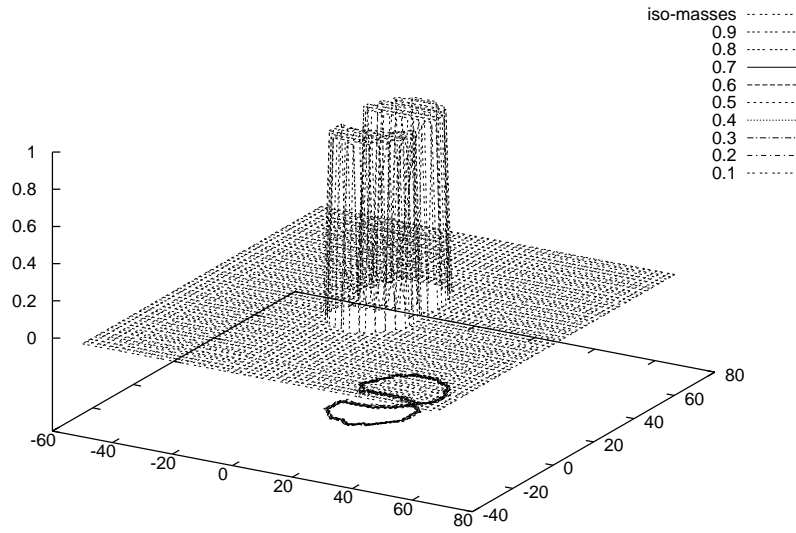


FIG. 3.14 – Influence de T sur les frontières de décisions de la classe 2 pour $T = 3.5$ (et $\alpha = 0.1$).

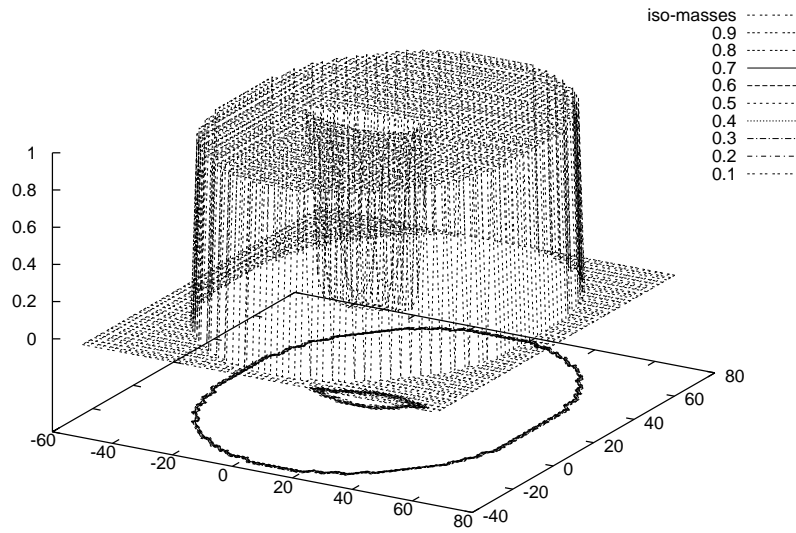


FIG. 3.15 – Influence de T sur les frontières de décisions de la classe 2 pour $T = 7$ (et $\alpha = 0.1$).

3.6.2 Influence du taux de rejet d'ambiguïté

Choix du taux α optimal

La Table ci-dessus révèle également que le meilleur résultat est obtenu pour des valeurs de n_a proches de la moitié du nombre d'individus total du nuage, c'est-à-dire, pour une valeur de α proche de $\alpha = 0.05$. Le cas idéal est donc d'avoir un α qui permet de catégoriser la moitié des individus dans les différents CCA. Expérimentalement, le meilleur résultat est obtenu pour $\alpha = 0.048$ pour lequel on obtient $n_a = 104$, un taux de reconnaissance de 99,52 % (un individu confondu). La figure 3.16 met en évidence les iso-masses concernant la classe 2, tandis que la figure 3.17 met en évidence les masses affectées à la classe 1.

Remarque : C'est pour $\alpha = 0.046$ que l'algorithme de coalescence classe la moitié de la population (105 individus sur 210) dans le classe de rejet d'ambiguïté. Pourtant, cette valeur conduit à un taux de reconnaissance de 99.05 %, ce qui inférieur au taux obtenu pour $\alpha = 0.048$ (pour lequel $n_a = 104$). Ceci est du au fait que dans l'ensemble de test, les formes des deux classes sont légèrement différentes de celles de l'ensemble d'apprentissage.

3.6.3 Ambiguïté élevée

Pour un rejet d'ambiguïté important (α plus faible) l'ensemble d'apprentissage des classifieurs Bayesiens inclut plus d'individus, jusqu'à inclure la totalité de l'ensemble d'apprentissage de départ. La conséquence est que ces classifieurs deviennent identiques au fur et à mesure que α diminue ; on finit par combiner les résultats de deux Bayesiens identiques, d'où l'obtention de résultats similaires à un seul classifieur Bayesien. Les figures 3.18 et 3.19 montrent ce phénomène, respectivement sur la classe 2 et la classe 1. À l'inverse, si α est trop grand (ambiguïté faible), certains ensembles d'apprentissage risquent de ne pas inclure d'individus (ou en nombre suffisant) représentant les deux classes, la phase d'apprentissage ne peut alors se dérouler correctement.

3.6.4 Frontières de décisions

Les deux régions correspondants aux deux classes sont mis en évidence dans la figure 3.20. Pour un α faible, ces régions ressemblent de plus en plus à des régions obtenues avec un classifieur Bayesien, c'est-à-dire dans notre cas, des ellipsoïdes (cf. figure 3.21).

3.6.5 Autres types de combinaisons

Fusion Bayésienne

Nous avons effectué une fusion Bayésienne (voir page 42) qui dans le cas optimal ci-dessus a permis d'avoir un taux de reconnaissance de 99.52 % (comme dans le cas optimal).

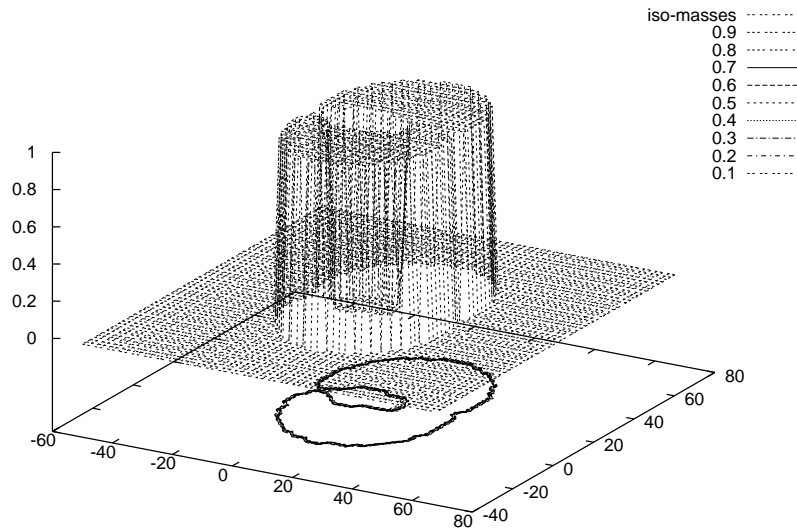


FIG. 3.16 – Iso-masses de la classe 2 pour $\alpha = 0.048$ ($T = 5$).

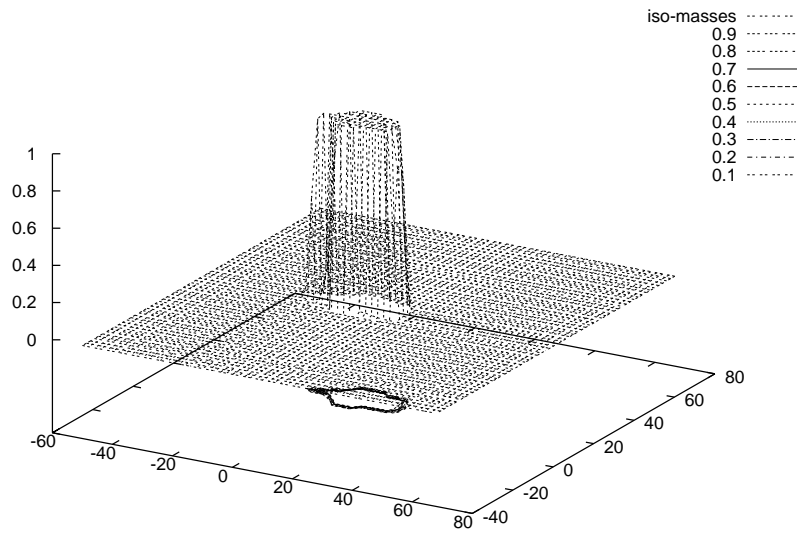


FIG. 3.17 – Iso-masses de la classe 1 pour $\alpha = 0.048$.

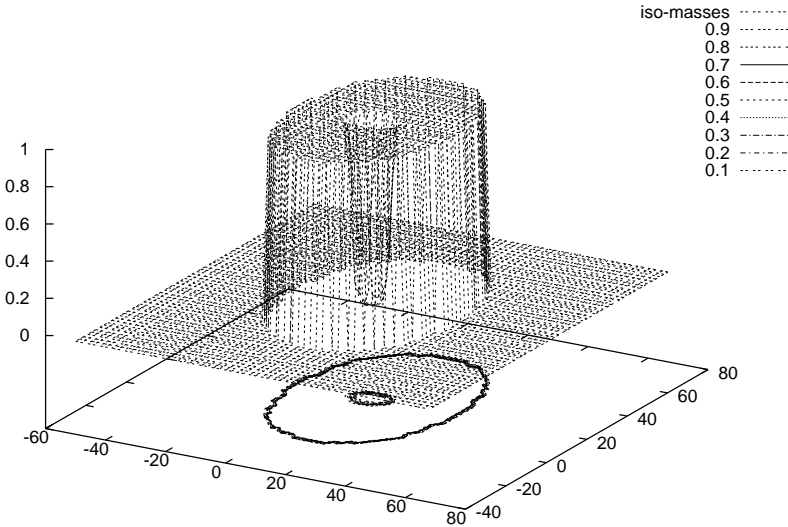


FIG. 3.18 – Iso-masses de la classe 2 pour $\alpha = 0.03$ ($T = 5$).

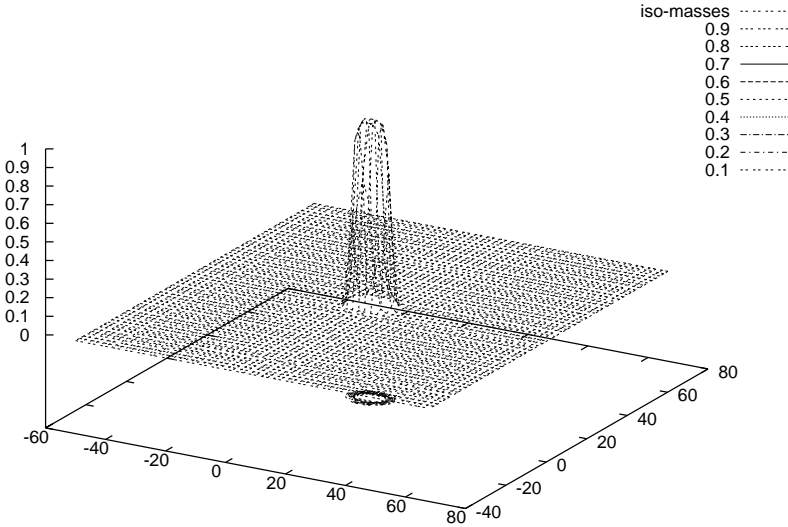
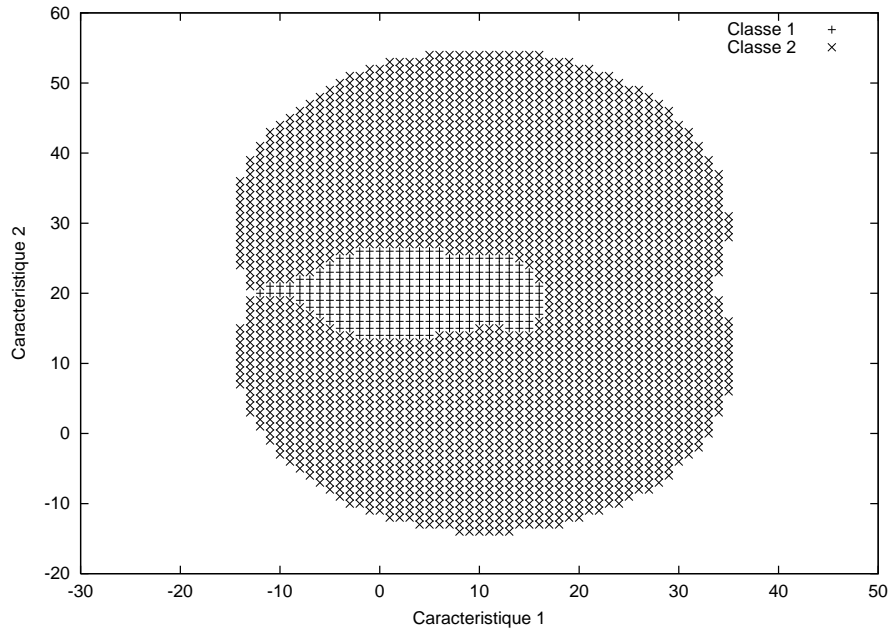
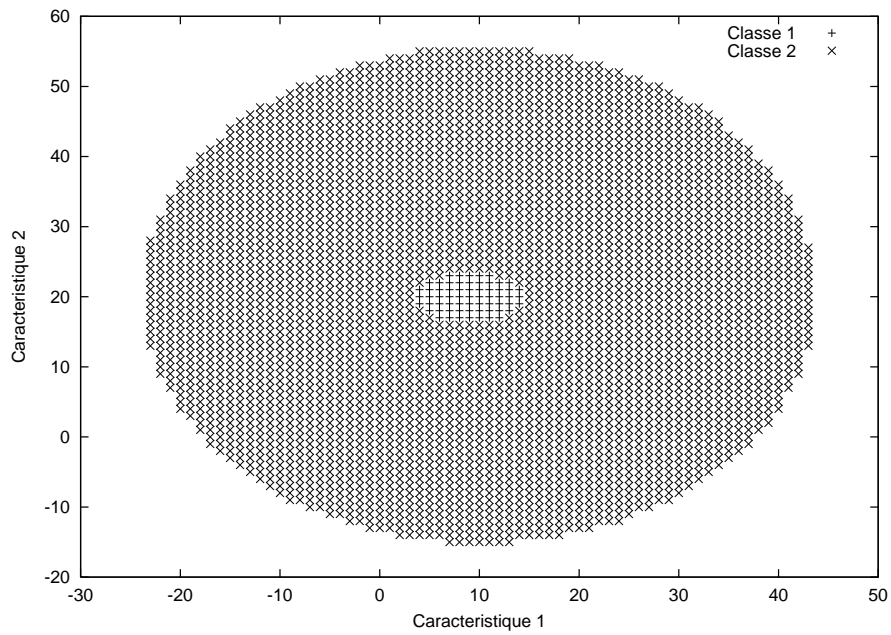


FIG. 3.19 – Iso-masses de la classe 1 pour $\alpha = 0.03$.

FIG. 3.20 – Les régions obtenues pour $\alpha = 0.048$ et $T = 5$.FIG. 3.21 – Les régions obtenues pour $\alpha = 0.02$ et $T = 5$.

Fusion avec le premier modèle défini par Appriou

Appriou a étudié les modèles de fusion permettant de respecter les trois propriétés fondamentales suivantes :

1. cohérence avec l'approche Bayésienne,
2. séparabilité de l'évaluation des hypothèses C_i ,
3. cohérence avec l'association probabiliste des sources.

Pour le premier des deux modèles obtenus, l'auteur propose de rechercher une modélisation des probabilités conditionnelles en masses pour chaque source (dans notre cas : les sorties d'un classifieur), en prenant en compte des coefficients de confiance q_{ik} (coefficient de confiance affecté à la classe i de la source k). Nous avons déterminé ces derniers en fonction des occurrences des deux classes réelles dans les ensembles d'apprentissage de chaque classifieur. Ainsi, nous avons $m_k(C_i) = q_{ik}p(x_k/C_i)$. Dans les mêmes conditions que ci-dessus, l'approche proposée par Appriou a conduit également au même taux de reconnaissance.

3.7 Conclusion

Dans ce chapitre, nous montrons l'utilité du rejet d'ambiguïté dans la phase de coalescence pour constituer les ensembles d'apprentissage de classifieurs plus spécialisés. Chaque classifieur est spécialisé sur une région particulière de l'espace de représentation, ce qui lui permet de mieux décider l'attribution d'un individu à une classe (les frontières sont mieux adaptées à la population locale). Leurs ensembles d'apprentissage sont plus simples et leurs décisions sont moins entachées d'ignorance.

L'algorithme des *c*-moyennes floues (sans rejet d'ambiguïté) donne de mauvais résultats concernant l'estimation des centres de classes. Bien qu'il soit possible d'utiliser d'autres algorithmes de coalescence intégrant le rejet d'ambiguïté, l'algorithme utilisé dans cette étude (Fc+2M) présente l'avantage de calculer des centres qui sont plus proches des centres physiques des classes de coalescence (les individus ambigus interviennent moins dans ce calcul). Cet avantage est important, car ces centres sont utilisés dans la première étape de classement, lors de la sélection des classifieurs les mieux adaptés.

Notre *Système de Classifieurs* est particulièrement intéressant pour gérer les classes multimodales et les classes qui se chevauchent (voir annexe A). Par conséquent, il est adapté au traitement statique des classes dynamiques, telles que nous les avons présentées dans l'approche 1 du chapitre précédent (page 69). Nos résultats montrent que, dans le cas des classes non-gaussiennes, un meilleur classifieur peut être obtenu (par rapport à un seul classifieur global) en combinant, par la règle de Dempster, les sorties des classifieurs les mieux adaptés. Il serait intéressant de comparer ces résultats avec ceux d'autres systèmes de classifieurs.

Si les classifieurs sont de différents types (certains permettant même le rejet d'ambiguïté), la combinaison devient plus complexe. Une des améliorations de notre algorithme peut venir de l'utilisation des modèles de fusion proposés par Appriou [4] et Smets [97]. En prenant en compte également le rejet de distance, notre approche permettra de traiter l'évolution des classes ainsi que l'apparition de nouvelles classes.

Chapitre 4

Une planification temporelle de systèmes de classifieurs

Lorsque les classes sont évolutives (dynamiques) et qu'il n'est pas possible de transformer un problème de RdF dynamique en un problème de RdF statique, la sélection dynamique de classifieurs semble être une solution pertinente. Ainsi, nous proposons, à travers ce chapitre, un système de RdF adapté à la reconnaissance de classes dynamiques. Les réseaux de Petri flous sont utilisés pour modéliser les changements d'états d'un tel système. Le caractère flou de ces changements permet d'anticiper sur le ou les états du système de RdF les mieux adaptés à l'instant considéré. Les performances et les limites de fonctionnement sont également présentées.

Sommaire

| | | |
|------------|--|------------|
| 4.1 | Introduction | 107 |
| 4.2 | Problématique | 107 |
| 4.3 | Représentation de connaissances dynamiques par des réseaux de Petri flous | 109 |
| 4.3.1 | RPF pour le raisonnement | 110 |
| 4.3.2 | RPF pour connaissances dynamiques | 111 |
| 4.3.3 | Exemple de RPF | 112 |
| 4.4 | Synthèse d'un Réseau de Petri Flou | 113 |
| 4.4.1 | Aspect dynamique de l'apprentissage | 113 |
| 4.4.2 | Algorithme de construction du système de RdF | 115 |
| 4.5 | Aspect statique de l'apprentissage | 117 |
| 4.6 | Fonctionnement du système de RdF | 117 |
| 4.7 | Validation | 117 |
| 4.7.1 | Données synthétiques | 117 |
| 4.7.2 | Apprentissage | 119 |
| 4.7.3 | Classement | 124 |
| 4.7.4 | Variation du nombre de positions dans une période. | 124 |
| 4.7.5 | Variation du seuil de confusion tolérée | 124 |
| 4.7.6 | Variation de l'écart-type des distributions | 128 |
| 4.7.7 | Autre type de trajectoire | 128 |
| 4.7.8 | Autres positions initiales | 132 |
| 4.8 | Conclusion | 137 |

4.1 Introduction

La RdF concernant les objets naturels évolutifs fait intervenir des classes dynamiques. La difficulté principale vient de la modélisation de ce type de classes et/ou de la modélisation de la prédiction de ce type de classes. Ce problème peut être contourné en rendant dynamique le système de RdF qui peut alors entrer dans différents états en fonction de l'évolution des classes (c'est l'approche 2 du chapitre 2, page 69). Ainsi, c'est le système de RdF qui change ou modifie son comportement en fonction du temps.

Les trajectoires de ces classes peuvent s'intersecter. Dans ce chapitre, nous traitons les cas où ces intersections sont uniquement spatiales. Face à la difficulté de réaliser un système de RdF général, permettant de traiter tous les problèmes de ce type, nous nous limitons dans cette étude aux problèmes où les classes subissent des évolutions périodiques.

Pour cela, nous proposons un système de reconnaissance de classes dynamiques fondé sur deux types d'apprentissage. L'aspect statique de l'apprentissage est assuré par des classifieurs ou des systèmes de classifieurs. Dans les travaux qui suivent, ces classifieurs sont obtenus en utilisant les résultats d'algorithmes de type *c*-moyennes floues en classement (une itération supplémentaire avec les individus à classer). Quant à l'aspect dynamique, il est traduit par l'apprentissage de la planification des différents états par un réseau de Petri flou¹⁸. Le réseau de Petri obtenu est un réseau cyclique. Ainsi, toute transition possède une seule place d'entrée et une seule place de sortie. Dans la phase de reconnaissance, la particularité du RPF proposé est qu'il doit être initialisé à l'instant considéré. Cette initialisation se fait en fonction des fonctions d'appartenance des jetons (construites lors de la phase d'apprentissage).

4.2 Problématique

Lorsque, pour une application donnée, les caractéristiques des objets à reconnaître ne dépendent pas du temps, les classes associées aux différents objets sont dites statiques. Les méthodes de RdF telles que les *c*-moyennes, les *k* plus proches voisins et bien d'autres sont alors bien adaptées à la reconnaissance de ce type d'objets.

Cependant, de nombreuses applications du monde réel font apparaître un glissement des classes dans l'espace de représentation (usure des machines-outils, dérive des références des instruments de mesure, évolution de la température ambiante, etc.). De même, certaines applications sont, par nature, dépendantes du temps. Ainsi, les objets naturels sont-ils souvent tributaires d'une évolution (la composante tendance d'une classe dynamique, qui correspond, dans ce cas, au vieillissement) elle-même assujettie à des variations périodiques (saisonniers ou journaliers).

¹⁸Ce chapitre a fait l'objet d'une publication dans une revue internationale [42] et d'autres publications [41] [43].

Les modifications de l'appareil qui observe un phénomène et l'évolution du phénomène lui-même nécessitent que les outils de RdF suivent ces variations. En réalité, du fait que la représentation du temps est très liée à son utilisation (choix de la granularité de l'unité de temps), les méthodes telles que la reconnaissance de séquences temporelles (Rombaut et al. [89]), dans lesquelles le temps sous-jacent est celui des séries d'états, sont bien adaptées.

Le problème étudié dans ce chapitre est la détection puis la reconnaissance des différents états d'un système de reconnaissance pour les applications où les caractéristiques subissent des évolutions périodiques (par exemple, saisonnières). Les classes ayant de telles caractéristiques sont des classes dynamiques appelées aussi classes évolutives. L'évolution des caractéristiques de ces classes est difficilement prédictible et il est communément admis que certaines trajectoires du monde réel ne peuvent être prédites fidèlement par les outils mathématiques modernes. Par conséquent, nous proposons de modéliser un système de reconnaissance fondé non pas sur une modélisation fine, mais au contraire, sur une généralisation de la trajectoire en la décrivant comme une liste finie d'états. Dans cette étude, nous ne traitons pas le problème des "changements brusques", auquel cas les classes peuvent être considérées comme étant statiques et multimodales.

Les réseaux de Petri sont des modèles de représentation bien adaptés à la modélisation des systèmes dynamiques intégrant des composants parallèles et asynchrones. Lorsque l'événement du système dynamique est discret, le système est dit Système Dynamique à Événement Discret (SDED) (Cardoso et al. [14]). Les réseaux de Petri sont couramment utilisés pour modéliser un SDED, chaque état du système étant représenté par un marquage du réseau associé. Cette modélisation présente pourtant l'inconvénient de considérer que le passage d'un état à un autre est un événement discret. Or, dans notre cas, ce passage est continu car les évolutions des classes traitées dans cette étude sont progressives.

En combinant un tel modèle avec la théorie des ensembles flous, la prise en compte des informations exogènes (connaissances de l'expert, connaissances a priori) et de l'incomplétude de l'information est explicite. En effet, on peut se situer plus ou moins de façon certaine dans un état donné (jeton flou) ou être plus ou moins certain qu'un événement se soit produit (transition floue). Ce type de réseau est connu, dans la littérature, sous le nom de réseau de Petri flou (RPF) et a été proposé pour la première fois par Looney [70]. Une telle approche convient très bien pour le cas où un état du système modélisé correspond à un marquage du RPF associé. Or, un système de reconnaissance dynamique nécessite un RPF dans lequel à un état du système correspond une place (voir la définition de ce mot dans la section suivante) du RPF, le changement d'état se décidant par un système à base de règles floues. Chen et al.[16] proposent également un modèle de RPF (présenté dans la section suivante), mais les auteurs remarquent que leur modèle ne peut pas être utilisé pour la représentation de connaissance temporelle (Jong et al. [62]). D'autre part, il a été montré que le caractère flou du marquage permet de bien décrire le changement d'état ou de la connaissance (Valette et Künzle [106]).

Par ailleurs, pour un problème dynamique tel que l'évolution périodique du système de reconnaissance, il serait utile de concevoir un système à base de connaissances qui

pourrait, comme le raisonnement humain, s'adapter à la dynamique des connaissances. La phase d'apprentissage d'un tel système consiste notamment à extraire les règles de production floues (Chen et al. [16]) d'un système à base de règles. Nous modélisons ces règles de production floues par un réseau de Petri flou. La représentation du système de reconnaissance par un RPF doit nous permettre d'effectuer une sélection dynamique de classifieurs (ou de *Systèmes de Classifieurs*). Par conséquent, notre travail se place au niveau décisionnel. Au niveau conceptuel (conception, spécification et vérification de composants logiciels), l'application de la théorie des réseaux de Petri a été traitée notamment par Estrailier et Girault [28].

Il est nécessaire de préciser que notre étude ne concerne pas l'analyse des propriétés du réseau, mais plutôt la modélisation d'un système de reconnaissance dynamique par un RPF. Ainsi, le RPF constitue une représentation du système qui évolue dans le temps. Pour une introduction aux propriétés et à l'analyse structurelle des Réseaux de Petri, le lecteur dispose notamment des travaux de Valette et al.[107] et de Murata [79].

4.3 Représentation de connaissances dynamiques par des réseaux de Petri flous

Dans notre approche (que nous nommons RCD-RPF), les propositions se réfèrent aux différents états du système de reconnaissance. Ainsi, les changements d'états de ce système sont-ils analogues aux transferts de jetons dans un réseau de Petri. Il est alors aisé d'associer un p-invariant au système de reconnaissance. Un p-invariant est un ensemble de places pour lequel la charge globale en jetons reste constante indépendamment des transitions franchies. Un des avantages de la théorie des réseaux de Petri est que les p-invariants peuvent être calculés par résolution de systèmes linéaires, sans énumération de tous les marquages possibles. Ce type de calcul permet de vérifier que le système de reconnaissance propose, à tout instant, un seul état possible du système. Les différentes places de ce p-invariant représentent alors les différents états du système.

Pour un RPF, ce type d'analyse ne peut être mis à profit que si l'évolution floue du marquage reste au moins partiellement cohérente avec le réseau de Petri ordinaire sous-jacent. Dans une telle application, l'analyse structurelle précédente et la modélisation par réseau de Petri d'un système de règles de production restent complètement cohérents (Valette et al. [105]).

Nous tenons compte du temps explicitement, en définissant les dates floues associées à la fois aux transitions et aux jetons (autant de jetons qu'il y a de places, chaque place ayant un jeton flou qui lui est associé). Ainsi, le temps courant correspond à une date possible pour les franchissements des transitions et les disponibilités des jetons.

4.3.1 RPF pour le raisonnement

La définition de Chen et al. [16] d'un RPF est la suivante :

Définition 8 *Un réseau de Petri flou est un 8-uplet $RPF = (\mathcal{P}, \mathcal{T}_{rpf}, \mathcal{D}, I, O, f, \alpha, \beta)$ où :*

- $\mathcal{P} = \{P_1, P_2, \dots, P_N\}$, un ensemble fini de places,
- $\mathcal{T}_{rpf} = \{T_1, T_2, \dots, T_M\}$, un ensemble fini de transitions,
- $\mathcal{D} = \{D_1, D_2, \dots, D_N\}$, un ensemble fini de propositions,
- $\mathcal{P} \cap \mathcal{T}_{rpf} \cap \mathcal{D} = \emptyset$ (vide), $|\mathcal{P}| = |\mathcal{D}|$ (nombre de places égal au nombre de propositions),
- $I : \mathcal{T}_{rpf} \rightarrow \mathcal{P}^\infty$ est la fonction d'entrée, qui associe à toute transition un ensemble de places,
- $O : \mathcal{T}_{rpf} \rightarrow \mathcal{P}^\infty$ est la fonction de sortie, qui associe à toute transition un ensemble de places,
- $f : \mathcal{T}_{rpf} \rightarrow [0, 1]$ est la fonction d'association, qui associe à toute transition un réel compris entre 0 et 1,
- $\alpha : \mathcal{P} \rightarrow [0, 1]$ est la fonction d'association, qui associe à toute place un réel compris entre 0 et 1,
- $\beta : \mathcal{P} \rightarrow \mathcal{D}$ est une fonction d'association bijective entre places et propositions.

En utilisant les réseaux de Petri flous, la règle de production floue R_i permettant de décrire une relation floue entre deux propositions D_j et D_k s'écrit de la manière suivante :

$$R_i : \text{SI } D_j \text{ ALORS } D_k (FC = \mu_i) \quad (4.1)$$

où μ_i est la valeur du Facteur de Certitude FC (Chen et al. [16]) associé à la transition T_i . Lorsqu'au moins deux propositions sont concernées, la règle contient des connecteurs logiques "et" ou "ou". Ces règles sont appelées règles de production floues composites et peuvent être divisées en 4 types (Looney et Alfize [71]) :

- *Type 1* : SI D_{j_1} et D_{j_2} et ... et D_{j_n} ALORS D_k ($CF = \mu_i$)
- *Type 2* : SI D_j ALORS D_{k_1} et D_{k_2} et ... et D_{k_n} ($CF = \mu_i$)
- *Type 3* : SI D_{j_1} ou D_{j_2} ou ... ou D_{j_n} ALORS D_k ($CF = \mu_i$)
- *Type 4* : SI D_j ALORS D_{k_1} ou D_{k_2} ou ... ou D_{k_n} ($CF = \mu_i$)

Pour les évolutions périodiques, le comportement du système de reconnaissance doit être cyclique : sur une période, la place de départ est la même que celle de la fin de la période. Pour cette raison et bien que le RPF que nous proposons dans la section suivante convienne à toutes ces règles, nous n'utiliserons dans la suite que les règles de type R_i .

Dans (Jong et al. [62]), les auteurs remarquent que ce modèle ne peut pas être utilisé pour la représentation de connaissance temporelle. Dans leur approche, la partie condition (proposition D_j) d'une règle est consommée lorsque la règle est appliquée, de la même manière que le jeton de la place d'entrée est consommé par le franchissement de la transition. Ceci oblige à considérer qu'une proposition reste vraie uniquement si elle

n'est pas consommée. Le jeton étant flou, la décision de consommer un jeton (i.e., de franchir une transition) dépend du seuil qui est associé à sa place. Dans ce cas, on ne peut prétendre que le franchissement d'une transition soit flou. Looney [70] utilise également un seuil associé à chaque transition pour décider de transmettre le jeton flou. Il propose d'en transmettre une copie à la place de sortie, pour que la proposition associée à la place d'entrée reste vraie. Ceci se traduit par une boucle simple dans le graphe du réseau de Petri. L'inconvénient est qu'il n'apparaît pas dans la matrice d'incidence du réseau. Au lieu de se donner un choix uniquement entre les solutions "copier" ou "consommer" (type binaire : 0 ou 1), il serait plus pertinent d'alimenter correctement toute place avec une charge adéquate d'un jeton associé à cette place. Notre étude montre qu'il est possible de construire des fonctions d'appartenance de tels jetons lors de la phase d'apprentissage (voir la section 4.4.1).

4.3.2 RPF pour connaissances dynamiques

Dans l'hypothèse des évolutions périodiques déterministes, les parties antécédent et conséquence d'une règle de production floue sont uniques, car les propriétés du temps sous-jacent sont celles de \mathbb{N} (l'ensemble des entiers naturels) relativement à la relation précédente (ordre total discret) : étant donné un premier état, chaque état suivant induit un successeur unique. Ainsi, nous utilisons uniquement les règles qui sont semblables à celle décrite par R_i (4.1). Cependant, le modèle de RPF proposé ci-dessous reste valable pour toutes les autres règles. C'est le cas dans l'hypothèse des problèmes non-déterministes, dans lequel le temps peut avoir plusieurs futurs, même si pour tous les états d'une série donnée, il a un passé unique.

Nous proposons de modifier le modèle de Chen et al. [16] de manière à tenir compte explicitement du temps. Chaque jeton associé à une place est un ensemble flou dont la fonction d'appartenance qui le définit dépend du temps. Chaque transition n'est plus associée à un facteur de certitude (indépendant du temps), mais est un ensemble flou dont la fonction d'appartenance qui le définit dépend du temps. Notre RPF n'est pas asynchrone dans le sens où une transition peut être franchie à un instant inconnu (contrairement au réseau de Petri classique), mais il est asynchrone dans le sens où le RPF peut être utilisé à des instants inconnus : c'est l'application des règles qui se produit à des instants indéterminés.

Pour la phase de classement, nous proposons de réinitialiser le marquage en fonction des valeurs prises par les jetons à l'instant du classement, avant de faire appel au système de reconnaissance par l'application des règles. La définition que nous proposons pour un RPF représentant un système de reconnaissance dynamique est la suivante :

Définition 9 *Un réseau de Petri flou est un 8-uplet $RPF = (\mathcal{P}, \mathcal{T}_{rpf}, \mathcal{D}, I, O, f, \alpha, \beta)$ où :*

- $\mathcal{P} = \{P_1, P_2, \dots, P_N\}$, un ensemble fini de places,
- $\mathcal{T}_{rpf} = \{T_1, T_2, \dots, T_M\}$, un ensemble fini de transitions,
- $\mathcal{D} = \{D_1, D_2, \dots, D_N\}$, un ensemble fini de propositions,
- $\mathcal{P} \cap \mathcal{T}_{rpf} \cap \mathcal{D} = \emptyset$, $|\mathcal{P}| = |\mathcal{D}|$,

- $I : \mathcal{T}_{rpf} \rightarrow \mathcal{P}^\infty$ est la fonction d'entrée, qui associe à toute transition un ensemble de places,
- $O : \mathcal{T}_{rpf} \rightarrow \mathcal{P}^\infty$ est la fonction de sortie, qui associe à toute transition un ensemble de places,
- $f : \mathcal{T}_{rpf} \times \mathcal{T} \rightarrow [0, 1]$ est la fonction d'association, qui associe à toute transition un réel compris entre 0 et 1, à tout instant $t \in \mathcal{T}$,
- $\alpha : \mathcal{P} \times \mathcal{T} \rightarrow [0, 1]$ est la fonction d'association, qui associe à toute place un réel compris entre 0 et 1, à tout instant $t \in \mathcal{T}$,
- $\beta : \mathcal{P} \rightarrow \mathcal{D}$ est une fonction d'association bijective entre places et propositions.

où \mathcal{T} représente le temps écoulé depuis l'état initial du RPF. La règle de production floue R_i devient alors :

$$R_i : \text{SI } D_j \text{ ALORS } D_k (FC = \mu_{T_j}(t)) \quad (4.2)$$

où $\mu_{T_j}(t) = f(T_j, t)$ est la valeur du Facteur de Certitude. De même, $\mu_{P_j}(t) = f(P_j, t)$ est la "valeur" (i.e., le degré d'appartenance) du jeton de la place P_j à l'instant t .

Si $\alpha(P_j) = y_j$, $y_j \in [0, 1]$ et $\beta(P_j) = D_j$, alors le degré de vérité de la proposition D_j est y_j , i.e., le degré de vérité de la validité de l'état j du système de reconnaissance est y_j . Cette règle peut être modélisée comme dans la figure 4.1, où P_j et P_k sont respectivement la place d'entrée et la place de sortie de la transition T_j . N'utilisant que des règles de type R_i (voir section précédente), y_j et y_k ne dépendent que de leur transition d'entrée et de la place d'entrée de cette transition. Ainsi, le degré de vérité de la proposition D_k (système de reconnaissance dans l'état k) est alors :

$$y_k = \alpha(P_k) = y_j \times \mu_{T_j}(t) = 0.8 * 0.9 = 0.72$$

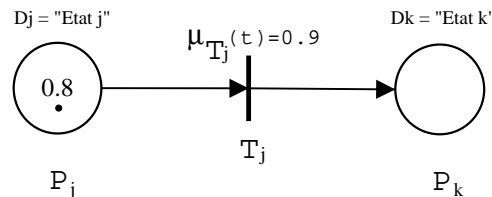


FIG. 4.1 – Représentation d'une règle de production floue par un RPF.

4.3.3 Exemple de RPF

La figure 4.2 présente un exemple de RPF (cas cyclique) pour la représentation de connaissances dynamiques, dont la définition est :

$$RPF_1 = (\mathcal{P}, \mathcal{T}_{rpf}, \mathcal{D}, I, O, f, \alpha, \beta) \text{ où :}$$

- $\mathcal{P} = \{P_1, P_2, P_3, P_4\}$,
- $\mathcal{T}_{rpf} = \{T_1, T_2, T_3, T_4\}$,
- $\mathcal{D} = \{Etat\ 1, Etat\ 2, Etat\ 3, Etat\ 4\}$,
- $I(T_1) = \{P_1\}$, $I(T_2) = \{P_2\}$, $I(T_3) = \{P_3\}$, $I(T_4) = \{P_4\}$,
- $O(T_1) = \{P_2\}$, $O(T_2) = \{P_3\}$, $O(T_3) = \{P_4\}$, $O(T_4) = \{P_1\}$,
- $f(T_1) = \mu_{T_1}(t)$, $f(T_2) = \mu_{T_2}(t)$, $f(T_3) = \mu_{T_3}(t)$, $f(T_4) = \mu_{T_4}(t)$,
(Fonctions d'initialisation de $\{T_1, T_2, T_3, T_4\}$)
- $\alpha(P_1) = \mu_{P_1}(t)$, $\alpha(P_2) = \mu_{P_2}(t)$, $\alpha(P_3) = \mu_{P_3}(t)$, $\alpha(P_4) = \mu_{P_4}(t)$,
(Fonctions d'initialisation de $\{P_1, P_2, P_3, P_4\}$)
- $\beta(P_1) = Etat\ 1$, $\beta(P_2) = Etat\ 2$, $\beta(P_3) = Etat\ 3$, $\beta(P_4) = Etat\ 4$.

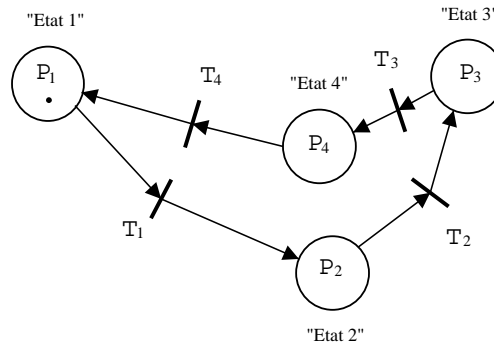


FIG. 4.2 – Représentation de connaissance dynamique par un RPF.

Pour ce réseau, les règles de production floues sont :

- $R_1 : SI\ Etat\ 1\ ALORS\ Etat\ 2 (FC = \mu_{T_1}(t))$,
- $R_2 : SI\ Etat\ 2\ ALORS\ Etat\ 3 (FC = \mu_{T_2}(t))$,
- $R_3 : SI\ Etat\ 3\ ALORS\ Etat\ 4 (FC = \mu_{T_3}(t))$,
- $R_4 : SI\ Etat\ 4\ ALORS\ Etat\ 1 (FC = \mu_{T_4}(t))$.

4.4 Synthèse d'un Réseau de Petri Flou

Un modèle de représentation de connaissances dynamiques étant proposé, les fonctions d'appartenance doivent être déterminées (i.e., apprises) lors de la phase d'apprentissage, i.e., elles doivent être apprises pour un système de reconnaissance donné.

4.4.1 Aspect dynamique de l'apprentissage

Afin de générer un réseau de Petri flou, il est nécessaire de détecter les changements d'états (les changements nécessaires de systèmes de classifieurs) et de déterminer les fonctions d'appartenance des transitions et des jetons.

Des mesures de changements fondées sur l'organisation des caractéristiques ont été proposées par Sarkar et Boyer [92]. A partir de valeurs et vecteurs propres de la matrice d'adjacences, leur méthode permet de quantifier le changement des données. Ces mesures, associées à des seuils permettent de créer un nouvel état. L'instant associé à cet événement peut servir alors au choix de la distribution des appartenances pour la transition d'entrée de la place associée au nouvel état.

Dans le cadre de la RdF, nous proposons une méthode moins coûteuse en temps de calcul, qui consiste à observer l'évolution de la matrice de confusion à partir de l'apprentissage effectué pour une configuration initiale. Une phase d'apprentissage permet de détecter les changements d'états nécessaires pour ne pas dépasser un certain seuil de confusion entre les classes. Les instants associés à ces changements servent à déterminer les dates flous pour les transitions et les jetons associés à chaque place, i.e., de construire leurs fonctions d'appartenance.

L'évolution étant périodique, à tout instant, une seule transition peut être franchie. Ceci implique que l'imprécision temporelle doit être faible, du fait que nous sommes amenés à choisir un seul état à tout instant, bien qu'il soit imaginable de prévoir le cas où l'on déciderait de l'affectation à une classe suite à la fusion des résultats de plusieurs *systèmes* de classifieurs. Un *système* de classifieurs peut être composé de plusieurs classifieurs. De tels *systèmes* s'avèrent utiles lorsque les classes sont multimodales ou lorsqu'elles se chevauchent (par exemple, l'algorithme 3C, décrit au chapitre 3). Puisque, à tout moment, le système de reconnaissance peut être, au plus, entre deux états, nous limitons cette étude au cas où seulement deux *systèmes* doivent être choisis.

L'apprentissage permet de construire ces fonctions d'appartenance en tenant compte implicitement des connaissances de l'expert. Chaque instant associé à un changement d'état correspond alors à l'instant auquel le degré d'appartenance de la transition franchie est maximal ($\mu_{T_n}(t) = 1$). Il est raisonnable de considérer que cette fonction d'appartenance doit s'annuler aux instants correspondants aux changements d'états précédent et suivant (i.e., correspondants aux instants pour lesquelles les degrés d'appartenance des transitions précédente et suivante sont maximales).

Pour les jetons flous, le raisonnement est un peu différent. En effet, entre deux changements d'états n et $n + 1$, $\mu_{P_n}(t)$ devrait avoir des degrés d'appartenance élevés et ne devrait pas être nulle à ces deux instants. Car c'est, précisément, à ces deux instants que le système de reconnaissance devra faire un choix permettant d'anticiper sur l'évolution des classes. Par conséquent, nous proposons d'annuler $\mu_{P_n}(t)$ avant le changement n et après le changement $n + 1$. En pratique, ces deux écarts sont déterminés en divisant par deux les durées de validité des états précédent et suivant. Ainsi, ces deux instants correspondent aux maxima de $\mu_{P_{n-1}}(t)$ et de $\mu_{P_{n+1}}(t)$. De plus, ce choix permet d'avoir un RPF p-invariant ($\forall t, \sum_{n=1}^N \mu_{P_n}(t) = 1$).

Dans cette étude, nous choisissons des fonctions d'appartenance triangulaires, définies ci-après. Ces deux fonctions d'appartenance sont représentées dans la figure 4.3.

Définition 10 Soit $t_{T_n} \in \mathcal{T}, \forall n \in [1, N]$, l'instant correspondant au changement de l'état n vers l'état $n + 1$. La fonction d'appartenance $\mu_{T_n}(t)$ de l'ensemble flou "la transition T_n est à peu près valide" est définie par :

$$\mu_{T_n} : \mathcal{T} \rightarrow [0, 1]$$

$$\mu_{T_n}(t) = \begin{cases} \frac{t-t_{T_{n-1}}}{t_{T_n}-t_{T_{n-1}}}, & \text{si } t_{T_{n-1}} \leq t \leq t_{T_n} \\ \frac{-t+t_{T_{n+1}}}{t_{T_{n-1}}-t_{T_n}}, & \text{si } t_{T_n} < t \leq t_{T_{n+1}} \\ 0, & \text{sinon.} \end{cases} \quad (4.3)$$

Définition 11 Soit $t_{T_n} \in \mathcal{T}, \forall n \in [1, N]$, l'instant correspondant au changement de l'état n vers l'état $n + 1$ et Δ_n , la durée entre $t_{T_n} \in \mathcal{T}$ et $t_{T_{n+1}} \in \mathcal{T}$ divisée par 2. La fonction d'appartenance $\mu_{P_n}(t)$ de l'ensemble flou "le jeton de la place P_n est à peu près valide" est définie par :

$$\mu_{P_n} : \mathcal{T} \rightarrow [0, 1]$$

$$\mu_{P_n}(t) = \begin{cases} \frac{t-(t_{T_n}-\Delta_{n-1})}{\Delta_{n-1}+\Delta_n}, & \text{si } (t_{T_n} - \Delta_{n-1}) \leq t \leq (t_{T_n} + \Delta_n) \\ \frac{-t+(t_{T_{n+1}}+\Delta_{n+1})}{\Delta_n+\Delta_{n+1}}, & \text{si } (t_{T_n} + \Delta_n) < t \leq (t_{T_{n+1}} + \Delta_{n+1}) \\ 0, & \text{sinon.} \end{cases} \quad (4.4)$$

Ainsi, l'état du réseau de Petri flou change-t-il en fonction de l'évolution des transitions floues et de l'évolution des marquages flous associés aux places. De cette manière, nous prenons explicitement le temps en compte et nous effectuons l'apprentissage dynamique. Ce dernier nous permettra de reconnaître, à tout moment, le "meilleur" système à utiliser.

Construit dans ces conditions, le RPF est vivant et ne comporte, par conséquent, pas de blocages. Il est aussi 1-borné et sauf (safe, en Anglais), par conséquent. Cette dernière propriété implique que le réseau est cohérent, i.e., $\forall t \in \mathcal{T}$ et $\forall n \in [1, N], \mu_{P_n}(t) \leq 1$. Dans le cas contraire, cela poserait un problème d'interprétation de " $\mu_{P_n}(t) > 1$ ".

4.4.2 Algorithme de construction du système de RdF

L'algorithme suivant permet de déterminer ces instants de changements d'états, instants qui serviront à construire les fonctions d'appartenance. Soit n le nombre d'états, égal au nombre de classifieurs (ou de systèmes de classifieurs). La procédure d'apprentissage génère de manière itérative des états, jusqu'à ce que le nombre d'états trouvé ne change plus et que les centres de classes calculées pour les différents états ne se déplacent plus (à un seuil ϵ près).

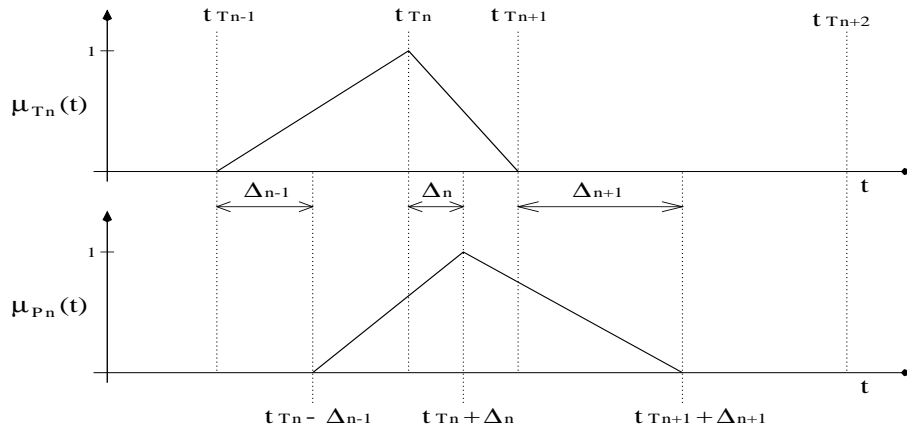


FIG. 4.3 – Construction des fonctions d'appartenance des transitions $\mu_{T_n}(t)$ et de celles des jetons $\mu_{P_n}(t)$.

Procédure Apprentissage sur première période (T)

Entrée : X^{lr} : l'ensemble d'apprentissage, T , Tolérance, ϵ

Sortie : Fonctions d'appartenance des jetons et des transitions $i \in [1, n]$, les classifieurs e_i ou systèmes SC_i $i \in [1, n]$

- 1 : Effectuer un apprentissage et un classement à $t = 0$ (classifieur e_1)
- 2 : Calculer la matrice de confusion (de référence) \mathcal{M}^0 , pour $t = 0$
- 3 : $m \leftarrow 0$; $n \leftarrow 1$
- 4 : **répéter**
- 5 : $t \leftarrow 0$
- 6 : **répéter**
- 7 : Effectuer un classement (classifieur e_n)
- 8 : Calculer la matrice de confusion \mathcal{M}_t
- 9 : **si** (Δ confusion $>$ Tolérance) **alors**
- 10 : $t_{T_n} \leftarrow t$; $n \leftarrow n + 1$
- 11 : Effectuer un apprentissage et un classement (classifieur e_n)
- 12 : Calculer la matrice de confusion \mathcal{M}^0
- 13 : **fin si**
- 14 : $t = t + \delta t$
- 15 : **jusqu'à** ($t > T$)
- 16 : $m_{ancien} \leftarrow m$; $m \leftarrow n$ ($m = n_{maximal}$ de l'itération)
- 17 : **jusqu'à** ($(m = m_{ancien})$ ET ($|Centres\ de\ classes| < \epsilon$))
- 18 : Construire les fonctions d'appartenance

4.5 Aspect statique de l'apprentissage

Les connaissances statiques (valables à un moment donné dans une période) sont apprises par les classifieurs ou les systèmes de classifieurs associés aux différents états du système de reconnaissance dynamique. Elles sont modélisées de différentes manières (centres de classes, matrices de variance-covariances, poids des entrées de neurones dans un réseau de neurones, etc.), en fonction du type de système de classifieurs.

4.6 Fonctionnement du système de RdF

Les fonctions d'appartenance des transitions T_i et celles des jetons des places P_i étant apprises lors de l'apprentissage, le degré de vérité d'un état du système (d'une proposition) est déterminé par la valeur du jeton associé à sa place (i.e., son degré d'appartenance). Les valeurs des jetons associés aux différentes places sont obtenues par l'application des règles de production floues, à l'instant considéré. En fonction de cet instant, l'individu inconnu x est classé par le ou les SCs adaptés. Dans le cas où deux SCs sont adaptés, leurs sorties sont combinées. La prise de décision est effectuée sur les résultats du SC retenu ou sur les résultats de la combinaison, en fonction du nombre de SCs adaptés à l'instant considéré.

La figure 4.4 illustre le fonctionnement global du système de RdF proposé. Un RPF est construit lors de la phase d'apprentissage en fonction de l'évolution des classes. De même, les transitions T_i et les jetons des places P_i sont appris lors de l'apprentissage. Ainsi, SC_i est spécialisé pour reconnaître les classes C_1 et C_2 pour une partie de leurs trajectoires.

4.7 Validation

Dans cette étude, chaque système de classifieurs est composé d'un seul classifieur. Au niveau du classifieur, nous utilisons également les ensembles flous. Il s'agit d'une méthode de classification non supervisée, appelée Fc+2M, présentée au chapitre précédent. Dans cette application, nous avons inhibé ses possibilités de rejet de distance et de rejet d'ambiguïté (cette méthode est alors équivalente aux c-moyennes floues définies par Bezdek [9]). Il va de soi que d'autres méthodes peuvent être employées.

4.7.1 Données synthétiques

La figure 4.5 présente un problème d'évolution périodique à deux classes dont les trajectoires sont elliptiques, avec croisement entre les deux trajectoires. La période est composée de $n_p = 256$ instants et chaque classe comporte 400 individus. Les confusions et/ou les ambiguïtés entre classes apparaissent dans certaines configurations (si on utilise un seul classifieur). Pour un problème du monde réel, les trajectoires sont en général plus compliquées, mais peuvent présenter les mêmes caractéristiques. La configuration des classes à $t = 196$, instant où les deux classes sont proches l'une de l'autre, est donnée par la figure 4.6. Dans ce cas, les classes restent bien séparées, mais nous verrons plus

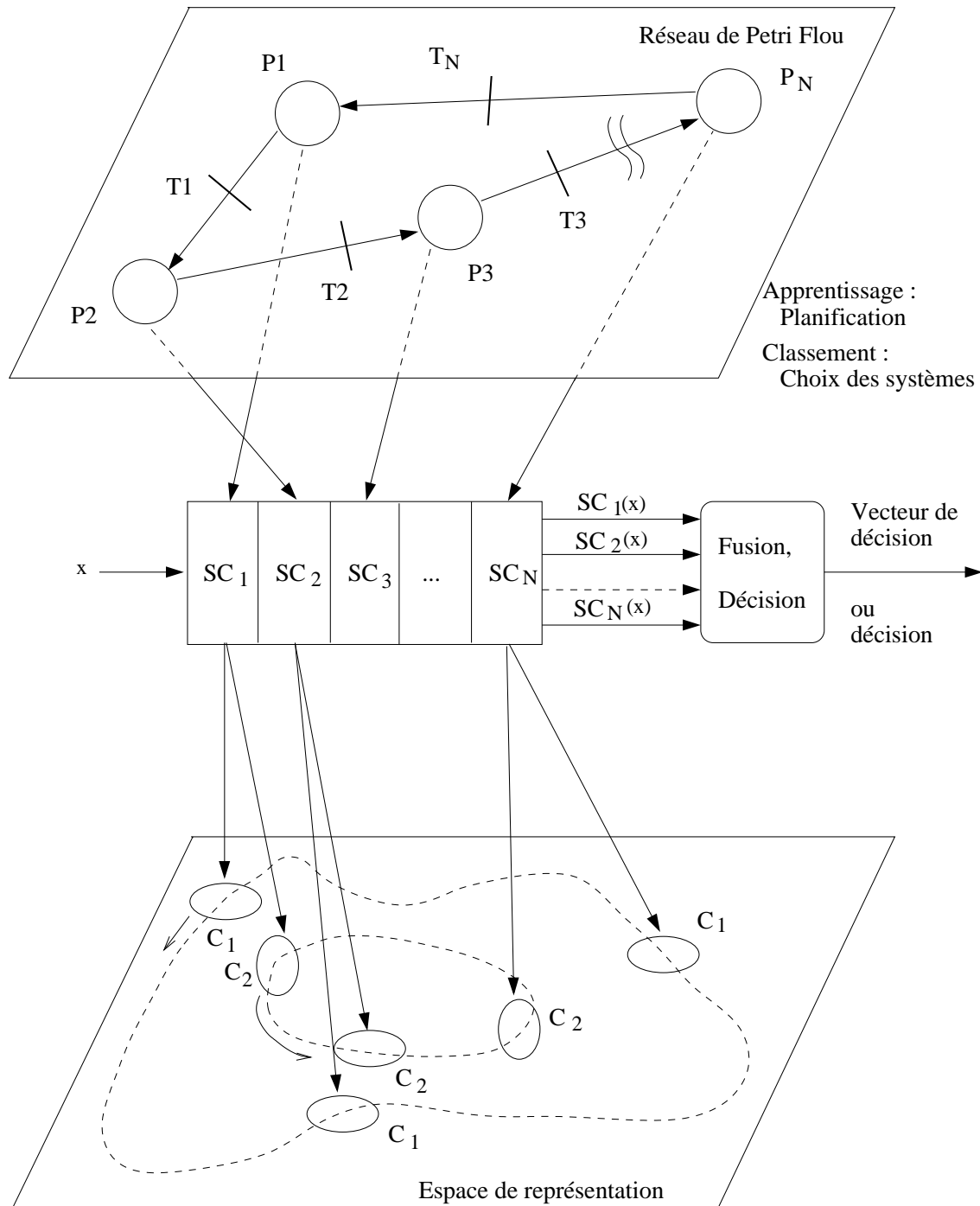


FIG. 4.4 – Fonctionnement du système de RdF proposé et principe de classement d'un individu inconnu x : chaque place (P_i) est associée à un état du système de RdF et chaque état du système est représenté par un classifieur ou un Système de Classifieurs (SC). L'individu x est classé par le SC dont l'état du système est adapté à l'instant considéré. Plusieurs SCs (les mieux adaptés) peuvent également contribuer à classer l'individu à travers une phase de fusion et de décision.

loin (en annexe B), un cas où il y a chevauchement partiel des classes. L'évolution de ces classes est définie par les n_p positions successives occupées, correspondant aux différents instants. Aux instants initial et final, les classes sont générées de la manière suivante (avec l'écart-type $\sigma = 2$) :

- ϕ_1 : 400 individus suivant une distribution Gaussienne ;

$$m_1 = \begin{pmatrix} 30 & 60 \end{pmatrix}^t, \Sigma_1 = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$$

- ϕ_2 : 400 individus suivant une distribution Gaussienne ($\Sigma_2 = \Sigma_1$) ;

$$m_2 = \begin{pmatrix} 90 & 50 \end{pmatrix}^t$$

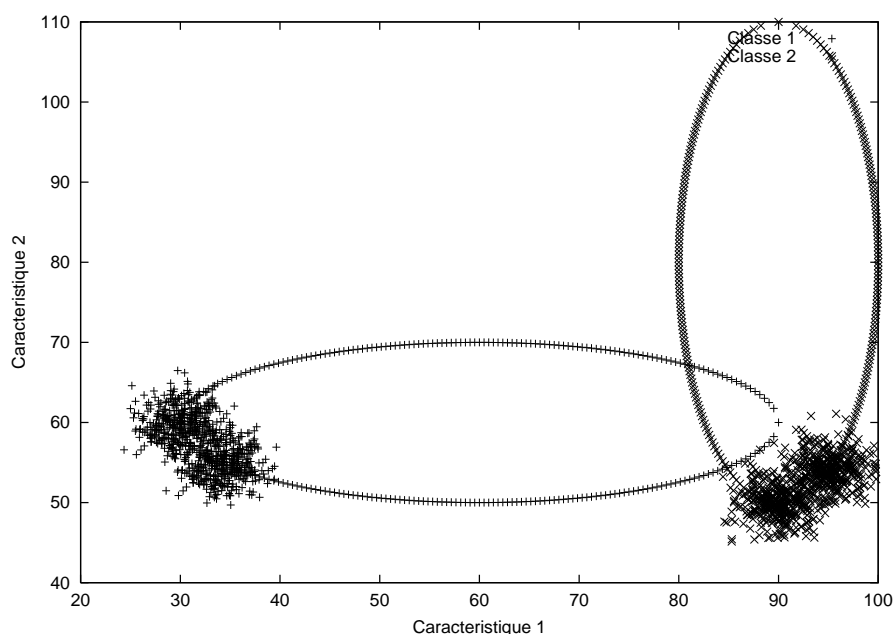


FIG. 4.5 – Exemple de problème à deux classes, dont les trajectoires sont elliptiques. La position 1 est superposée avec la position 10. Pour les autres positions, seules les centres de classes sont représentés.

4.7.2 Apprentissage

Après l'apprentissage et le classement (pour le suivi de l'évolution du taux de reconnaissance) dans la position initiale des classes, nous créons un nouvel état à chaque fois que la confusion sur une classe quelconque dépasse un seuil choisi de $cf_i = 10$ % de confusion. Ce seuil correspond au seuil de confusion toléré en classement. Lorsque ce dépassement se produit, le nouvel état est associé à un nouveau classifieur (créé) qui apprend sur les données associées à la position concernée et les reclasse (dans ce cas, le taux de reconnaissance

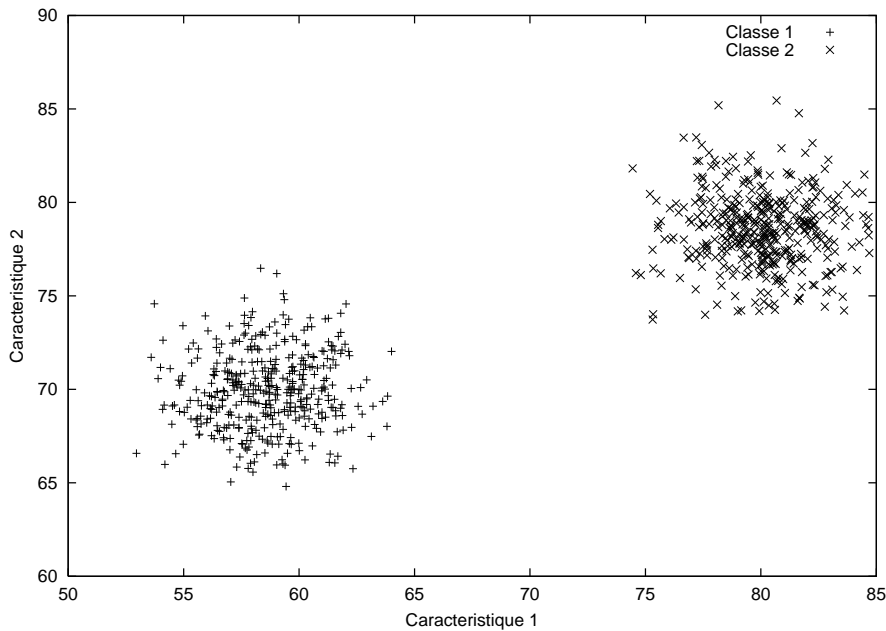


FIG. 4.6 – Configuration des deux classes dans la position 196.

du classifieur est évalué sur son ensemble d'apprentissage). Le classement s'effectue alors avec ce nouveau classifieur jusqu'au prochain dépassement du seuil de confusion. Le cycle se termine lorsque toute la période est ainsi traitée. Soit X^{lr} , l'ensemble des individus générés dans toutes les positions de la période, qui constitue l'ensemble d'apprentissage. L'ensemble de test X^t est généré dans les mêmes conditions.

Fonctionnement

Si nous réitérons le cycle d'apprentissage plusieurs fois sur une même période (période supposée fixe), au bout du 3ème cycle (dans nos expérimentations), le graphe obtenu devient stable (y compris les centres de classes associés). Dans ces conditions, nous obtenons l'évolution du taux de reconnaissance de la figure 4.7. Les centres de classes de chaque état (du système de reconnaissance) et les taux de confusion associés aux transitions sont présentés respectivement dans les Tableaux 4.1 et 4.2. Ces centres sont montrés dans la figure 4.8. Les places et transitions successives obtenues sont représentées par un réseau de Petri dans la figure 4.9.

A partir du 2ème cycle, chaque cycle débute en adoptant (en conservant) le dernier classifieur utilisé lors du cycle précédent. Ceci permet le suivi de l'évolution du taux de reconnaissance. Les figures 4.10 et 4.11, présentent respectivement les fonctions d'appartenance des transitions $\mu_{Tn}(t)$ et les fonctions d'appartenance des jetons $\mu_{Pn}(t)$ obtenues. On peut constater que ces fonctions sont non symétriques et de supports différents. Ceci montre l'adaptation de ces fonctions à l'application envisagée à travers l'adaptation de celles-ci à la dynamique des classes évolutives.

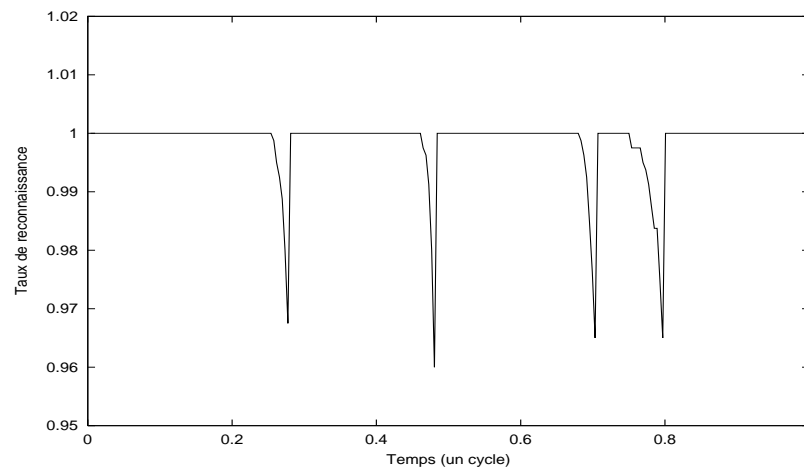


FIG. 4.7 – Evolution du taux de reconnaissance en tenant compte des changements d'états, lors de l'apprentissage. Les 4 changements d'états sont définis suites aux 4 chutes visibles sur la courbe.

| Etat (Place) | Classes | | | |
|-----------------|-------------|-------------|-------------|-------------|
| | C1 | | C2 | |
| | Paramètre 1 | Paramètre 2 | Paramètre 1 | Paramètre 2 |
| Etat 1 | 50.17 | 69.38 | 80.56 | 70.12 |
| Etat 2 | 63.29 | 49.99 | 99.94 | 83.24 |
| Etat 3 | 87.67 | 56.06 | 93.88 | 107.6 |
| Etat 4 | 65.64 | 69.75 | 80.18 | 85.59 |

TAB. 4.1 – Centres des classes obtenus pour les différents états (4) du système (après le cycle 3).

| Transitions | T_1 | T_2 | T_3 | T_4 |
|-----------------------|-------|-------|-------|-------|
| Taux de confusion (%) | 6.4 | 6.1 | 5.5 | 6.1 |

TAB. 4.2 – Taux global de confusion engendrant des changements d'états (le seuil de confusion tolérée sur une classe est de 10 %).

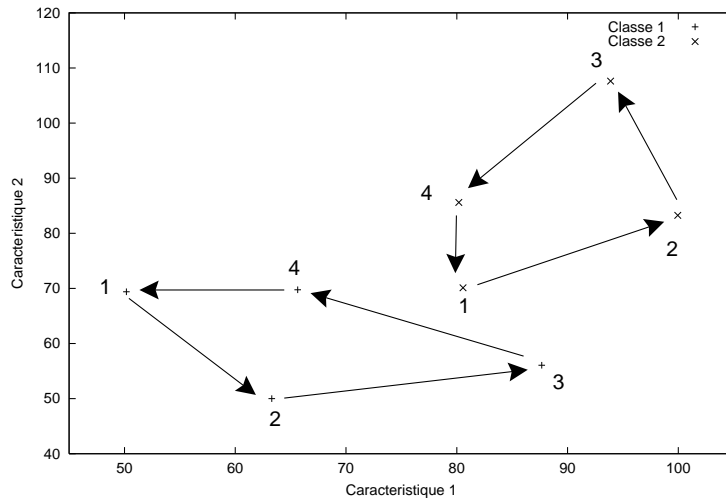


FIG. 4.8 – Résultat de l'apprentissage des différentes positions des centres de classes (pour les différents états indiqués sur la figure).

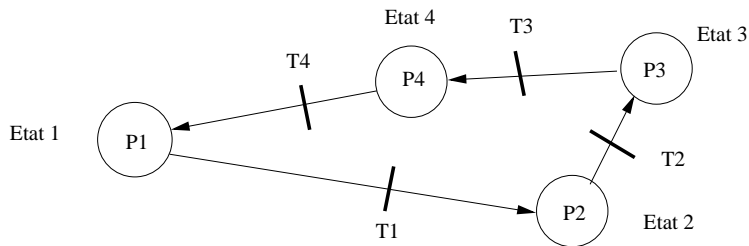


FIG. 4.9 – Places (états) successives et transitions (événements) obtenues à la fin du cycle 3.

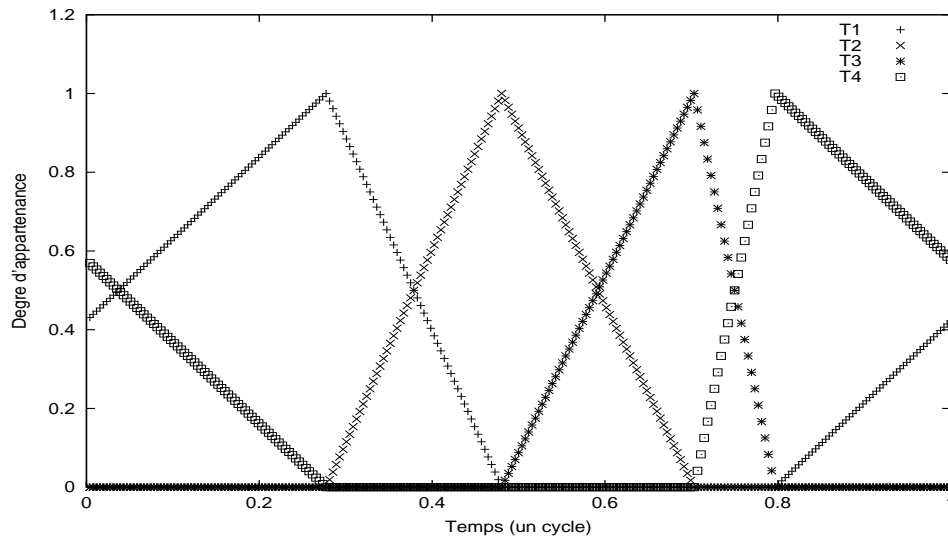


FIG. 4.10 – Fonctions d'appartenance des transitions $\mu_{T_n}(t)$ (pour les transitions T_1 à T_4). On peut constater que les fonctions générées sont non symétriques et de supports différents (donc adaptées à l'application envisagée).

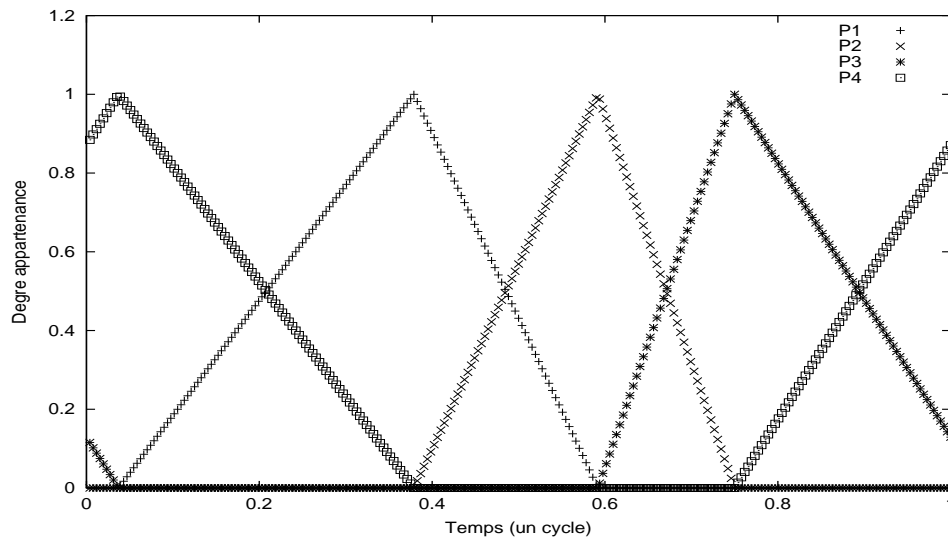


FIG. 4.11 – Fonctions d'appartenance des jetons $\mu_{P_n}(t)$ d'initialisation des places P_n . On peut constater que les fonctions générées sont non symétriques et de supports différents (donc adaptées à l'application envisagée).

Influence de la position initiale

Le système de reconnaissance obtenu (les états, le graphe, les fonctions d'appartenance et les classifieurs) diffère si le nombre de positions dans la période ou si le seuil utilisé sur le taux de confusion sont différents. Il est évident que le système obtenu diffère également selon la définition de l'instant initial de la période concernée. Il en est de même selon que l'on classe ou non les individus du premier instant, lors de la première itération (dans notre cas, ce classement est effectué), i.e., selon que la matrice de confusion de référence \mathcal{M}^0 soit évaluée sur l'ensemble d'apprentissage du classifieur ou sur l'ensemble correspondant à l'instant suivant. Dans la sous-section 4.7.8, nous étudions les changements de comportement possibles à travers le choix de deux autres instants initiaux.

4.7.3 Classement

La figure 4.12 présente les fonctions d'appartenance des marquages des différentes places P_n qui résultent de l'opération t-norme de type produit algébrique, entre $\mu_{P_{n-1}}(t)$ et $\mu_{T_{n-1}}(t)$. L'évolution du taux de reconnaissance en classement est donnée dans la figure 4.13 (avec $X^t = X^{lr}$). On peut constater que les résultats sont meilleurs que dans la phase de construction des classifieurs : le système RdF anticipe sur l'évolution des classes en adoptant le ou les classifieurs adaptés. Lorsque les degrés d'appartenance obtenus sont non nuls pour deux jetons (cf. figure 4.12), les deux classifieurs correspondants peuvent être appliqués. Leurs sorties peuvent être combinées par une somme pondérée (par les degrés d'appartenances des deux jetons). Dans ce cas, l'évolution est donnée dans la figure 4.14 (avec $X^t = X^{lr}$). Pendant le classement, afin d'étudier la robustesse du système, un bruit peut être ajouté aux centres des classes dans toutes les directions. Le bruit choisi est Gaussien, défini par $\mathcal{N}((0;0)(\sigma_{centres};\sigma_{centres};))$, avec $\sigma_{centres} = 1$. L'évolution sans fusion mais avec bruit est donnée par la figure 4.15 (avec $X^t \neq X^{lr}$). Le même cas avec fusion est présenté par la figure 4.16. Tous ces résultats montrent que le système de reconnaissance classe correctement plus de 98 % à tout moment de la période¹⁹.

4.7.4 Variation du nombre de positions dans une période.

Le tableau 4.3 montre que pour les valeurs de σ assez faibles (comparée aux distances interclasses), le nombre d'états en fonction du nombre de positions dans la période reste relativement stable. Ce même nombre augmente sensiblement lorsque la valeur de σ devient importante. La figure 4.17 montre cette augmentation pour $\sigma = 8$ (le chevauchement des classes est alors possible, cf. figure 4.18).

4.7.5 Variation du seuil de confusion tolérée

Lorsque σ est faible, le seuil cf_i a peu d'influence sur le nombre d'états générés. L'influence s'accroît avec l'augmentation de σ . Ainsi, pour $\sigma = 8$ (voir figure 4.18), on peut dépasser la dizaine d'états. La figure 4.19 montre que plus le seuil cf_i toléré est grand,

¹⁹Et même plus de 99 %, lorsqu'on autorise des combinaisons entre les résultats.

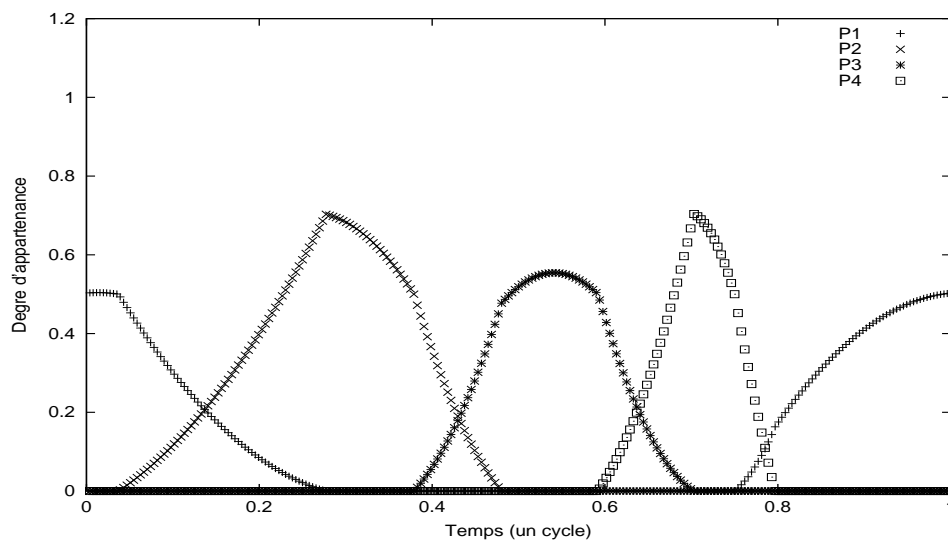


FIG. 4.12 – Fonctions d'appartenance des jetons $\mu_{P_n}(t)$ obtenues pour les différentes places P_n , résultants de l'opération t-norme entre la fonction de sa transition d'entrée et celle de la place d'entrée de cette transition.

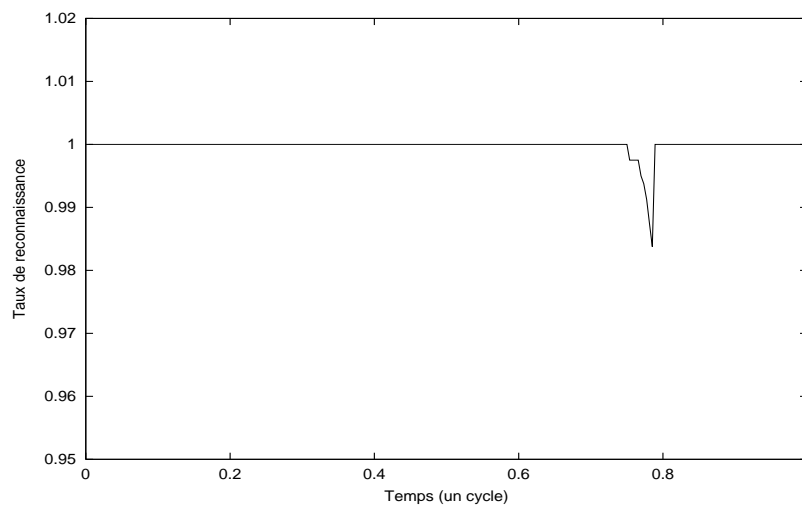


FIG. 4.13 – Evolution du taux de reconnaissance en classement (avec $X^t = X^{lr}$ et une même évolution dans le temps). Les chutes sont minimales car le système de RdF anticipe l'évolution des classes.

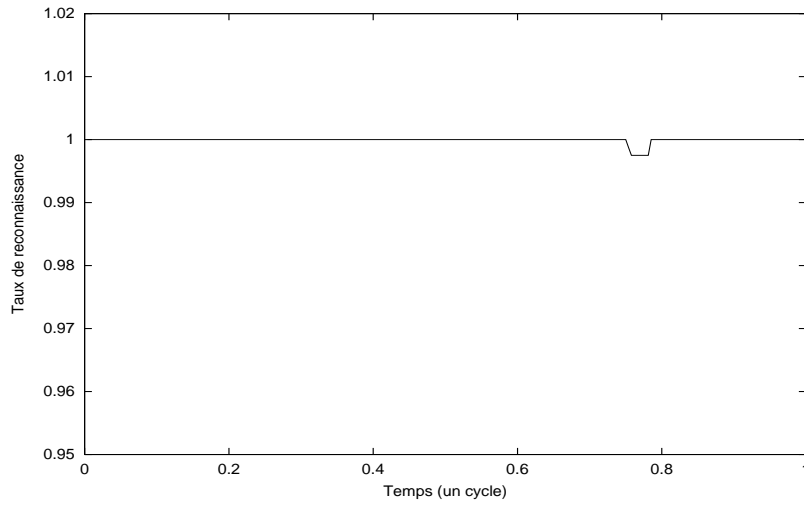


FIG. 4.14 – Evolution du taux de reconnaissance en classement avec fusion (avec $X^t = X^{lr}$). Les résultats sont améliorés par la fusion des sorties des classifieurs pour les deux états les plus valides.

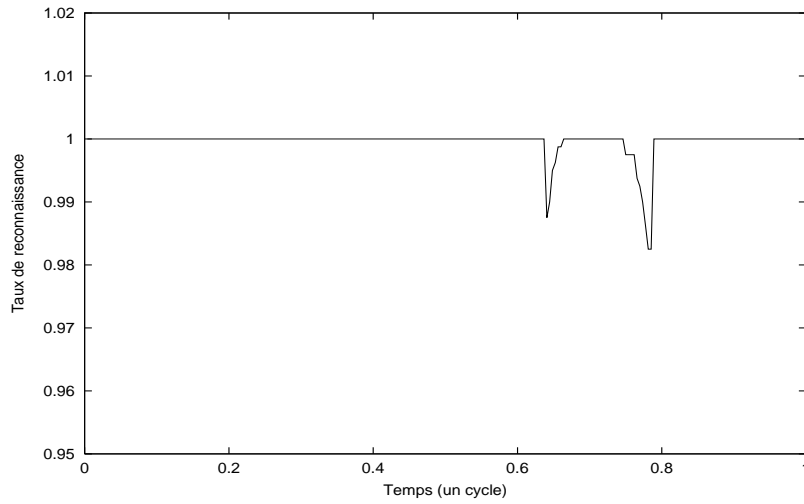


FIG. 4.15 – Evolution du taux de reconnaissance en classement avec bruit ($\sigma_{centres} = 1$, $X^t \neq X^{lr}$) et sans fusion. Peu de dégradation due au bruit : le système de RdF est robuste.

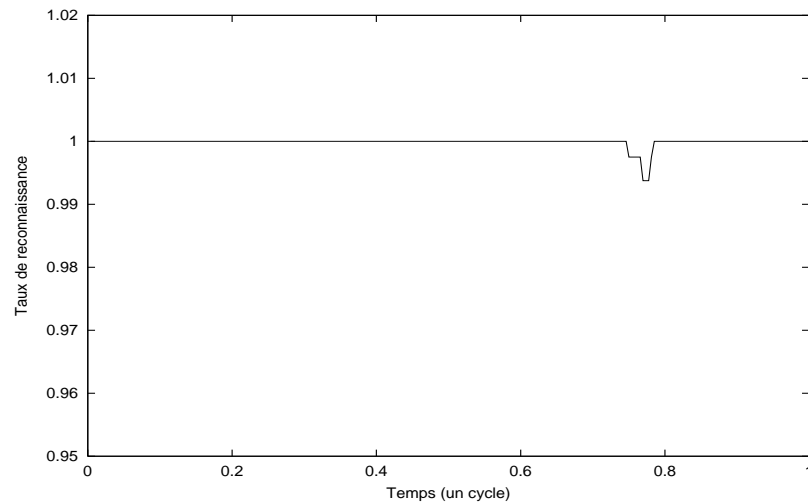


FIG. 4.16 – Evolution du taux de reconnaissance en classement avec bruit ($\sigma_{centres} = 1$, $X^t \neq X^{lr}$) et avec fusion. Le résultat est également amélioré.

| Nombre de positions | 4 | 8 | 16 | 32 | 64 | 128 | 256 | 512 |
|---------------------|---|---|----|----|----|-----|-----|-----|
| $\sigma = 2$ | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 4 |
| $\sigma = 3$ | 3 | 3 | 4 | 3 | 4 | 4 | 4 | 4 |
| $\sigma = 5$ | 3 | 4 | 4 | 4 | 4 | 6 | 6 | 6 |
| $\sigma = 8$ | 3 | 4 | 6 | 7 | 10 | 10 | 11 | 11 |

TAB. 4.3 – Nombre d'états dans une période pour différents σ (seuil de confusion cf_i toléré : 10 %).

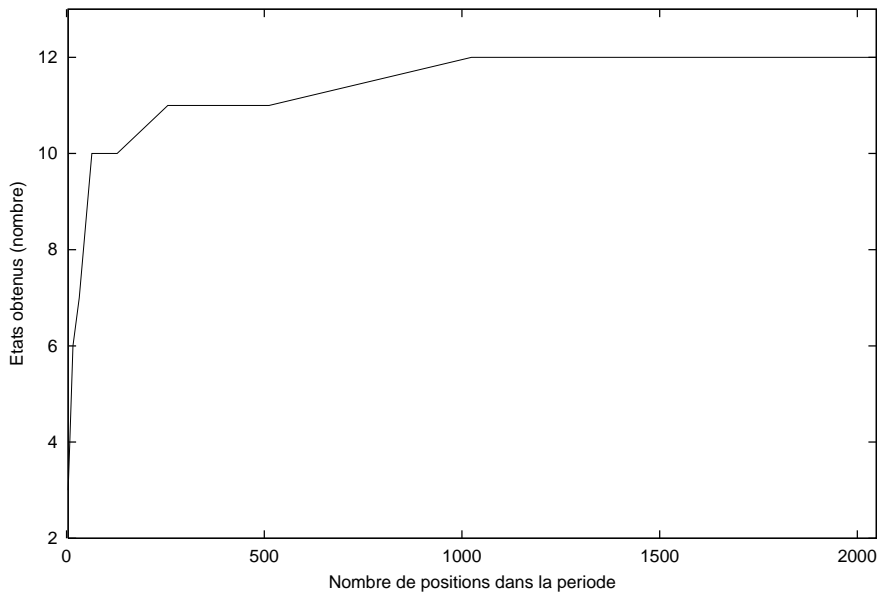


FIG. 4.17 – Influence du nombre de positions pour $\sigma = 8$ (avec $cf_i = 0.1$). Plus le chevauchement peut être important et plus le nombre d'états générés sera élevé. Dans ce cas, le nombre de positions influe peu.

plus le nombre d'états nécessaire est réduit et plus le taux de reconnaissance comporte des chutes importantes.

4.7.6 Variation de l'écart-type des distributions

Pour des valeurs élevées de l'écart-type σ des distributions, le système de reconnaissance a tendance à générer plus d'états afin de satisfaire aux contraintes (en particulier le seuil cf_i). Ceci est mis en évidence dans la figure 4.20. Jusqu'à une valeur $\sigma = 4.5$, 4 états suffisent pour effectuer la reconnaissance. Au delà de cette valeur, le nombre d'états nécessaires dépend du degré de chevauchement entre les deux classes (lorsque leurs positions s'approchent) et de l'évolution éventuelle de la forme des classes sur une intervalle de temps d'échantillonnage (temps période / n_p).

4.7.7 Autre type de trajectoire

La figure 4.21 présente le cas où en plus du croisement des trajectoires des deux classes, il y a croisement de la trajectoire de la classe 1 avec elle-même. Nous obtenons également 4 états, représentés dans la figure 4.22. Pour comparer aux résultats obtenus avec les trajectoires elliptiques (pour les deux classes), le tableau 4.4 présente les centres de classes obtenus à la troisième itération.

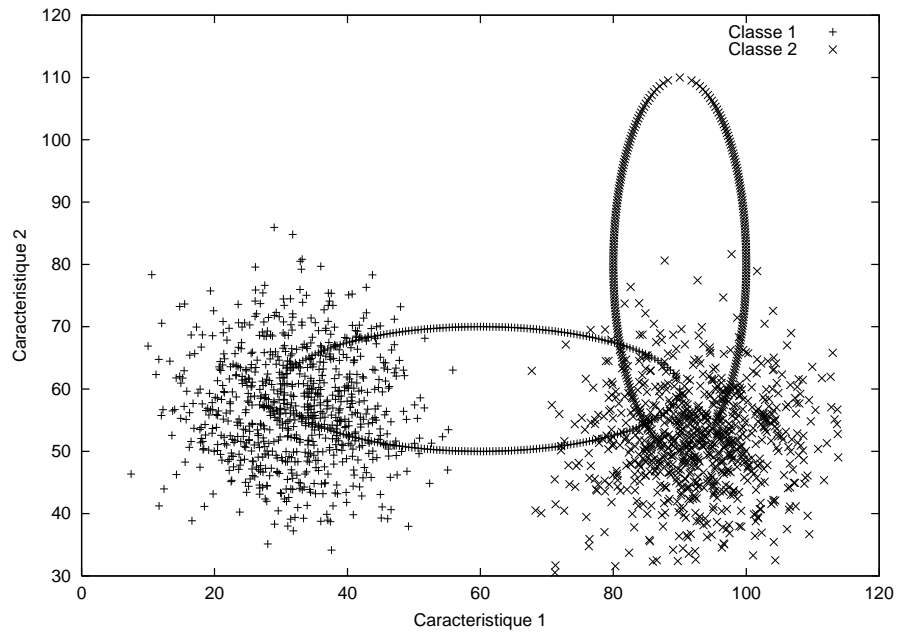


FIG. 4.18 – Les deux classes dans les positions 0 et 10 avec leurs trajectoires pour $\sigma = 8$ (et $n_p = 256$ positions).

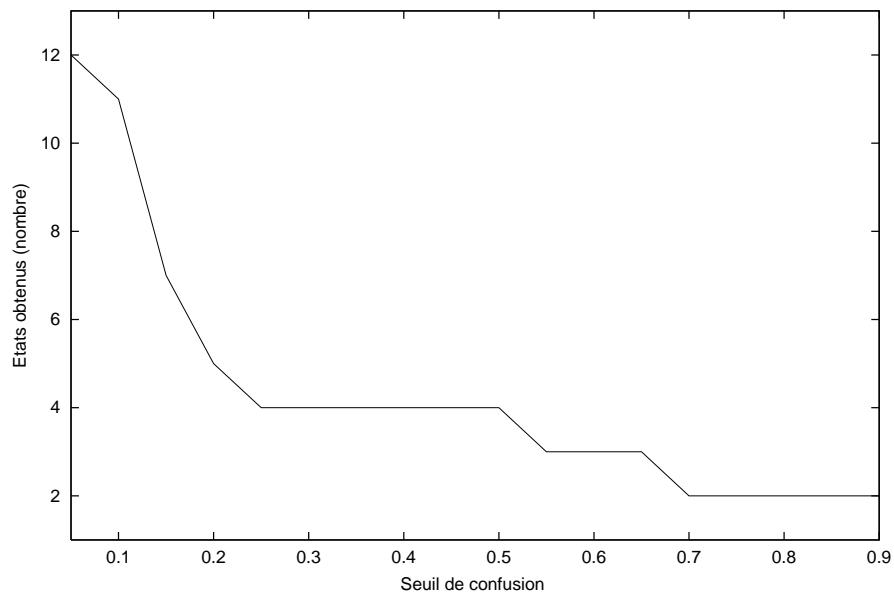


FIG. 4.19 – Influence du seuil de confusion tolérée (pour $\sigma = 8$ et $n_p = 256$ positions).

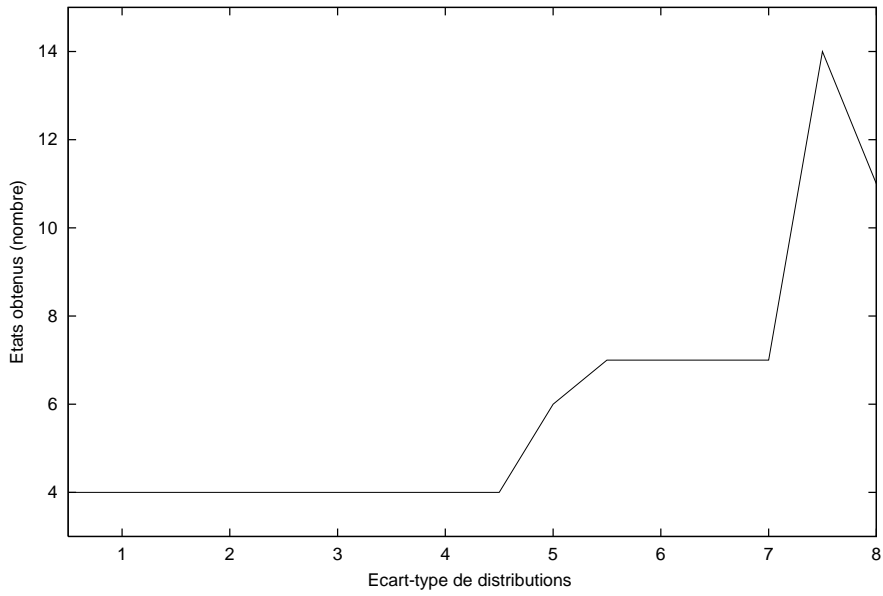


FIG. 4.20 – Influence de l'écart-type des distributions des deux classes (pour $cf_i = 0.1$ et $n_p = 256$ positions).

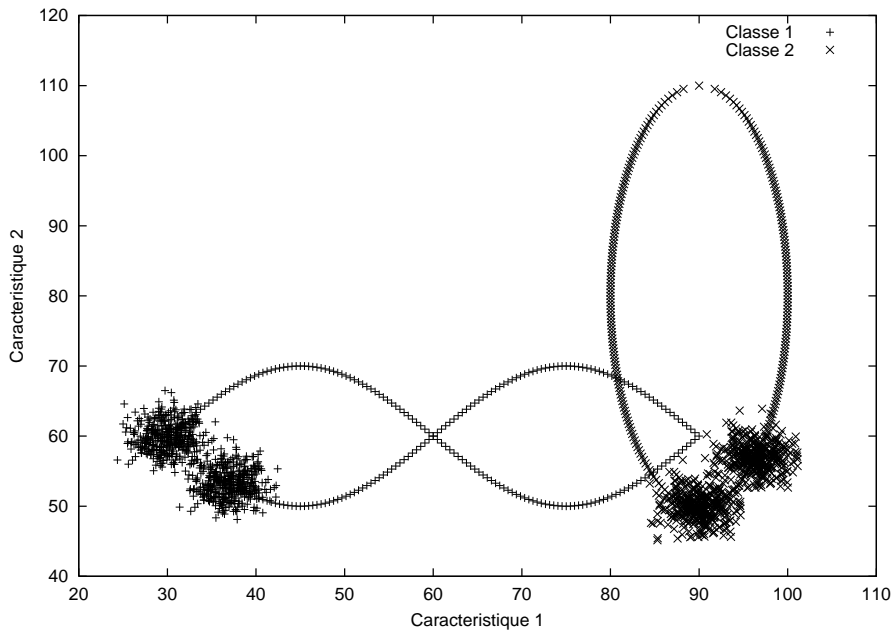


FIG. 4.21 – Les deux classes dans les positions 0 et 16 avec leurs trajectoires (pour $\sigma = 2$ et $n_p = 256$ positions).

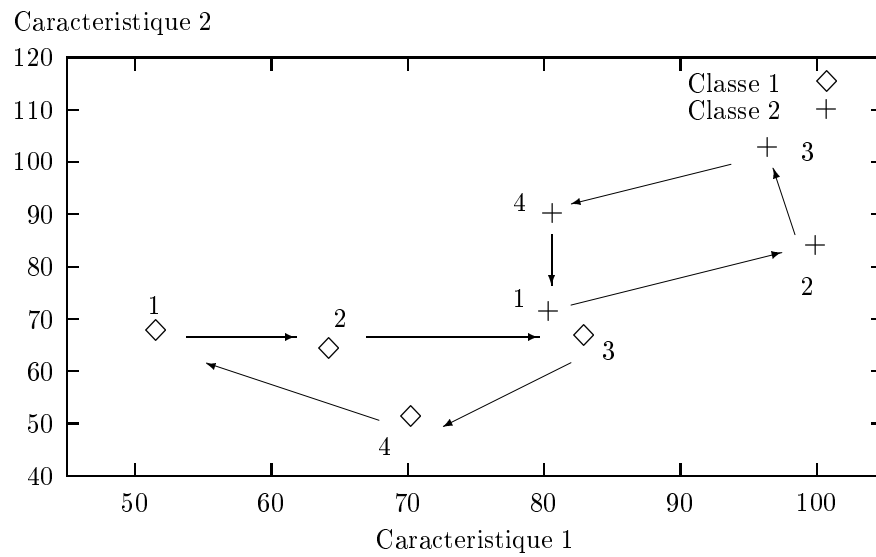


FIG. 4.22 – Positions des centres de classes pour les différents états.

| Etat (Place) | Classes | | | |
|-----------------|-------------|-------------|-------------|-------------|
| | C1 | | C2 | |
| | Paramètre 1 | Paramètre 2 | Paramètre 1 | Paramètre 2 |
| Etat 1 | 51.58 | 67.67 | 80.41 | 71.53 |
| Etat 2 | 64.23 | 64.21 | 99.91 | 84.19 |
| Etat 3 | 82.98 | 66.65 | 96.44 | 102.9 |
| Etat 4 | 70.33 | 51.12 | 80.62 | 90.28 |

TAB. 4.4 – Centres des classes obtenus pour les différents états (4) du système (après le cycle 3).

4.7.8 Autres positions initiales

Initialisation 1

La figure 4.23 présente un cas où les conditions initiales sont différentes (position initiale choisie dans la période). Nous obtenons également 4 états qui sont représentés dans la figure 4.24. La figure 4.25 montre clairement qu'il y a une périodicité dans l'évolution des marquages obtenus. En effet, on constate bien que les marquages obtenus pour les différentes places ont les mêmes valeurs au début et à la fin du cycle (voir les marquages de P1 et de P2, en particulier). La fonction d'appartenance du marquage de P5 représente l'état supplémentaire obtenu (par rapport au cas où $\sigma = 2$). L'évolution du taux de reconnaissance en apprentissage et en classement est montrée dans les figures 4.26 et 4.27.

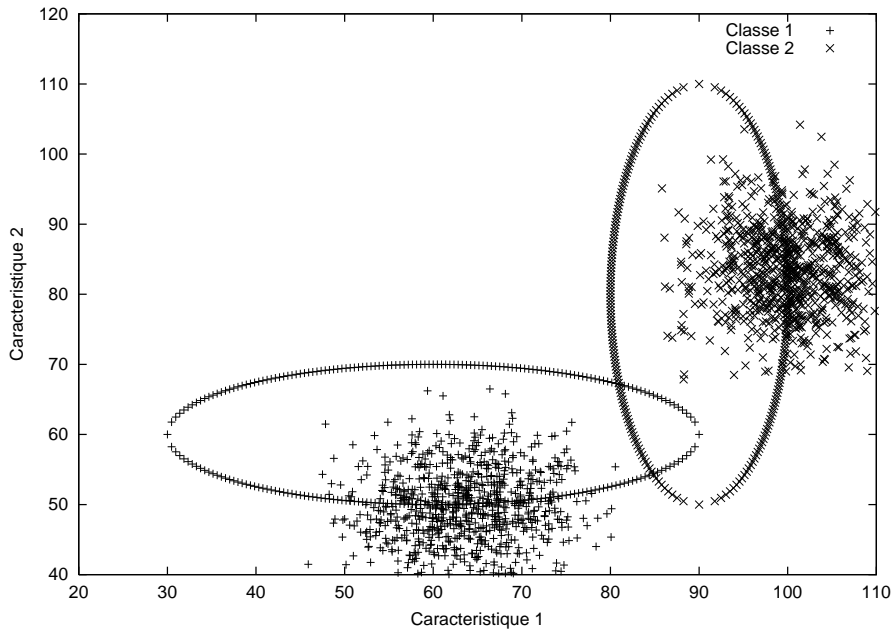


FIG. 4.23 – Initialisation 1 : les deux classes dans les positions 0 et 16 avec leurs trajectoires (pour $\sigma = 5$ et $n_p = 256$ positions).

Initialisation 2

La figure 4.28 présente une autre position initiale. Nous obtenons également 4 états qui sont représentés dans la figure 4.29. L'évolution du taux de reconnaissance en apprentissage et en classement est donné dans les figures 4.30 et 4.31.

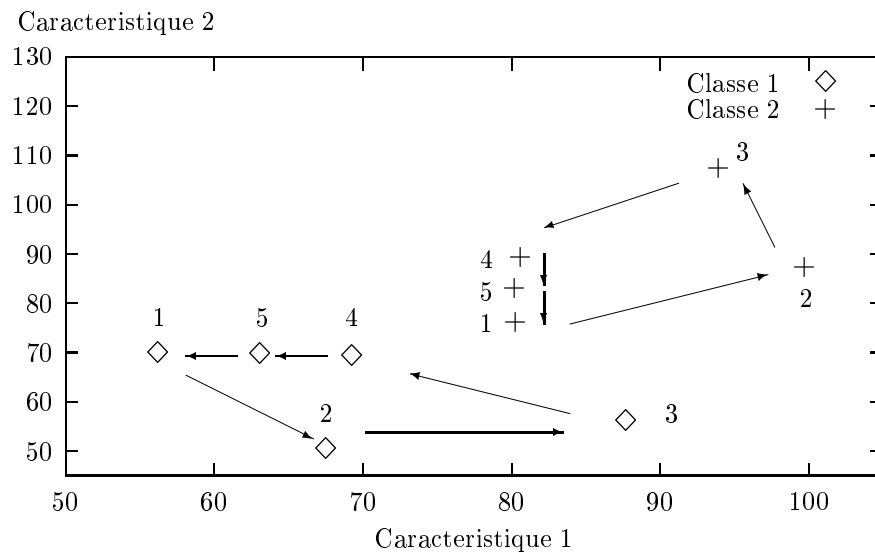


FIG. 4.24 – Initialisation 1 : positions des centres de classes pour les différents états (pour $\sigma = 5$ et $n_p = 256$ positions).

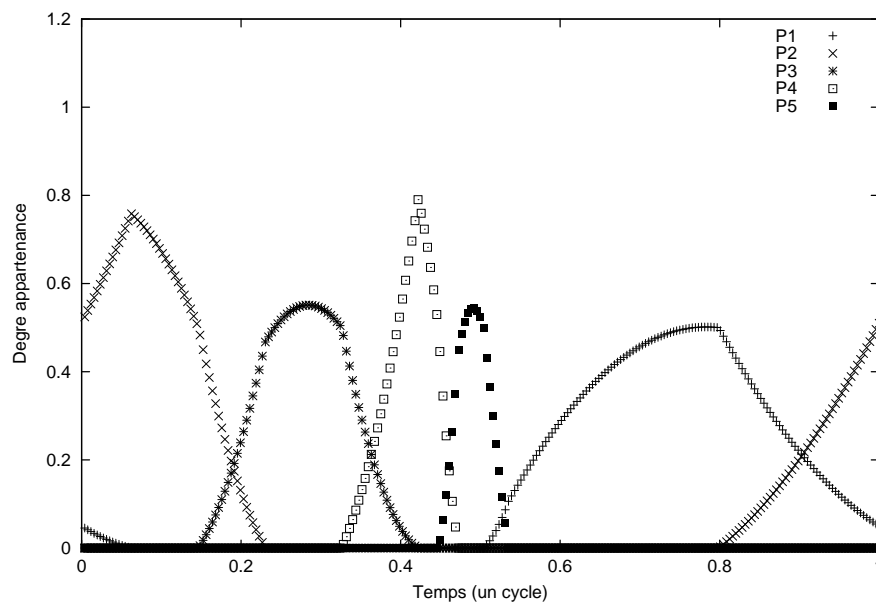


FIG. 4.25 – Initialisation 1 : marquages obtenus au cours du temps (pour $\sigma = 5$ et $n_p = 256$ positions).

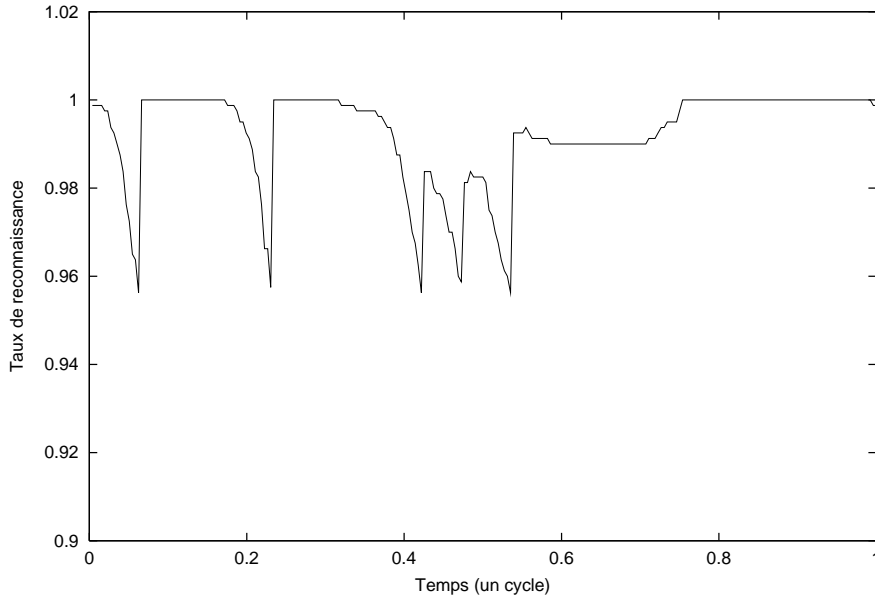


FIG. 4.26 – Initialisation 1 : taux de reconnaissance en apprentissage (pour $\sigma = 5$ et $n_p = 256$ positions). Le taux de confusion toléré sur une classe est de $cf_i = 10\%$.

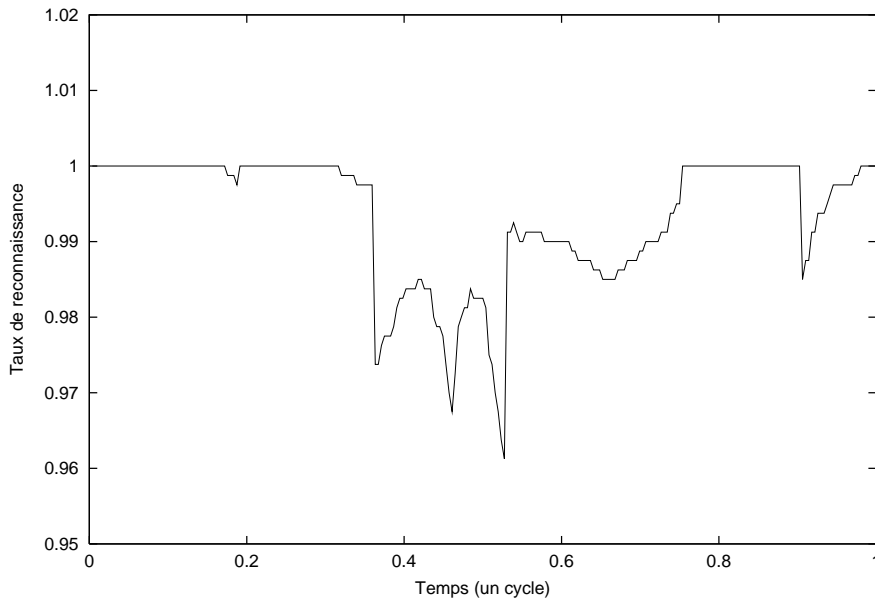


FIG. 4.27 – Initialisation 1 : taux de reconnaissance en classement (pour $\sigma = 5$ et $n_p = 256$ positions). Le taux de reconnaissance global reste supérieur ou égal à 96 %.

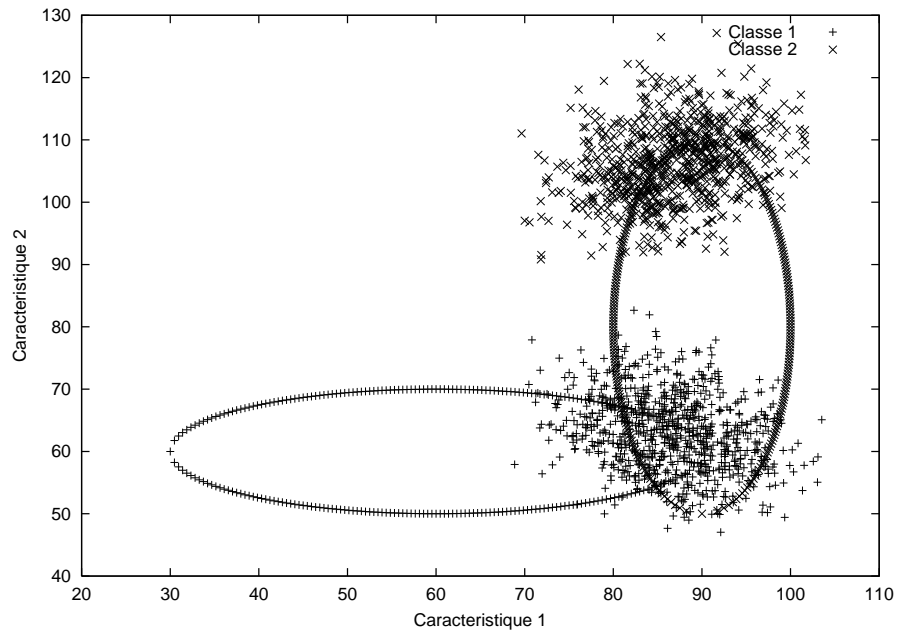


FIG. 4.28 – Initialisation 2 : les deux classes dans les positions 0 et 16 avec leurs trajectoires (pour $\sigma = 5$ et $n_p = 256$ positions).

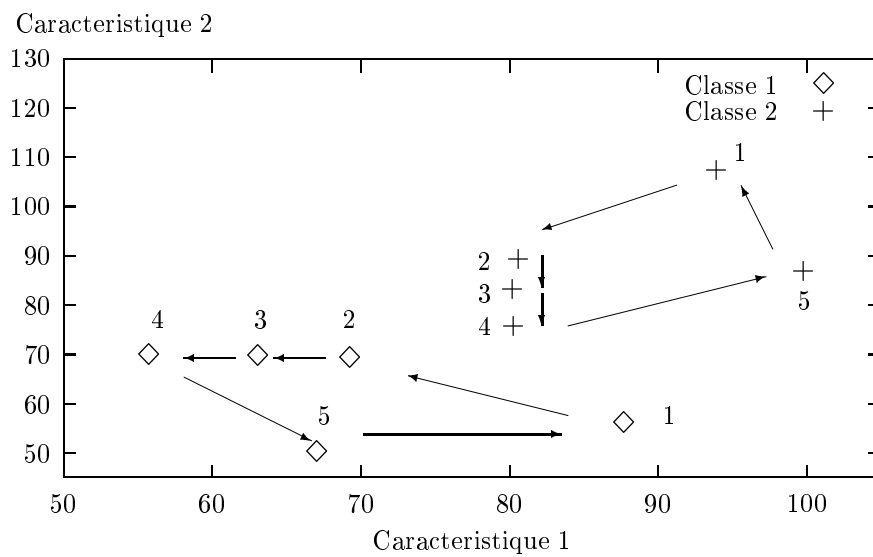


FIG. 4.29 – Initialisation 2 : positions des centres de classes pour les différents états ($\sigma = 5$ et $n_p = 256$ positions).

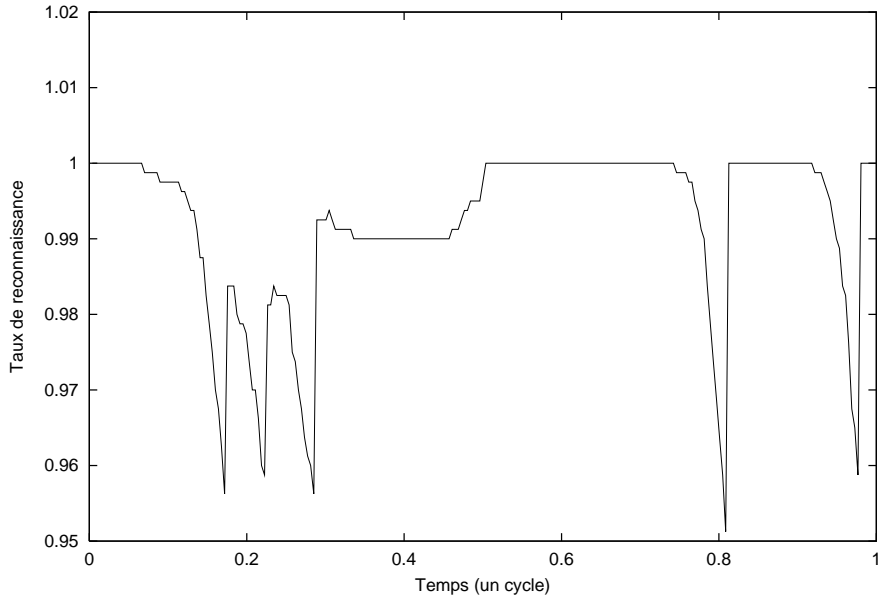


FIG. 4.30 – Initialisation 2 : taux de reconnaissance en apprentissage ($\sigma = 5$ et $n_p = 256$ positions). Le taux de confusion toléré sur une classe est de $cf_i = 10\%$.

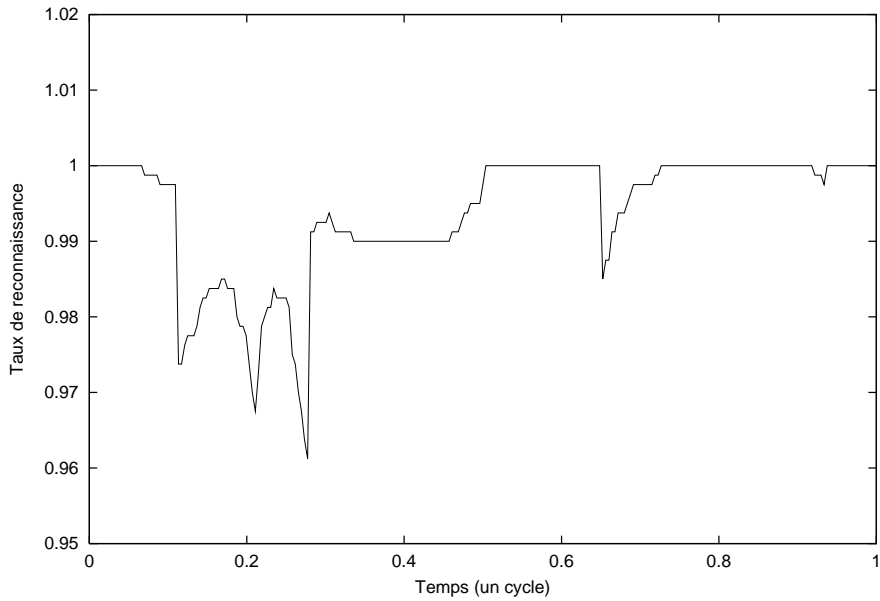


FIG. 4.31 – Initialisation 2 : taux de reconnaissance en classement ($\sigma = 5$ et $n_p = 256$ positions). Le taux de reconnaissance global reste supérieur ou égal à 96 %.

4.8 Conclusion

Dans ce chapitre, nous avons montré comment des Formes issues de classes dynamiques (d'évolutions périodiques) peuvent être reconnues. Le RPF s'avère être un formalisme adéquat pour la représentation de connaissances dynamiques et en particulier pour la représentation des différents états d'un système de reconnaissance. L'avantage de l'approche est de permettre de s'affranchir de la modélisation de l'évolution des classes. Cette évolution est apprise par le RPF.

La construction des fonctions d'appartenance des jetons et des transitions en fonction d'un ensemble d'apprentissage permet d'anticiper l'évolution des classes, c'est-à-dire, de prédire l'état du système de RdF par un RPF. Ces fonctions ont les propriétés d'être non symétriques et de supports différents. Ceci met en évidence l'adaptation du système de reconnaissance à l'application. De plus, la méthode est assez robuste au bruit. Elle permet la combinaison de classifieurs dans le but d'améliorer les résultats, ce que nous avons montré par une combinaison simple. On peut résumer les avantages de l'approche par les 4 points suivants :

1. apprentissage de connaissances statiques et dynamiques,
2. application aisée aux classes dynamiques (d'évolutions périodiques),
3. adaptation à l'évolution des classes par la construction de fonctions d'appartenance,
4. possibilité d'analyse du système obtenu, à l'aide d'outils mathématiques utilisés pour l'analyse des réseaux de Petri.

Cependant, la méthode peut aussi être améliorée par :

- une optimisation des états obtenus lors de la phase d'apprentissage (quels sont les critères les mieux adaptés pour décider de l'arrêt des itérations),
- une prise en compte d'informations exogènes dans les transitions (informations ne provenant pas du phénomène observé mais pouvant influencer les classements),
- une possibilité de gérer plusieurs futurs.

Conclusions et perspectives

Sans disposer de plusieurs caractéristiques visuelles d'un objet, telles que celles qui représentent sa couleur, sa forme et sa luminosité, peut-on identifier l'objet en question ? Il est évident que plus nous en disposons, plus nous avons de chances d'identifier correctement l'objet, et ces chances sont d'autant plus grandes que la caractéristique supplémentaire est plus indépendante (décorrélée) de celles dont on dispose déjà²⁰.

Sans nos cinq sens, peut-on prendre les "bonnes" décisions ? ou entreprendre les "bonnes" actions ? ou même raisonner ? Bonnes ou mauvaises, nos réflexions, nos décisions et nos actions dépendent de la perception que nous avons de notre environnement et ne peuvent être cohérentes avec notre environnement que si nous disposons d'au moins un de ces cinq sens, à moins d'avoir un sixième sens ! Puisque nous en disposons de plusieurs, nos décisions sont-elles prises en fonction de la combinaison de ces sens ? Lorsqu'un événement se produit autour de nous, pour peu que nous n'ayons pas le regard en direction de cet événement, c'est notre ouïe qui nous permet de prendre la première décision ; après analyse du bruit (en particulier, sa direction de provenance), il est possible de diriger notre regard en direction de l'événement. Ensuite, généralement, c'est la perception visuelle qui permet d'identifier l'événement en tenant compte, éventuellement, du bruit perçu auparavant. Après l'ouïe, on peut également être amené à faire un choix, par exemple, entre diriger notre regard ou palper l'objet en cause. Ce choix peut différer en fonction de la distance à laquelle l'objet se trouve et/ou en fonction de l'instant considéré. Plus généralement, il s'agit d'adopter une sélection adaptative et/ou une sélection dynamique d'actions à entreprendre²¹. A une échelle plus grande, celle de la Société, en plus de la combinaison des avis (élections), les décisions ne sont-elles pas prises, souvent, par un processus de coopération et de sélection entre les différents avis ?

N'est-il pas possible d'appliquer ce type de stratégies d'identifications à la RdF en utilisant une association de classifieurs ? C'est ce que nous avons tenté d'adopter comme ligne directrice dans nos travaux, en étudiant et en proposant des *Systèmes de Classifieurs*. Notre premier système (3C) met en œuvre la coopération et la sélection adaptative de classifieurs. Un système effectuant la sélection de classifieurs en fonction du temps est également proposé au chapitre 4.

²⁰Il n'est donc pas étonnant que l'on puisse mieux identifier en disposant d'un sens supplémentaire (par exemple, l'ouïe).

²¹Dans le cadre de la RdF, ces actions correspondent à des classements de différents types.

Le doute étant une caractéristique des systèmes à apprentissage, il est souvent difficile de le gérer, en tout cas, avec les méthodes actuelles de gestion des connaissances. Ces méthodes ne fournissent, en général, que des mesures de "certitude". Par le fait même de ne pas tenir compte du "doute" ou de "l'ignorance", ces mesures se disqualifient en s'entachant d'incertitude. Si la part de doute est ignorée, il est cependant possible de faire en sorte que son importance relative par rapport aux "certitudes" devienne faible, voire négligeable. Une manière d'y parvenir est de consulter les classifieurs qui sont spécialisés pour le type de décisions à prendre²². D'où la nécessité de gérer un ensemble de classifieurs avec différents types d'associations entre les classifieurs individuels.

Lorsqu'il s'agit de reconnaître des objets naturels évolutifs, l'indisponibilité de certaines caractéristiques ou de certaines sources, la complexité même du problème de reconnaissance (ou encore le fait qu'un problème de reconnaissance soit considéré comme statique à tort) font que, souvent, les classes ont des formes complexes. Pour résoudre ces problèmes, des *Systèmes de Classifieurs* ont été proposés comme une alternative à la recherche de modèles de classes de plus en plus complexes. Cependant, la plupart des systèmes proposés ne mettent en œuvre que la combinaison de classifieurs. Or, cette approche ne suffit plus.

En effet, la combinaison de classifieurs possède certaines limites :

- espace de discernement commun entre les différents classifieurs ; cette contrainte oblige les différents classifieurs (en apprentissage et en classement) à traiter une quantité imposante de données, engendrant un coût de calcul élevé. Tous les classifieurs doivent fournir leurs décisions à un module de combinaison et pour avoir des avis pertinents, chacun d'eux doit gérer la quasi-totalité de l'espace des caractéristiques et de représentation. De plus, une approche répartie d'un *Système de Classifieurs*, dans de telles conditions, est contraignante en termes d'échanges d'informations provenant des différents classifieurs. Les techniques de coopération de classifieurs, en particulier les *Systèmes de Classifieurs* multi-niveaux, peuvent surmonter ces deux problèmes.
- l'architecture de la prise de décision est rigide et ne permet pas une adaptation du système à son environnement. En particulier, lorsque les classes (ou certaines classes) traitées sont susceptibles de subir des évolutions temporelles, cette architecture nécessite une mise à jour de l'apprentissage de l'ensemble des classifieurs. Réeffectuer la phase d'apprentissage et dans le cas d'un système de classifieurs réparti, celle d'échanger une importante quantité d'informations, sont des contraintes sévères. La sélection spatiale et/ou spatio-temporelle (i.e., en fonction de l'individu à classer et/ou en fonction de l'individu et du temps) de classifieurs permet de résoudre ces types de problèmes.

Il est donc plus prudent d'utiliser à la fois la combinaison, la coopération et la sélection

²²Dans la vie courante, il est souvent conseillé de ne parler que de ce que l'on connaît.

tion de classifieurs. L'idéal serait plutôt de combiner, de faire coopérer et de sélectionner de manière dynamique différents classifieurs. Nous avons mis en œuvre deux systèmes de reconnaissance mettant partiellement cet idéal en pratique. Le premier met en œuvre la combinaison, la coopération et la sélection adaptative de classifieurs. Les performances de ce système diminuent lorsque les trajectoires des classes s'intersectent ou lorsque les classes comportent de nombreux modes et que l'ensemble d'apprentissage n'est pas suffisant pour les distinguer correctement. Le second met en œuvre la sélection dynamique et la combinaison de classifieurs. De plus, il peut facilement être adapté pour effectuer également de la coopération, ce que nous espérons réaliser dans un proche avenir. La limite de fonctionnement de ce système est atteinte lorsque deux ou plusieurs classes évolutives se chevauchent fortement dans l'espace de représentation, à un instant considéré (croisement spatio-temporel).

Nos objectifs à court terme sont de tester et d'adapter le premier système sur des données comportant plus de classes (les classifieurs individuels pouvant alors avoir des cadres de discernements différents). Ensuite, on pourra envisager d'associer ce type de système à certains états générés dans le second système (en particulier, lorsqu'il y a chevauchement des classes). Enfin, le second système pourrait être modifié de manière à gérer plusieurs futurs. A partir d'un état donné, on peut ainsi avoir le choix entre plusieurs autres états, en tenant compte d'informations exogènes.

Les *Systèmes de Classifieurs* pourront servir dans les applications à données réparties car ils sont prédisposés à l'implémentation selon ce type d'architecture. En effet, lorsque les données distantes sont obtenues avec un haut débit/faible distance, la stratégie de la combinaison de classifieurs peut être la plus efficace du fait de sa précision. Lorsqu'elles sont obtenues selon un débit moyen/distance moyenne, la coopération avec des classifieurs distants peut être la stratégie la mieux adaptée (possibilité de réduire la quantité des informations échangées). Enfin, lorsque les données sont obtenues avec un faible débit/longue distance ou lorsqu'elles peuvent subir des évolutions temporelles, la sélection de classifieurs distants peut être la meilleure stratégie.

A long terme, les modes de fonctionnement ou de surcharge des réseaux de télécommunications pourraient être identifiés ou détectés au niveau régional, national et/ou international. Les récessions économiques ou crises financières pourraient être identifiées et/ou prédites à partir d'un *Système de Classifieurs* réparti sur les principales places boursières internationales en exploitant divers indices économiques mesurés régionalement, nationalement ou internationalement.

Bibliographie

- [1] B. Achermann and H. Bunke. Combination of classifiers on the decision level for face recognition. Rapport interne IAM-96-002, Institut für Informatik, Université de Bern, Suisse, Jan. 1996.
- [2] Association AFIA. Dossier IA et Logique. In *Bulletin de l'AFIA n° 35*, pages 15–35. AFIA, Oct. 1998.
- [3] M.D. Alder. *Principles of Pattern Classification : Statistical, Neural Net and Syntactic Methods of Getting Robots to See and Hear*. UWA Centre for Intelligent Information Processing Systems, ftp ://ciips.ee.uwa.edu.au/pub/syntactic/book/, August 1994.
- [4] A. Appriou. Probabilités et incertitude en fusion de données multi-senseurs. *Revue Scientifique et Technique de la Défense*, pages 27–40, Jan. 1991.
- [5] A. Appriou. La fusion de données : de la recherche aux développements opérationnels. In *Approche méthodologique pour la fusion de données - JSO2000*, Chatillon, France, Déc. 2000.
- [6] D. Bahler and L. Navarro. Methods for combining heterogeneous sets of classifiers. In *17th Natl. Conf. on Artificial Intelligence (AAAI 2000), Workshop on New Research Problems for Machine Learning*, 2000.
- [7] A. Ben-Hur, D. Horn, H.T. Siegelmann, and V. Vapnik. A support vector clustering method. In *Proceedings of ICPR2000, 15th International Conference on Pattern Recognition*, pages 728–731, Barcelone, Espagne, Sept. 2000.
- [8] A.M. Bensaïd, L.O. Hall, J.C. Bezdek, and L.P. et al. Clarke. Validity-Guided (Re)Clustering with Applications to Image Segmentation. *IEEE Transactions on Fuzzy Systems*, 4(2) :112–123, May 1996.
- [9] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New-York, 1981.
- [10] I. Bloch. Incertitude, impécision et additivité en fusion de données : point de vue historique. *Traitement du Signal*, 13(4) :267–287, 1996.
- [11] I. Bloch and H. Maître. Fusion of Image Information under Imprecision. In B. Bouchon-Meunier, editor, *Aggregation and Fusion of Imperfect Information*, Series Studies in Fuzziness, pages 189–213. Physica Verlag, Springer, 1997.
- [12] E. Boutleux. *Diagnostic et suivi d'évolution de l'état d'un système, par reconnaissance des formes floues. Application au modèle du réseau téléphonique français*. Thèse, Université de Compiègne, France, Nov. 1996.
- [13] J. Cao, M. Ahmadi, and M. Shridhar. Handwritten numerals with multiple features and multistage classifiers. *IEEE International Journal on Circuits and Systems*, 6(London) :323–326, June 1994.

-
- [14] J. Cardoso, R. Valette, and D. Dubois. Fuzzy Petri Nets : an overview. In *13th IFAC World Congress*, pages 443–448, San Francisco, USA, 1996.
- [15] K. Chen, L. Wang, and H. Chi. Methods of combining multiple classifiers with different features and their applications to text-independant speaker identification. *International Journal of Pattern Recognition and Artificial Intelligence*, 11(3) :417–445, 1997.
- [16] S.M. Chen, J.S. Ke, and J.F. Chang. Knowledge Representation using Fuzzy Petri Nets. *IEEE Transactions on Knowledge and Data Engineering*, 2(3) :311–319, 1990.
- [17] S.B. Cho and J.H. Kim. Combining multiple neural networks by fuzzy integral for robust classification. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-25(2) :380–384, Fév. 1995.
- [18] C.K.Chow. On optimum recognition error and reject tradeoff. *IEEE Transactions on Information Theory*, 16 :41–46, January 1990.
- [19] C.K.Chow. Recognition error and reject trade-off. In *Proceedings of Third Annual Symposium on Document Analysis and Information Retrieval*, volume 13, pages 1–8, University of Nevada, Las Vegas, USA, April 1994.
- [20] G. Cron. *Diagnostic par reconnaissance des formes floue d'un système dynamique et réparti. Application à la gestion en temps réel du trafic téléphonique français*. Thèse, Université de Compiègne, France, Nov. 1999.
- [21] D. Demigny, M. Paindavoine, and S. Weber. Architecture à reconnaissance dynamique pour le traitement temps réel des images. *Techniques et Sciences Informatiques*, 1999.
- [22] C. Demko, P. Loonis, and M. Ménard. Les c+1-moyennes floues : introduction du rejet en classification. In *Actes des sixièmes rencontres de la société francophone de classification*, Montpellier, France, Sept. 1998.
- [23] A.P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.*, 38 :325–339, 1967.
- [24] A. Djouadi and E. Bouktache. A fast algorithm for the nearest-neighbor classifier. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(3) :277–282, 1997.
- [25] C. Dousson and M. Ghallab. Suivi et reconnaissance de chroniques. Technical Report 94082, Université de Toulouse, LAAS, Mars 1994.
- [26] D. Dubois and H. Prade. Possibility theory. *Plenum Press. New York*, 1988.
- [27] B. Dubuisson. Decision with reject option. In *European Signal Processing Conference*, Barcelone, SPAIN, 1990.
- [28] P. Estraillier and C. Girault. *Applying Petri Net Theory to the Modelling, Analysis and Prototyping of Distributed Systems. In Manufacturing Research and Technology, Modern Tools for Manufacturing Systems*. Elsevier, 1993.
- [29] P. Fabiani. *Représentation dynamique de l'incertain et stratégie de prise d'information pour un système autonome en environnement évolutif*. Thèse, Université de Toulouse, France, http://www.cert.fr/fr/dcsd/PUB/THESES/fabiani/manuscrit_fabiani/, Déc. 1996.
- [30] R. A. Fisher. The statistical utilisation of multiple measurements. *Annals of Eugenetics*, 8, 1938.
- [31] B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical Recipes in C, The Art of Scientific Computing*. Cambridge University Press, 1990.

-
- [32] J. Franke and E. Mandler. A comparison of two approaches for combining the votes of cooperating classifiers. In *Proceedings of 11th International Conference on Pattern Recognition*, volume 2, pages 611–614, The Hague, Netherlands, 1992.
- [33] G. Giacinto and F. Roli. Adaptive selection of image classifiers. In *ICIAP'97, 9th International Conference on Image Analysis and Processing*, pages 38–45, Florence, Italy, Sept. 1997. Lecture Notes in Computer Science 1310, Springer.
- [34] G. Giacinto and F. Roli. Ensembles of neural networks for soft classification of remote-sensing images. In *Proceedings of the European Symposium on Intelligent Techniques*, pages 166–170, Bari, Italy, Mars 1997.
- [35] G. Giacinto and F. Roli. Methods for dynamic classifier selection. In *ICIAP '99, 10th International Conference on Image Analysis and Processing*, pages 659–664, Venice, Italy, Sept. 1999.
- [36] G. Giacinto and F. Roli. An approach to the automatic design of multiple classifier systems. *Pattern Recognition Letters*, 22 :25–33, 2001.
- [37] G. Giacinto, F. Roli, and Giorgio Fumera. Design of effective multiple classifier systems by clustering of classifiers. In *Proceedings of ICPR2000, 15th International Conference on Pattern Recognition*, pages 160–163, Barcelone, Espagne, Sept. 2000.
- [38] A. Giani. *Parallel Cooperative Classifier Systems*. PhD thesis, Université de Pise, Italie, Mars 1999.
- [39] B. Gosselin. Cooperation of multilayer perceptron classifiers. In *Proceedings of the 8th Workshop on Circuits, Systems and Signal Processing*, pages 187–190, Mierlo, Pays-Bas, 1997.
- [40] V. Gunes, P. Loonis, and M. Ménard. Fusion adaptative multi-classifieurs appliquée aux classes évolutives. In *Journée GDR-ISIS GT 6 Fusion : Fusion et Décision*, Paris, France, Fév. 2000.
- [41] V. Gunes, P. Loonis, and M. Ménard. Reconnaissance de classes dynamiques : une modélisation par réseau de petri flou. In *LFA2001, Logique Floue et Applications*, Mons, Belgique, Nov. 2001.
- [42] V. Gunes, P. Loonis, and M. Ménard. A fuzzy petri net for pattern recognition : Application to dynamic classes. *Knowledge And Information Systems (KAIS)*, Springer-Verlag, A paraître en 2002.
- [43] V. Gunes, P. Loonis, and M. Ménard. Représentation de classes dynamiques par des réseaux de petri flous. In *Journée GDR-ISIS GT 6 Fusion : Méthodologie en fusion d'informations*, Paris, France, Jan. 2001.
- [44] V. Gunes, M. Ménard, and P. Loonis. Fuzzy clustering with ambiguity for multi-classifiers fusion : Clustering-Classification Cooperation. In *1999 EUSFLAT-ESTYLF Joint Conference*, pages 505–508, Palma (Mallorca), Spain, 1999.
- [45] V. Gunes, M. Ménard, and P. Loonis. Coalescence floue avec ambiguïté pour la fusion multi-classifieurs : Cooperation Classification-Classement. In *RFIA2000, Reconnaissance des Formes et Intelligence Artificielle*, Paris, France, Fév. 2000.
- [46] V. Gunes, M. Ménard, and P. Loonis. A multiple classifier system using ambiguity rejection for Clustering-Classification Cooperation. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems (IJUFKS)*, 8(6) :747–762, World Scientific, Déc. 2000.

- [47] T.M. Ha. The optimum class-selective rejection rule. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol.19, No.6, pages 608–614, Yokohama, Japan, March 1997.
- [48] D. Happel and P. Bock. Overriding the experts : A stacking method for combining marginal classifiers. In *Proceedings of the 13th International FLAIRS Conference*, Menlo Park, CA, USA, 2000.
- [49] R.M. Haralick. The table look-up rule. *Communications in Statistics - Theory and Methods*, A5(12) :1163–1191, 1976.
- [50] R.J. Hathaway, J.W. Davenport, and J.C. Bezdek. Relational duals of the c-means algorithms. *Pattern Recognition*, 22 :205–212, 1989.
- [51] J.P. Haton. Connectionist and hybrid models for automatic speech recognition. In K. Ponting, editor, *Computational Models of Speech Pattern Recognition*. Springer, 1998.
- [52] T. K. Ho. *A Theory of Multiple Classifier Systems And Its Application to Visual Word Recognition*. PhD thesis, State University of New York at Buffalo, May 1992.
- [53] T. K. Ho, J.J. Hull, and S.N. Srihari. Decision combination in multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16 :66–75, Jan 1994.
- [54] T.K. Ho. Random decision forests. In *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, pages 278–282, Montreal, Canada, August 1995.
- [55] T.K. Ho. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8) :832–844, Aug. 1998.
- [56] L. Holmström, P. Koistinen, J. Laaksonen, and E. Oja. Comparison of neural and statistical classifiers - theory and practice. Technical Report A13, Helsinki University of Technology, Finland, <http://www.cis.hut.fi/jorma/papers/>, Jan. 1996.
- [57] T.S. Huang and C.Y. Suen. Combination of multiple experts for the recognition of unconstrained handwritten numerals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(1) :90–94, Jan 1995.
- [58] Y.S. Huang and Suen C.Y. A method of combining multiple classifiers - a neural network approach. In *Proceedings of 12th International Conference on Pattern Recognition and Computer Vision*. Jerusalem, pages 473–475, 1994.
- [59] A. Jaimes and S.F. Chang. Automatic selection of visual features and classifiers. In *IS&T/SPIE Storage and Retrieval for Image and Video Databases VIII*, San Jose, CA, U.S.A., Jan. 2000.
- [60] A. Jaimes and S.F. Chang. Integrating multiple classifiers in visual objects detectors learned from user input. In *4th Asian Conference on Computer Vision (ACCV 2000)*, Tapei, Taiwan, Jan. 2000.
- [61] A.K. Jain, R.P.W. Duin, and J. Mao. Statistical pattern recognition : A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1) :4–37, Jan 2000.
- [62] W.T. Jong, Y.S. Shiau, Y.J. Horng, H.H. Chen, and S.M. Chen. Temporal knowledge representation and reasoning techniques using time Petri nets. *IEEE Transactions on Systems, Man and Cybernetics.*, 29(4) :541–545, August 1999.
- [63] F. Kimura and M. Shridar. Handwritten numerical recognition based on multiple algorithms. *Pattern recognition*, 24(10) :969–983, 1991.
- [64] J. Kittler, M. Hatef, R.P.W. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3) :226–239, 1998.

-
- [65] C. Koch. Towards the neuronal substrate of visual consciousness. In *Towards a Science of Consciousness : The First Tucson Discussions and Debates*, Tucson, USA, 1994.
- [66] T. Kohonen. Self-Organizing Maps. *Springer Series in Information Sciences*, 30, Oct. 1995.
- [67] R. Krishnapuram and J.M. Keller. The possibilistic c-means algorithm : Insights and recommendations. *IEEE Transactions on Fuzzy Systems*, 4(3) :385–393, August 1996.
- [68] L.I. Kuncheva, C.A. Whitaker, C.A. Shipp, and R.P.W. Duin. Is independence good for combining classifiers ? In *Proceedings of ICPR2000, 15th International Conference on Pattern Recognition*, pages 168–171, Barcelone, Espagne, Sept. 2000.
- [69] L. Lam and C.Y. Suen. A theoretical analysis of the application of majority voting to pattern recognition. *International Conference of Pattern Recognition, Jerusalem*, pages 418–420, 1994.
- [70] C.G. Looney. Fuzzy Petri nets for rule-based decision making. *IEEE Trans. on Systems, Man and Cybernetics*, 18(1) :178–183, Jan-Feb 1988.
- [71] C.G. Looney and A.R. Alfize. Logical controls via boolean rule matrix transformations. *IEEE Trans. on Systems, Man and Cybernetics*, SMC-17(6) :1077–1082, Nov-Dec 1987.
- [72] P. Loonis. *Contribution à la minimisation de l'a priori en Reconnaissance des Formes*. Thèse, Université de La Rochelle, France, Jan. 1996.
- [73] P. Loonis and M. Ménard. Fusion connexionniste et algorithme génétique pour adapter les systèmes de reconnaissance des formes à leur environnement. *Traitement du Signal*, Nov. 1996.
- [74] P. Loonis and S. Sochacki. Contrôle non destructif de la qualité des produits de la mer. In *GRETSI 2001*, Toulouse, Sept. 2001.
- [75] S. Mathevet, L. Trassoudaine, P. Checchin, and J. Alizon. Combinaison de segmentations en régions. *Traitement du Signal*, 16(2) :93–104, 1999.
- [76] G.J. McLachlan and K.E. Basford. *Mixture Models : Inference and Applications to Clustering*. Marcel Dekker, New York, U.S.A., July 1988.
- [77] M. Menard. Le rejet d'ambiguïté dans les fc+2m : extension à toutes les classes composées. *Traitement du Signal*, 15(6) :24, 1999.
- [78] M. Ménard, C. Demko, and P. Loonis. The fuzzy c+2-means : Solving the ambiguity reject in clustering. *Pattern recognition*, 33(7) :1219–1237, 2000.
- [79] T. Murata. Petri nets : Properties, analysis and applications. *Proceedings of the IEEE*, 77(4) :541–580, April 1989.
- [80] G.S. Ng and H. Singh. Data equalisation with evidence combination for pattern recognition. *Pattern Recognition Letters*, 19 :227–235, 1998.
- [81] N.J. Nilsson. *Learning Machines*. McGraw-Hill, 1965.
- [82] J.R. Parker. Rank and response combination from confusion matrix data. Technical Report 2000-660-12, University of Calgary, Canada, June 2000.
- [83] Z. Pawlak. Rough sets. Theoretical aspects of reasoning about data. *Kluwer Academic Publishers. Dordrecht, Boston, London*, 1991.
- [84] J.G. Postaire. *De l'image à la décision*. DUNOD Informatique, Paris, France, 1987.
- [85] J.G. Postaire and L. Macaire. Exploitation de l'information couleur pour la segmentation d'images. In *Journée GDR-ISIS GT 3 Imagerie couleur*, Paris, France, Mai 1998.

- [86] D. Puzeat. Priming an artificial neural classifier. In *International Work Conference on Artificial Neural Networks (IWANN'95)*, pages 559–565, Malaga, Spain, June 1995.
- [87] J.R. Quinlan. Bagging, boosting, and c4.5. In *Proceedings of 13th National Conference on Artificial Intelligence*, pages 725–730, Portland, Oregon, USA, Aug. 1996.
- [88] G. Rogova. Combining the results of several neural network classifiers. *Neural Networks*, 7(5) :777–781, 1994.
- [89] M. Rombaut, S. Lorette-Rougegrez, J.M. Nigro, and I. Jarkass. Temporal sequence recognition using uncertain sensor data. In *FUSION2000*, Paris, France, July 2000.
- [90] J. Romeder. *Méthodes et programmes d'analyses discriminantes*. Dunod, Paris, 1973.
- [91] S.L. Salzberg. On comparing classifiers : A critique of current research and methods. *Data Mining and Knowledge Discovery*, 1 :1–12, 1999.
- [92] S. Sarkar and K.L. Boyer. Quantitative measures of change based on feature organisation : Eigenvalues and eigenvectors. *Computer Vision and Image Understanding.*, 71(1) :110–136, 1998.
- [93] G. Shafer. A mathematical theory of evidence. *Princeton University Press, Princeton, N.J.*, 1976.
- [94] A.J.C. Sharkey. Multi-net systems. In : *Combining Artificial Neural Nets, Ensemble and Modular Multi-Net Systems*, pages 1–27, 1999.
- [95] A.J.C. Sharkey and N.E. Sharkey. The "test and select" approach to ensemble combination. In *Proceedings of First International Workshop on Multiple Classifier Systems, Lecture Notes in Computer Science*, pages 30–44, 2000.
- [96] J.C. Simon. La reconnaissance des formes à l'épreuve des faits. *La Recherche*, 312, Sept. 1998.
- [97] P. Smets. The combination of evidence in the transferable belief model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5) :447–458, 1990.
- [98] C. Y. Suen, C. Nadal, R. Legault, T. A. Mai, and L. Lam. Computer recognition of unconstrained handwritten numerals. *Proceedings of the IEEE*, 80 :1162–1180, 1992.
- [99] C.Y. Suen, T.A. Nadal, T.A. Mai, R. Legault, and L. Lam. Recognition of totally unconstrained handwritten numerals based on the concept of multiple experts. *Frontiers in Handwriting Recognition*, pages 131–143, Montreal : Corcordia University, 1990.
- [100] D.M.J. Tax, M. Van Breukelen, R.P.W. Duin, and J. Kittler. Combining multiple classifiers by averaging or by multiplying. *Pattern recognition*, 33 :1475–1485, 2000.
- [101] S. Theodoridis and K Koutroumbas. *Pattern Recognition*. Academic Press, San Diego, USA, 1999.
- [102] V. Tresp and M. Taniguchi. Combining estimators using non-constant weighting functions. *Advances in Neural Information Processing Systems*, 7, 1995.
- [103] K. Tumer and J. Ghosh. Error correlation and error reduction in ensemble classifiers. *Connection Science*, 8 :385–404, Déc 1996.
- [104] J.M. Twomey and A.E. Smith. Committee networks by resampling. In *Intelligent Neural Engineering Systems Through Artificial Networks*, volume 5, pages 153–158. ASME Press, 1995.
- [105] R. Valette, D. Andreu, J. Cardoso, and J.C. Pascal. Fuzzy Petri Nets and their application in CIME. *Special Issue of Recent Applications of Petri Nets of the Trans. of the Inst. of Electrical Engineers of Japan*, 114-C(9) :876–880, 1994.

-
- [106] R. Valette and L.A. Künzle. Réseaux de Petri pour la détection et le diagnostic. In *Journées Nationales : Sécurité, Surveillance, Supervision*, Paris, France, 1994.
- [107] R. Valette, B. Pradin-Chèzalviel, and F. Girault. An introduction to Petri net theory. *Studies in Fuzziness and Soft Computing*, 22(1) :3–24, 1999.
- [108] M. Van Breukelen, R.P.W. Duin, and D.M.J. Tax. Combining classifiers for the recognition of handwritten digits. In *1st IAPR TC1 Statistical Techniques in Pattern Recognition*, pages 13–18, Prague, Tchéquie, Juin 1997.
- [109] A. Verikas, A. Lipnickas, K. Malmqvist, M. Bacauskiene, and A. Gelzinis. Soft combination of neural classifiers : A comparative study. *Pattern Recognition Letters*, 20 :429–444, 1999.
- [110] D. H. Wolpert. Stacked generalization. *Neural Networks*, 5(2) :241–259, 1992.
- [111] K. Woods, W.P. Kegelmeyer, and JR.K. Bowyer. Combination of multiple classifiers using local accuracy estimates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(4), 1997.
- [112] L. Xu, A. Krzyzak, and C.Y. Suen. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man and Cybernetics*, 22(3) :418–435, 1992.
- [113] M.S. Yang. A survey of fuzzy clustering. *Mathl. Comput. Modelling*, 18(11) :1–16, 1993.
- [114] L. A. Zadeh. Fuzzy sets. *Information Control*, 8 :338–353, 1965.
- [115] E. Zahzah. *Contribution à la représentation des connaissances et à leur utilisation pour l'interprétation automatique des images satellites*. PhD thesis, Université Paul Sabatier, Toulouse, Septembre 1992.

Annexe A

Résultats de l'algorithme 3C avec les données ELENA

A.1 Les données ELENA utilisées

Nous testons l'algorithme 3C sur des classes multimodales et qui se chevauchent. Les données *clouds* de la base de données ELENA (programme de recherches ESPRIT 6891) conviennent très bien pour nos tests. Ce jeu de données (artificielles) est constitué de deux classes. La classe ϕ_1 est la somme de trois distributions Gaussiennes différentes :

$$p(x/\phi_1) = \frac{1}{2} \left(\frac{p_1(x)}{2} + \frac{p_2(x)}{2} + p_3(x) \right), \quad (\text{A.1})$$

où

$$p_j(x) = \frac{1}{2\pi\sigma_{j1}\sigma_{j2}} \exp \left(-\frac{(x - m_{j1})^2}{2\sigma_{j1}^2} - \frac{(x - m_{j2})^2}{2\sigma_{j2}^2} \right). \quad (\text{A.2})$$

Les valeurs m_{j1} et m_{j2} sont les moyennes de la distribution j selon les caractéristiques 1 et 2 et σ_{j1} et σ_{j2} sont les écart-types correspondants. Les paramètres sont résumés dans la Table A.1.

| j | m_{j1} | m_{j2} | σ_{j1} | σ_{j2} |
|-----|----------|----------|---------------|---------------|
| 1 | 0.0 | 0.0 | 0.2 | 0.2 |
| 2 | 0.0 | 2.0 | 0.2 | 0.2 |
| 3 | 2.0 | 1.0 | 0.2 | 1.0 |

TAB. A.1 – Paramètres des 3 distributions de la classe ϕ_1 .

La classe ϕ_2 est constituée d'une seule distribution Gaussienne :

$$p(x/\phi_2) = \frac{1}{2\pi} \exp \left(-\frac{(x^2 + y^2)}{2} \right). \quad (\text{A.3})$$

Les données disponibles sur le site ftp anonyme :

ftp.dice.ucl.ac.be, dans le répertoire *pub/neural/ELENA/databases*,

comportent 2500 individus pour chacune des deux classes. Nous avons utilisé les 200 premiers individus de chaque classe. Les deux classes obtenues sont représentées dans la figure A.1.

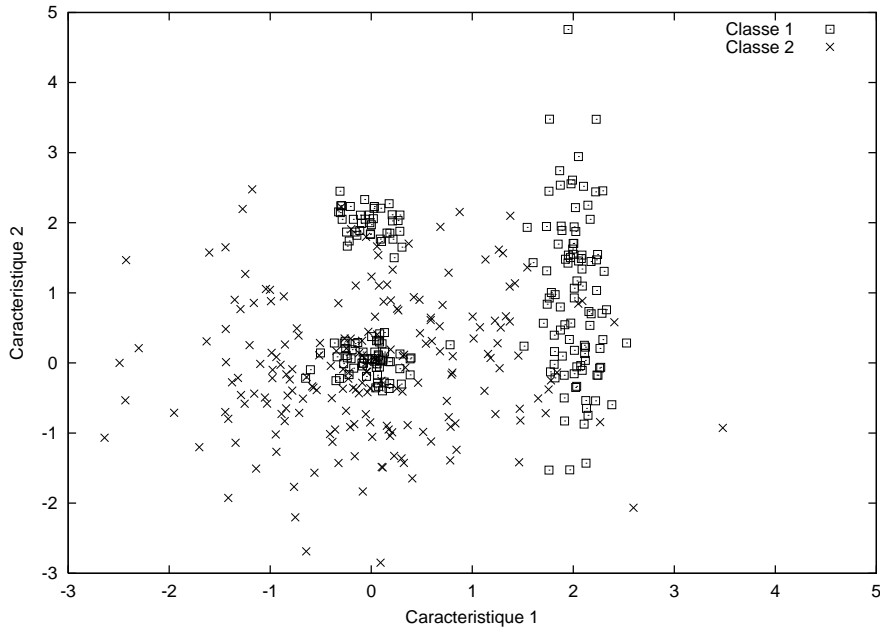


FIG. A.1 – Les deux classes des données *clouds*.

A.2 Influence de l'ambiguïté

Soit n_a , le nombre d'individus classés dans des CCAs lors de la phase de coalescence. Pour tous les résultats obtenus ci-après, le facteur de flou vaut $m = 1.5$, $T = 5$ et $c = 3$. La Table A.2 montre que le meilleur résultat de classement de l'ensemble d'apprentissage est obtenu pour $\alpha = 0.7$ (α optimal).

| α | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.9 |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|
| % | 67.75 | 81.25 | 82.50 | 82.50 | 84.75 | 86.25 | 86.75 | 86.00 |
| n_a | 371 | 299 | 244 | 182 | 91 | 40 | 21 | 0 |

TAB. A.2 – Influence de α sur le taux de reconnaissance de l'ensemble d'apprentissage (pour $c = 3$).

A.3 Résultats pour α optimal

Les ensembles d'apprentissage des différents classifieurs Bayesiens (au nombre de 3 pour $c = 3$) sont obtenus par la procédure de coalescence (Fc+2M). Les degrés d'appartenances des individus à ces différents ensembles sont représentés par la figure A.2.

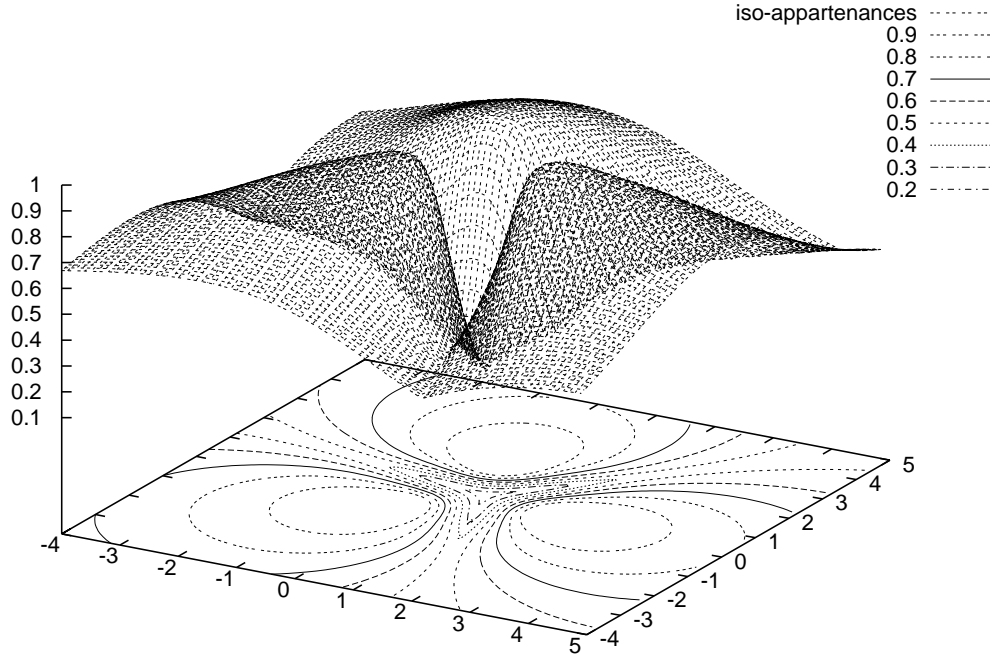


FIG. A.2 – Degrés d'appartenance maximales parmi les appartenances aux 3 ensembles d'apprentissage.

Les masses affectées à la classe ϕ_1 sont représentées par la figure A.3. Les frontières de décisions sont mis en évidence dans la figure A.4. Cette valeur de α conduit au meilleur taux de reconnaissance de l'ensemble d'apprentissage (pour $c = 3$).

A.4 Compromis entre précision et généralisation

Dans ce cas ci-dessus, la classe ϕ_2 est privilégiée par rapport à ϕ_1 en terme de régions d'affectations. Les classes sont précisément délimitées par des frontières de décisions qui épousent les formes des classes. Si les individus qui constituent l'ensemble d'apprentissage sont bien représentatifs des classes réelles, cette précision est bénéfique. Dans le cas contraire, cette précision conduit à une sur-adaptation du système à l'ensemble d'apprentissage et privilégie ainsi une des classes sans justification. De manière générale (si on a peu ou pas d'informations sur la représentativité de l'ensemble d'apprentissage), il ne faut

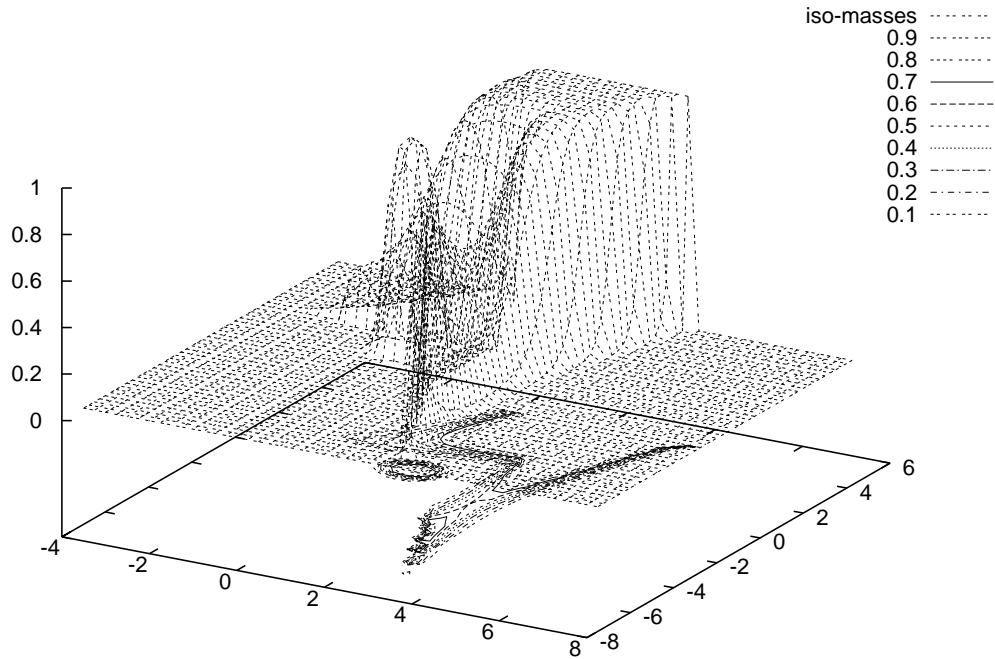


FIG. A.3 – Iso-masses de la classe ϕ_1 pour $\alpha = 0.7$ ($T = 5$).

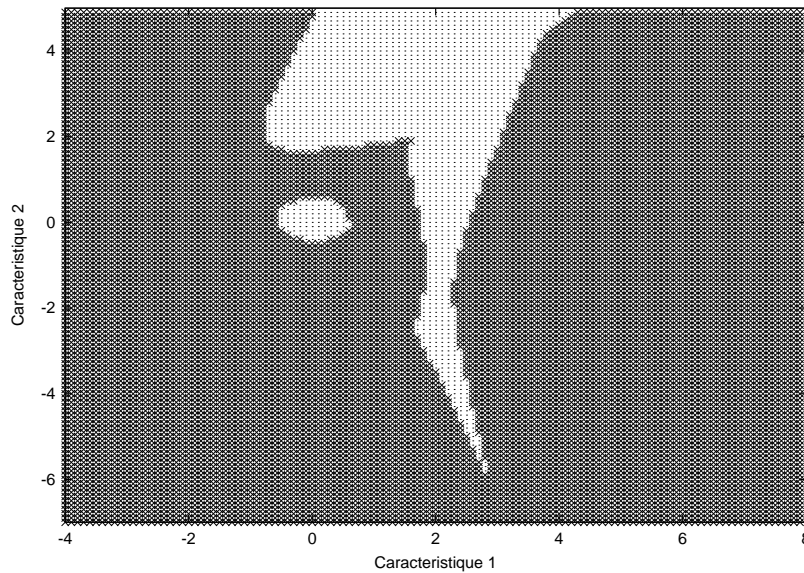


FIG. A.4 – Régions/frontières obtenues pour $\alpha = 0.7$ ($T = 5$). Les régions claires correspondent à la classe 1.

privilégier aucune classe.

Pour cela, il faut que la moitié de la population soit classée en ambiguïté. Ainsi, une valeur $\alpha = 0.38$ permet de classer 200 individus dans les classes de rejet d'ambiguïté, mais avec un taux de reconnaissance plus faible (82.5 % sur l'ensemble d'apprentissage). Les figures A.5 et A.6 représentent alors les iso-masses et les régions obtenues. La dernière figure peut être comparée à celle obtenue avec un seul classifieur Bayésien (cf. figure A.7) pour lequel le premier mode de la classe 1 est complètement ignoré.

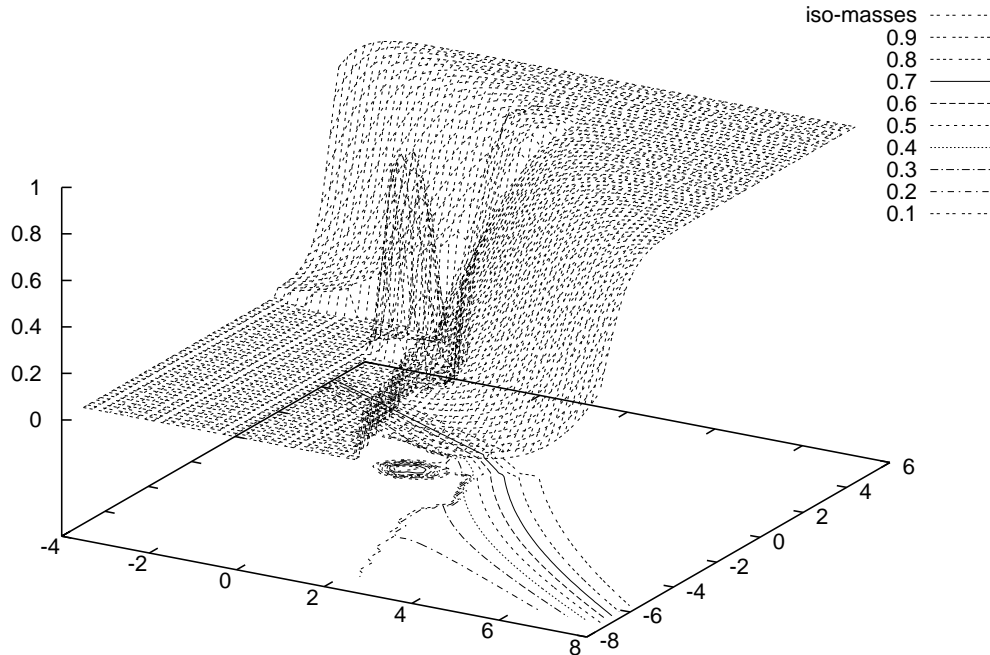


FIG. A.5 – Iso-masses de la classe ϕ_1 pour $\alpha = 0.38$ ($T = 5$).

A.5 Résultats sur un ensemble de test

Nous constituons un ensemble de test en sélectionnant les 200 individus suivants de chaque classe (les 200 premiers ont été utilisés pour l'apprentissage). Le taux de reconnaissance global obtenu est de 85,75 %. Des taux plus précis (par classes) sont données dans la Table A.3. La Table A.4 montre que le résultat obtenu est bien meilleur que celui d'un seul classifieur (ici, un classifieur Bayésien).

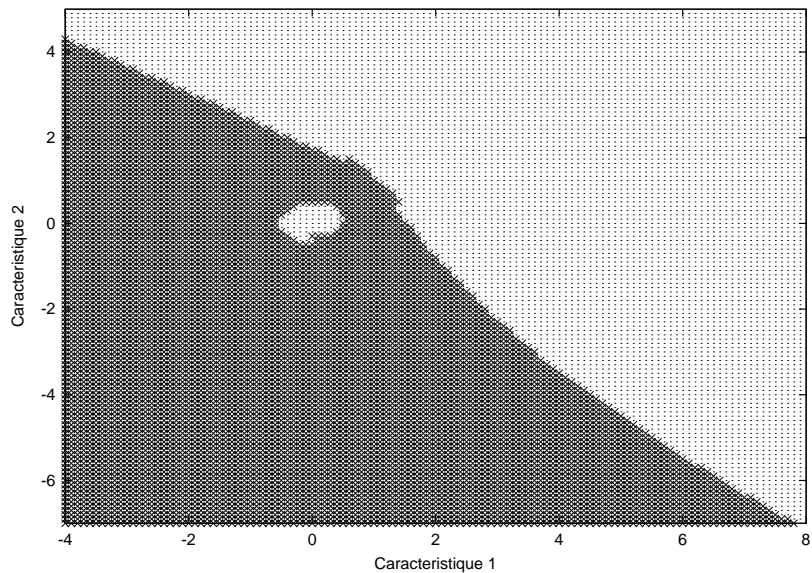


FIG. A.6 – Régions/frontières obtenues pour $\alpha = 0.38$ ($T = 5$). Les régions claires correspondent à la classe 1.

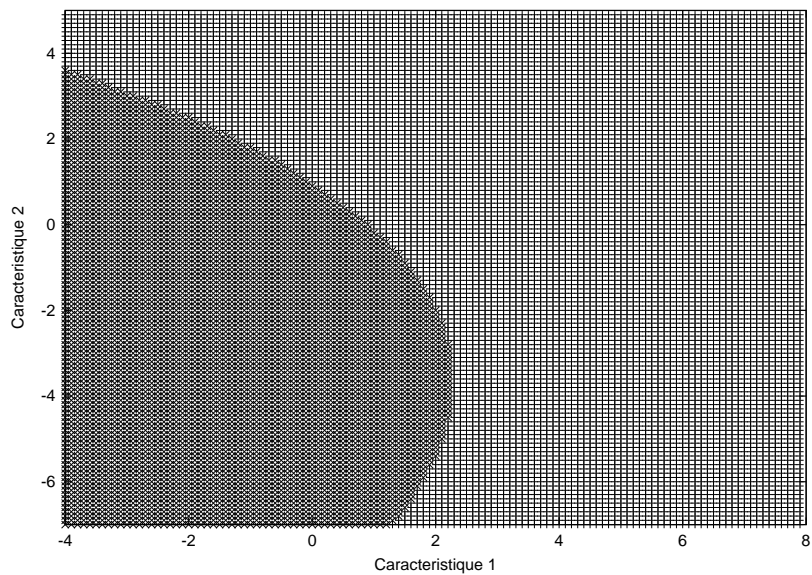


FIG. A.7 – Régions/frontières obtenues avec un seul classifieur Bayésien. Les régions claires correspondent à la classe 1. On constate que le premier mode de la classe 1 est complètement ignoré.

| | | |
|------------|----------|----------|
| \nearrow | ϕ_1 | ϕ_2 |
| ϕ_1 | 95.0 | 5.0 |
| ϕ_2 | 23.5 | 76.5 |

TAB. A.3 – Matrice de confusion sur un ensemble de test (pour $c = 3$), avec un taux de reconnaissance global de 85.75 %.

| | | |
|-----------|-------|----------|
| | Bayes | 3C (c=3) |
| Succès | 72.75 | 85.75 |
| Confusion | 27.25 | 14.25 |

TAB. A.4 – Comparaison avec le classifieur Bayésien : le mauvais résultat du classifieur Bayésien vient du fait que la première classe comporte 3 modes.

Annexe B

Robustesse de l'algorithme RCD-RPF

B.1 Robustesse aux chevauchements des classes

Lorsqu'il y a chevauchement des classes de telle manière que deux classes soient partiellement confondues dans l'espace de représentation, les états générés ne suffisent plus (pour cela, nous avons translaté de 20 les individus de la classe 1 selon la caractéristique 2). La figure B.1 illustre ce cas (position 156). On remarque bien que le taux de reconnaissance en apprentissage (figure B.2) est acceptable, mais lorsqu'il s'agit de classer des individus inconnus, la chute du taux de reconnaissance est assez brutale (le taux devient parfois inférieur à 70 %, figure B.3). L'algorithme est donc assez sensible aux ambiguïtés.

Cependant, dans les applications réelles, les formes des classes devraient peu varier d'une période à une autre (pour une position donnée). Dans ce cas, un algorithme tel que celui proposé dans la chapitre 3 convient mieux qu'un seul classifieur. Il faudra alors déterminer des critères de choix de classifieurs ou de systèmes de classifieurs lors de la génération de chaque nouvel état.

B.2 Robustesse aux évolutions des trajectoires

D'une période à une autre, la trajectoire des classes peut évoluer. L'étude ci-dessous donne une idée de la robustesse du système proposé à ces évolutions. Pour cela, nous avons translaté (dans la phase de classement) les centres des classes 1 et 2 respectivement de +5 et de -5 (selon la caractéristique 1, cf. figure B.4).

L'évolution du taux de reconnaissance en apprentissage reste inchangée (figure 4.7), puisque les individus ayant servis à l'apprentissage sont les mêmes que dans l'étude initiale. Lorsqu'il s'agit de classer des individus inconnus, la confusion augmente du fait qu'il y a chevauchement partiel des classes sur une partie de la trajectoire. Bien qu'il y ait une baisse du taux de reconnaissance (jusqu'à devenir inférieur à 80 %, figure B.5), il faut préciser que relativement aux dispersions des classes ($\sigma = 2$), la modification effectuée sur la trajectoire est assez importante.

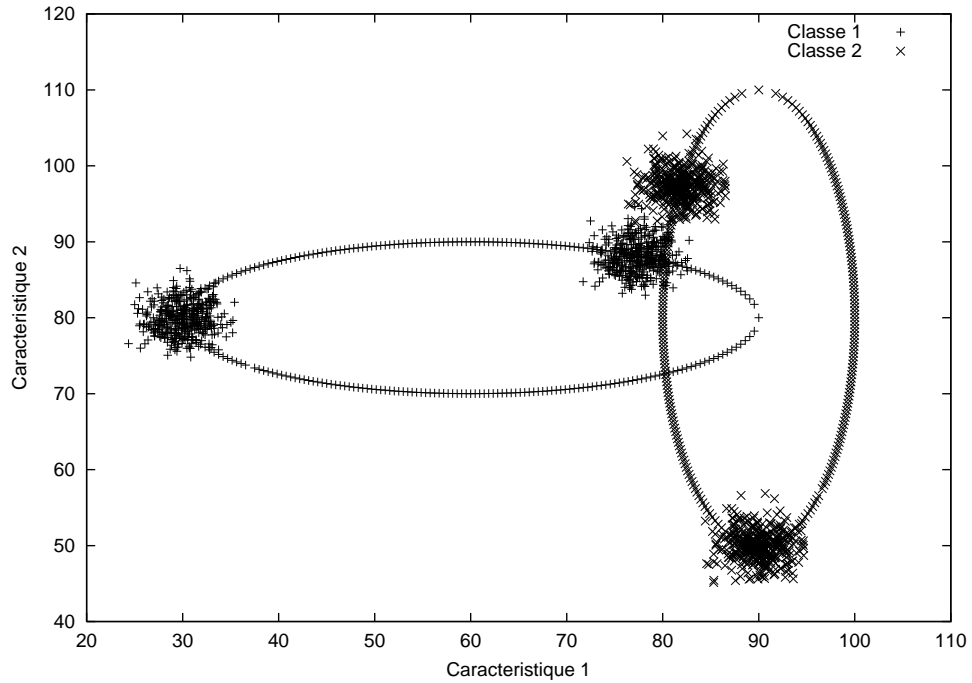


FIG. B.1 – Les deux classes dans la position initiale et dans la position 156 (avec la trajectoire de chaque classe).

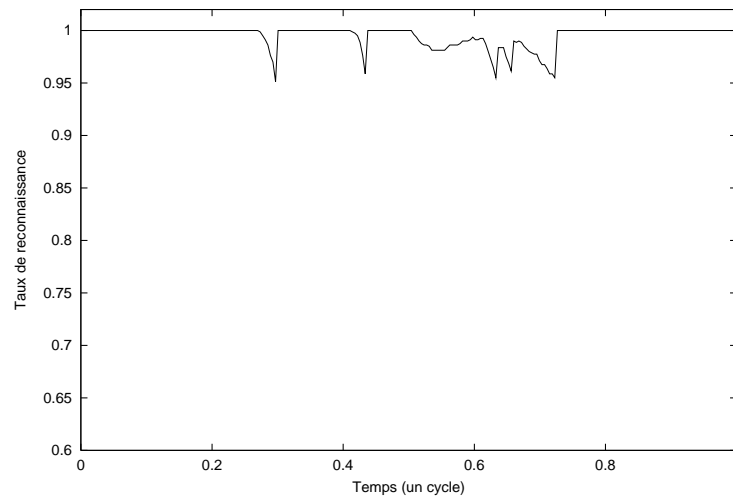


FIG. B.2 – Evolution du taux de reconnaissance en apprentissage lorsqu'il y a chevauchement des classes.

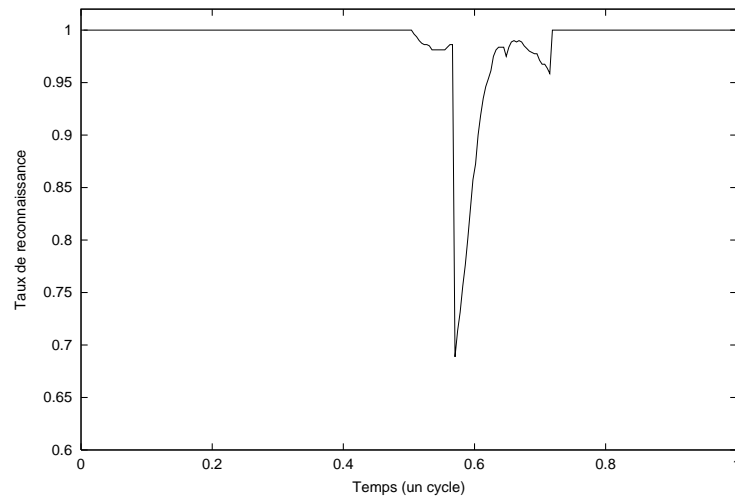


FIG. B.3 – Evolution du taux de reconnaissance en classement lorsqu’il y a chevauchement des classes.

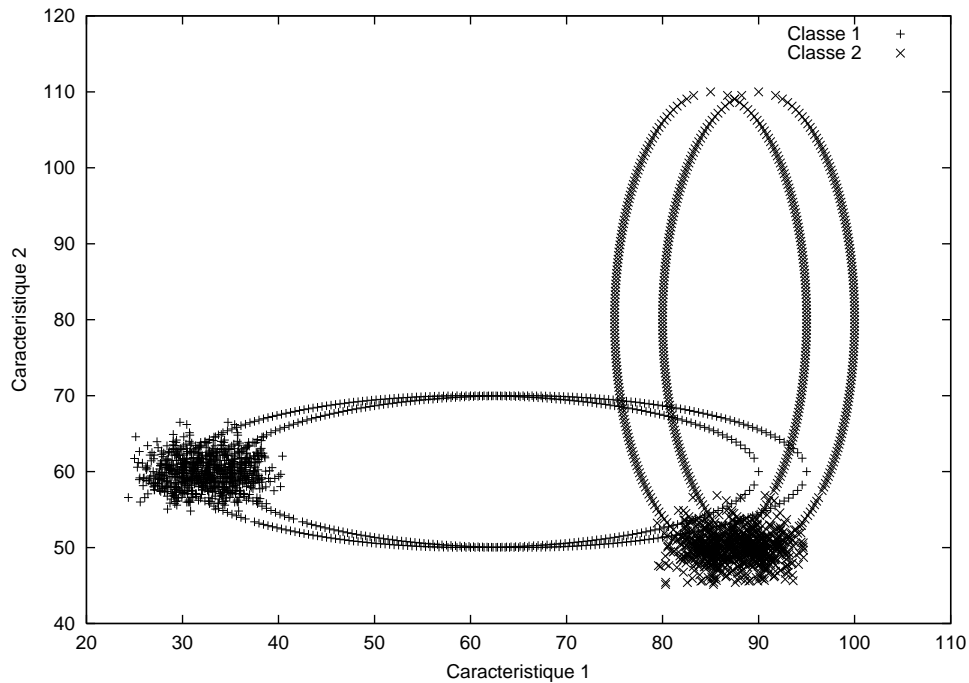


FIG. B.4 – Les trajectoires de classes lors de l’apprentissage et lors du classement (+5 et -5 sur la caractéristique 1 de la classe 1 et 2 respectivement). Les deux classes sont représentées dans la position initiale en apprentissage et en classement.

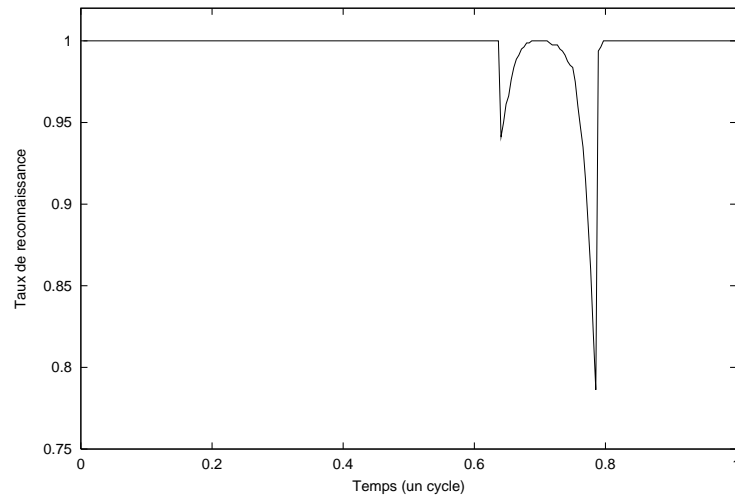


FIG. B.5 – Evolution du taux de reconnaissance en classement lorsque les trajectoires évoluent.

Index

| | | |
|--|---------------------------------|--------|
| A | Fusion | 38 |
| A posteriori | 62 | |
| A priori | 62 | |
| Adaptative | 47 | |
| Ambiguïté | 81 | |
| B | G | |
| Bayésien (classement) | 61 | |
| | Gaussienne | 62 |
| C | H | |
| C moyennes floues | 81 | |
| C+2 moyennes floues | 86 | |
| Cadre de discernement | 25, 27 | |
| Caractéristiques | 25 | |
| Classe de Coalescence (CC) | 78 | |
| Classe de Coalescence d'Ambiguïté (rejet) .. | 78 | |
| Classe de rejet | 81 | |
| Classes dynamiques | 55, 67, 108 | |
| Classes statiques | 56 | |
| Classifieur | 28 | |
| Coalescence floue | 79 | |
| Combinaison | 38, 44 | |
| Coopération | 45, 78 | |
| Croyance | 43 | |
| D | I | |
| Décision (règle) | 41 | |
| Distance | 81, 85 | |
| | Ignorance | 19 |
| E | J | |
| Ensembles de classifieurs | 36 | |
| Etat | 112 | |
| Evidence | 42 | |
| | Jeton | 115 |
| F | M | |
| Facteur de Certitude | 110 | |
| Fonction objective | 81 | |
| Fonctions d'appartenance | 86 | |
| Fonctions d'appartenances | 114 | |
| Forme | 25 | |
| | Masse | 43 |
| | Matrice de confusion | 37 |
| | O | |
| | Objets dynamiques | 55, 57 |
| | Objets statiques | 56 |
| | P | |
| | P-invariant | 109 |
| | Paramètre | 25 |
| | Performances | 37 |
| | Places | 110 |
| | Plausibilité | 43 |
| | Propositions | 110 |
| | R | |
| | Réseau de Petri Flou | 110 |
| | Règle de production floue | 110 |
| | Règle orthogonale | 44 |
| | Raisonnement | 110 |
| | RdF | 19, 25 |
| | Reconnaissance | 25 |
| | Rejet d'ambiguïté | 81 |
| | S | |
| | Sélection adaptative | 47 |

Sélection de classifieurs.....46
Sélection dynamique 47
Sélection statique 47
Système de classifieurs 32, 114

T

Taux de confusion.....38
Taux de reconnaissance 38
Taxonomies 28, 36
Transition.....110, 115

V

Vecteurs de Formes 26
Vote 42

Abstract

In the field of Pattern Recognition (PR), imperfect information handling is often carried out by applying several theories dealing with its imprecision, its uncertainty, its incompleteness and/or its contradictions. The decision making would then be influenced by this imperfection. An alternative to this handling consists of reducing the effect of this unperfectness by associating several classifiers through several strategies.

Combination is one of the possible strategies for the association of several classifiers. However, it is also possible to imagine two other strategies : the first one is to involve several classifiers into cooperation in order to make a final decision, with the possibility that the decision of a classifier could be influenced by another classifier decision. The second consists in selecting the decisions of one or several classifiers according to the situation and/or to time. We propose a *classifier system* approach which can involve one or several of these strategies.

In the PR framework, we propose, in this thesis, some approaches and some solutions involving the combination, the cooperation and/or the selection of classifiers strategies, while bearing in mind the temporal aspect of the evolving natural objects features. Through the study of the static and dynamic aspects of PR, we suggest that to recognize dynamic classes, there are two possible approaches.

When the trajectories of the classes do not intersect and these classes are multi-modal, our approach consists of transforming these dynamic classes into static ones. The classes have then complex shapes. To deal with such classes, an algorithm of classifiers cooperation is proposed. It involves an unsupervised PR method which is able to carry out an adaptive classifier selection and several supervised PR methods.

When there are no intersections and the classes evolve in time in a continuous way, the proposed approach consists in transforming the PR system in a dynamic one. A method, based on the modelisation of the system states changes with a fuzzy Petri net, is proposed. The fuzzy specification of these changes allows the prediction of the system states which are the most adapted to the PR problem, at the involved instant.

Keywords :

| | | |
|----------------------------|----------------------------|--------------------------|
| Pattern recognition | Classifier systems | Fuzzy Petri nets |
| Combination of classifiers | Cooperation of classifiers | Selection of classifiers |
| Fuzzy clustering | Ambiguity rejection | Fuzzy sets |

Résumé

Lorsque plusieurs classifieurs sont amenés à concourir à une même tâche de reconnaissance, plusieurs stratégies de décisions, impliquant ces classifieurs de différentes manières, sont possibles. Une première stratégie consiste à décider suite à différents avis : il s'agit de la combinaison de classifieurs. Une deuxième stratégie consiste à utiliser un ou plusieurs avis pour mieux guider d'autres classifieurs dans leurs phases d'apprentissages, et à utiliser un ou plusieurs avis pour améliorer la prise de décisions d'autres classifieurs dans la phase de classement : il s'agit de la coopération de classifieurs. Enfin, la troisième et dernière stratégie consiste à privilégier un ou plusieurs classifieurs en fonction de divers critères ou en fonction de la situation : il s'agit de la sélection de classifieurs.

L'aspect temporel de la RdF, c'est-à-dire l'évolution possible des classes à reconnaître, est traité par la stratégie de la sélection. En étudiant les aspects statiques et dynamiques de la RdF, nous montrons que pour reconnaître des classes dynamiques, deux approches sont possibles. Ces deux approches sont validées sur un ensemble de test.

Dans le cas où les trajectoires des classes ne s'intersectent pas et que ces classes sont multimodales, l'approche proposée consiste à transformer ces classes dynamiques en classes statiques. En intégrant l'évolution de ces classes dans le temps, les classes obtenues deviennent alors complexes. Pour traiter ce type de classes, un algorithme de coopération des classifieurs est proposé. Il met en œuvre, d'une part, une méthode de classification non-supervisée effectuant une sélection adaptative de classifieurs et, d'autre part, plusieurs méthodes de RdF supervisées.

Lorsqu'il n'y a pas d'intersection et que les classes évoluent de manière continue dans le temps, l'approche proposée consiste à rendre dynamique le système de RdF. Une méthode, fondée sur la modélisation des changements d'états du système par un réseau de Petri flou, est proposée. La méthode permet de prédire le ou les états du système les mieux adaptés au problème de RdF, à l'instant considéré.

Discipline : Informatique

Mots-clés :

| | | |
|-----------------------------|-----------------------------|---------------------------|
| Reconnaissance des formes | Systèmes de classifieurs | Réseaux de Petri flous |
| Combinaison de classifieurs | Coopération de classifieurs | Sélection de classifieurs |
| Coalescence floue | Rejet d'ambiguïté | Ensembles flous |

Université de La Rochelle - Laboratoire L3i (UPRES EA 2118)