

Regroupement d'attributs par règles d'association dans les systèmes d'inférence floue

Ilef Ben Slima*, Amel Borgi**

*Université de Tunis El Manar, Faculté des Sciences de Tunis,
LR11ES14 LIPAH, Tunis, Tunisie
ilef.benslima@fst.rnu.tn,

**Université de Tunis El Manar, Institut Supérieur d'Informatique,
Faculté des Sciences de Tunis, LR11ES14 LIPAH, Tunis, Tunisie
Amel.Borgi@insat.rnu.tn

Résumé. Dans les systèmes d'apprentissage supervisé par construction de règles de classification floues, un nombre élevé d'attributs descriptifs conduit à une explosion du nombre de règles générées et peut affecter la précision des algorithmes d'apprentissage. Afin de remédier à ce problème, une solution est de traiter séparément des sous-groupes d'attributs. Cela permet de décomposer le problème d'apprentissage en des sous-problèmes de complexité inférieure, et d'obtenir des règles plus intelligibles car de taille réduite. Nous proposons une nouvelle méthode de regroupement des attributs qui se base sur le concept des règles d'association. Ces règles découvrent des relations intéressantes entre des intervalles de valeurs des attributs. Ces liaisons locales sont ensuite agrégées au niveau des attributs mêmes en fonction du nombre de liaisons trouvées et de leur importance. Notre approche, testée sur différentes bases d'apprentissage et comparée à l'approche classique, permet d'améliorer la précision tout en garantissant une réduction du nombre de règles.

1 Introduction

Ce travail s'inscrit dans le cadre de l'apprentissage supervisé et plus précisément des systèmes de classification à base de règles floues (Ishibuchi et al. (1992, 1994)). Ces systèmes ont la spécificité d'être facilement interprétables grâce à l'utilisation de termes linguistiques.

Dans ces systèmes, les règles floues peuvent être fournies par un expert humain. Comme l'acquisition des connaissances humaines est une tâche complexe, plusieurs travaux se sont consacrés à l'automatisation de la construction des règles à partir des données numériques (Ishibuchi et al., 1992), (Dehzangi et al., 2007). Cette construction comprend deux phases : une partition floue de l'espace des entrées puis la construction d'une règle floue pour chaque sous-espace flou issu de cette partition. Dans ces systèmes, un nombre élevé d'attributs conduit à une explosion du nombre de règles générées, ce qui entraîne une dégradation de la compréhensibilité des systèmes, et affecte le temps de réponse nécessaire aussi bien à la phase d'apprentissage qu'à la phase de classification.

De ce fait, l'optimisation du nombre de règles floues ainsi que du nombre d'antécédents paraît comme une clé pour améliorer les systèmes de classification à base de règles floues. Dans ce cadre, plusieurs approches ont été proposées dans la littérature. On peut citer l'approche de sélection des règles pertinentes par algorithme génétique (Ishibuchi et al., 1995) ou par le concept d'oubli (Nozaki et al., 1994). Une autre approche consiste à réduire le nombre d'attributs par une sélection des attributs les plus significatifs (Lee et al., 2001).

Dans ce papier, nous nous intéressons à la technique de regroupement des attributs dans les prémisses des règles. Dans cette approche, initialement introduite dans un cadre non flou par Borgi (1999), les attributs prédictifs sont regroupés en blocs, les attributs de chaque bloc sont traités séparément et apparaissent ensemble dans une même prémisse. Une première extension de ce travail dans un cadre flou, pour la génération de règles dans les systèmes d'inférence floue, a été réalisée par Soua et al. (2012). Cette approche de regroupement des attributs, nommée SIFCO, présente l'avantage de décomposer le problème d'apprentissage en des sous-problèmes de complexité inférieure, et de réduire ainsi le nombre de règles générées. De plus, cette approche permet d'obtenir des règles plus intelligibles car de taille réduite.

Le regroupement d'attributs dans (Borgi, 1999) et (Soua et al., 2012) se fait par recherche de corrélation linéaire : les attributs linéairement corrélés sont regroupés et traités séparément. Dans cet article, nous proposons une méthode qui se base sur le concept des règles d'association (RA) introduit par Agrawal et al. (1993). Les RA vont nous permettre de déterminer les attributs "liés" ou "associés" qui seront regroupés dans les mêmes règles.

L'article est organisé comme suit : dans la partie 2, nous présentons les systèmes de classification à base de règles floues. Nous décrivons, dans la partie 3, le principe de regroupement des attributs comme présenté dans (Borgi, 1999) et (Soua et al., 2012). Notre approche de regroupement des attributs par RA est décrite dans la partie 4 et les résultats des tests expérimentaux sont présentés dans la partie 5. Nous concluons l'article en présentant les principales perspectives de ce travail.

2 Apprentissage à base de règles floues

On se place dans le cadre des problèmes de classification supervisée dont le but est d'affecter une classe à un objet décrit par des variables descriptives (des attributs). Nous nous intéressons au système de classification floue proposé dans (Ishibuchi et al., 1992). Afin de simplifier les notations, nous désignons ce système par l'acronyme SIF. Deux phases sont à distinguer dans ce système : la phase d'apprentissage dans laquelle on construit le modèle de classement à partir des données d'apprentissage, et la phase de classification qui sert à associer une classe à un objet inconnu en utilisant ce modèle.

2.1 Phase d'apprentissage

La méthode de génération des règles floues que nous adoptons correspond à l'utilisation d'une grille floue simple, proposée par Ishibuchi et al. (1992). Pour illustrer cette approche, nous supposons, par souci de clarté, que notre problème d'apprentissage est un problème bi-dimensionnel (2 attributs : X_1 et X_2). Les m exemples d'apprentissage considérés sont notés $E_p = (X_{p1}, X_{p2})$; $p = 1, 2, \dots, m$, ils appartiennent chacun à l'une des C classes :

y_1, y_2, \dots, y_C . Il est à noter que dans les SIF les attributs considérés sont numériques ; chacun des attributs X_1 et X_2 est partitionné en k sous-ensembles flous $\{A_1, A_2, \dots, A_k\}$ où chaque sous-ensemble A_i est défini par une fonction d'appartenance triangulaire symétrique. Un exemple de grille floue simple est présenté dans Fig. 1, dans le cas d'une partition floue de taille 5 ($k = 5$) ; le nombre de régions floues est égal à 5^2 . Dans le cas général, pour un problème de n attributs et k partitions pour chaque attribut, on obtient une grille floue de k^n sous espaces flous. Pour chaque région $A_i \times A_j$ de la grille, une règle floue R_{ij}^k est construite :

$$R_{ij}^k : \text{Si } X_1 \text{ est } A_i \text{ et } X_2 \text{ est } A_j \text{ alors } X = (X_1, X_2) \text{ appartient à } y_{ij}^k \text{ avec } CF_{ij}^k \quad (1)$$

où $i, j = 1, 2, \dots, k$

- y_{ij}^k est la conclusion de la règle, elle correspond à l'une des C classes
- CF_{ij}^k est le degré de certitude de la règle, il traduit sa validité.

La conclusion et le degré de certitude de chaque règle sont déterminés comme suit :

1. Pour chaque classe y_t , calculer la somme des compatibilités des exemples d'apprentissage appartenant à cette classe, par rapport à la prémisse de la règle :

$$\beta_{y_t} = \sum_{Y(X_p)=y_t} \mu_i^k(X_{p1}) * \mu_j^k(X_{p2}) \quad ; t = 1, 2, \dots, C \quad (2)$$

- $X_p = (X_{p1}, X_{p2})$: un exemple d'apprentissage
 - $Y(X_p)$: la classe associée à X_p
 - μ_i^k, μ_j^k : les fonctions d'appartenance respectives de A_i et A_j
2. Trouver la classe y_a qui a la plus grande valeur de compatibilité

$$\beta_{y_a} = \max\{\beta_{y_1}, \beta_{y_2}, \dots, \beta_{y_C}\} \quad (3)$$

3. Déterminer le degré de certitude CF_{ij}^k

$$CF_{ij}^k = \frac{|\beta_{y_a} - \beta|}{\sum_{t=1}^C \beta_{y_t}} \text{ avec } \beta = \sum_{y_t \neq y_a} \frac{\beta_{y_t}}{C-1} \quad (4)$$

Dans les travaux portant sur la construction de règles de classification floues, les attributs ne sont pas nécessairement partitionnés en un même nombre de sous-ensembles flous. Plusieurs types de grilles floues ont été étudiés comme par exemple la grille floue rectangulaire (Ishibuchi et al., 1994).

2.2 Phase de classification

Dans cette phase, le système décide, à partir de la base de règles générées, notée S_R , de la classe y_a à associer à un individu $E' = (X'_1, X'_2)$ de classe inconnue.

1. Pour chaque classe y_t ; $t = 1, 2, \dots, C$, calculer α_{y_t} par :

$$\alpha_{y_t} = \max\{\mu_i^k(X'_1) * \mu_j^k(X'_2) * CF_{ij}^k / y_{ij}^k = y_t \text{ et } R_{ij}^k \in S_R\} \quad (5)$$

2. Trouver la classe y_a qui maximise α_{y_t} :

$$\alpha_{y_a} = \max\{\alpha_{y_1}, \alpha_{y_2}, \dots, \alpha_{y_C}\} \quad (6)$$

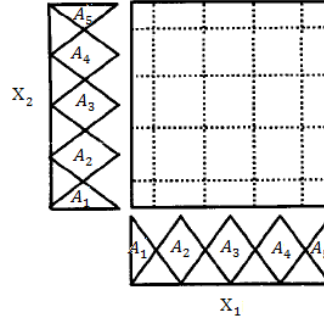


FIG. 1 – Grille floue simple.

3 Regroupement d'attributs

L'approche de regroupement d'attributs se base sur le concept des ensembles d'apprentissage artificiel, qui repose sur la combinaison des décisions de plusieurs apprenants pour améliorer l'exécution du système global (Valentini et Masulli, 2002). L'idée de ce concept est de répartir l'information - qui peut correspondre aux exemples d'apprentissage, aux attributs descriptifs ou encore aux classes - entre plusieurs apprenants, chaque apprenant réalise la phase d'apprentissage sur l'information qui lui a été fournie, et les opinions "individuelles" des différents apprenants sont ensuite combinées pour atteindre une décision finale. Dans notre cas, l'information à répartir correspond aux attributs descriptifs : chaque classifieur utilise un sous-ensemble des attributs initiaux et construit une base de règles locale, puis les différentes bases locales obtenues sont combinées pour former le modèle final (voir Fig. 2).

Cette approche, vérifiée expérimentalement dans (Soua et al., 2012) et (Borgi, 1999), permet de garantir une réduction conséquente du nombre de règles sans trop altérer les taux de bonnes classifications. Pour un problème de n attributs et k sous ensembles flous pour chaque attribut, le nombre de règles générés par les SIF, noté NR_{SIF} , vaut k^n . Lorsqu'on découpe le problème d'apprentissage en g sous-problèmes et on applique sur chacun d'eux la même démarche de génération de règles que les SIF, on obtient un nombre de règles NR_{regrp} égal à :

$$NR_{regrp} = \sum_{i=1}^g k^{n_i} \quad (7)$$

où n_i est le nombre d'attributs liés dans le $i^{\text{ème}}$ groupe g_i .

Il a été démontré dans (Soua et al., 2012) que si les groupes d'attributs, issus de l'approche de regroupement, forment une partition de l'ensemble des attributs de départ ($\sum_{i=1}^g n_i = n$), alors :

$$NR_{regrp} = \sum_{i=1}^g k^{n_i} \leq \prod_{i=1}^g k^{n_i} = k^{\sum_{i=1}^g n_i} = k^n = NR_{SIF} \quad (8)$$

Par conséquent :

$$NR_{regrp} \leq NR_{SIF} \quad \text{si} \quad \sum_{i=1}^g n_i = n \quad (9)$$

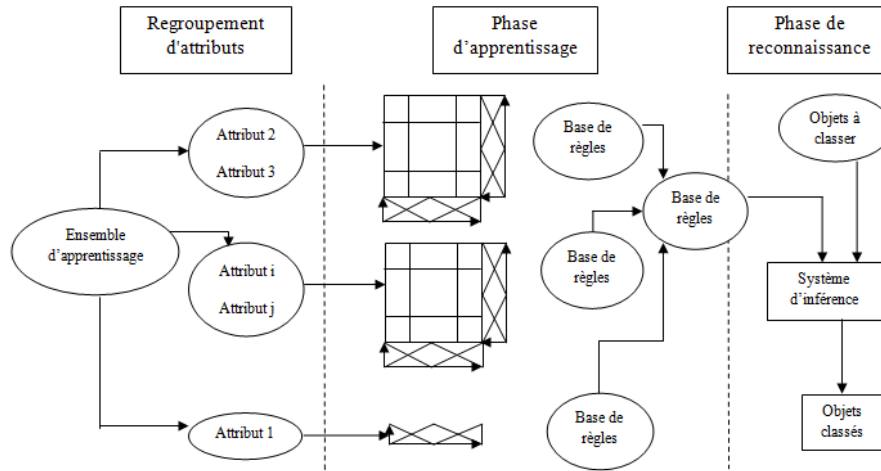


FIG. 2 – Approche de génération de règles par regroupement des attributs (Soua et al., 2012).

4 Approche proposée : regroupement des attributs par RA

Notre contribution réside au niveau de la méthode de regroupement d'attributs ; nous proposons une nouvelle méthode n'utilisant pas la recherche de corrélation linéaire, mais qui se base sur le concept des règles d'association (Agrawal et al., 1993). Les algorithmes d'extraction des RA déterminent les associations intéressantes entre les attributs en analysant leurs apparitions simultanées dans les enregistrements de la base de données. Cette méthode peut être très intéressante pour les bases de données pour lesquelles il n'existe aucune relation de type corrélation linéaire entre les attributs.

Généralement, les algorithmes d'extraction des RA déterminent les associations entre des variables de type booléen. Comme les données traitées dans les SIF sont quantitatives, il est nécessaire de commencer par les transformer en des valeurs booléennes, puis d'appliquer le concept des RA sur ces valeurs. Ensuite, à partir des associations trouvées entre ces valeurs booléennes, nous déduisons les associations entre les attributs de départ. Enfin et dans le but de garantir une réduction du nombre de règles, nous nous proposons de filtrer les groupes d'attributs associés de manière à obtenir une partition de l'ensemble des attributs de départ. Nous décrivons dans ce qui suit ces différentes étapes.

4.1 Génération des itemsets fréquents : liaisons locales entre attributs

L'existence d'une liaison entre deux variables dépend de la réponse à la question : est-ce que la connaissance des valeurs de l'une permet de prédire les valeurs de l'autre ? Le concept des RA répond à cette question en associant les valeurs qui apparaissent souvent ensemble dans les transactions de la base de données considérée.

Les RA ont été introduites par Agrawal et al. (1993) dans le but d'analyser les transactions de ventes dans un supermarché pour déterminer les articles souvent achetés ensemble. Une

Regroupement d'attributs par application des règles d'association

transaction consiste en un ensemble d'articles achetés par un client ; un article est appelé item ou attribut. Soit D une base de données transactionnelle ; $I = \{I_1, I_2, \dots, I_n\}$ l'ensemble d'attributs ou d'items. Un sous ensemble X de I est appelé motif ou itemset. Chaque itemset X est caractérisé par un support qui est la fraction des transactions de D qui contiennent l'itemset X . Une RA est une implication $X \rightarrow Y$ où X et Y sont deux itemsets et $X \cap Y = \emptyset$. Cette implication signifie que les transactions qui vérifient l'itemset X ont tendance à vérifier également l'itemset Y . Une RA est caractérisée par deux mesures : support et confiance.

- Le support de $X \rightarrow Y$ est la fraction de la base qui contient simultanément X et Y . Il est égal au support de l'itemset $X \cup Y$.
- La confiance de $X \rightarrow Y$ est la proportion des transactions qui contiennent Y sachant qu'elles contiennent aussi X .

Le problème de découverte des RA est décomposé en deux sous-problèmes :

- la génération des itemsets fréquents (tous les itemsets ayant un support supérieur à un seuil prédéfini *minSupp*).
- la génération des règles d'association à partir de ces itemsets fréquents ; une RA doit avoir une confiance supérieure à un seuil prédéfini par l'utilisateur *minConf*.

Dans ce travail, nous nous intéressons au premier sous-problème et nous cherchons à déterminer les groupes d'attributs liés. Pour déterminer les itemsets fréquents, plusieurs algorithmes ont été proposés (Agrawal et al., 1993) (Agrawal et Srikant, 1994) (Savasere et al., 1995). L'algorithme Apriori proposé dans (Agrawal et Srikant, 1994) est le plus connu et il est largement utilisé mais il ne traite que des données booléennes. Dans les problèmes courants, la majorité des données sont quantitatives et qualitatives et nécessitent des algorithmes applicables à ce type de données. Une extension de Apriori a été proposée par Srikant et Agrawal (1996) ; ils ont proposé de faire une correspondance entre des variables quantitatives ou qualitatives et des variables booléennes par le codage disjonctif complet. Pour une variable qualitative, chaque catégorie correspond à un élément booléen. Pour une variable quantitative, on discrétise l'attribut en des intervalles, puis on fait correspondre une variable booléenne à chaque intervalle.

Dans notre cas, les attributs étant continus, nous recourons au codage disjonctif complet des attributs. Le partitionnement des attributs se fait par une discrétisation régulière à intervalles égaux. Nous obtenons donc des intervalles que nous assimilons à des valeurs booléennes. Nous appliquons ensuite l'algorithme Apriori sur ces intervalles et obtenons ainsi des itemsets fréquents ou des groupes d'intervalles liés.

4.2 Détermination des attributs liés : liaisons globales entre attributs

Dans l'étape précédente, nous avons déterminé les groupes d'intervalles liés. Notre but étant de faire un regroupement des attributs et non pas de leurs intervalles, on se propose de développer une procédure qui permet de déterminer la liaison entre un groupe d'attributs à partir des liaisons trouvées entre leurs intervalles.

Nous définissons pour cela une grille d'association qui représente les associations entre les valeurs (intervalles) d'un groupe d'attributs. Chaque axe de la grille concerne un attribut. La Fig. 3 présente 3 exemples de grille avec deux attributs X_1 et X_2 ; X_1 est décomposé en 6 valeurs ($val_{X_1}^1, \dots, val_{X_1}^6$), et X_2 en 5 valeurs ($val_{X_2}^1, \dots, val_{X_2}^5$). Dans cet exemple, les itemsets obtenus à l'issue de l'application de l'algorithme Apriori concernent des couples

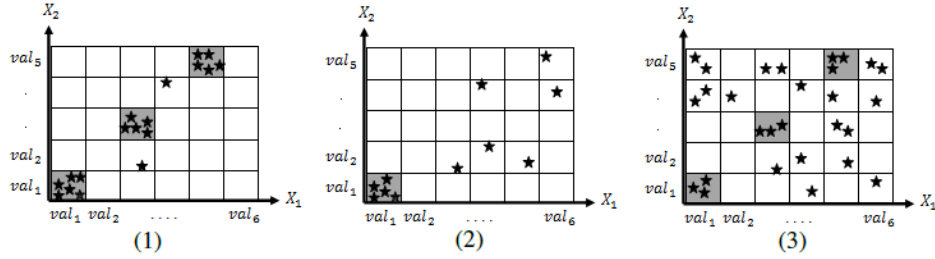


FIG. 3 – Trois exemples de grille d'association.

de valeurs des attributs X_1 et X_2 . Quand deux valeurs forment un itemset fréquent, la case correspondante à leur intersection est grisée : on appelle cette case une région liée.

La liaison entre deux valeurs de deux attributs n'entraîne pas forcément la liaison entre les deux attributs puisque d'une part, ces attributs peuvent avoir très peu de régions liées (exemple (2) de la Fig. 3) et d'autre part, le nombre de données dans ces régions peut être très faible par rapport au nombre total de données (exemple (3) de la Fig. 3).

Dire que, plus le nombre de régions liées est grand, plus l'association entre les attributs est forte, n'est pas toujours suffisant. En effet, le principe de RA détermine si une région est liée en analysant son support, et ce dernier reflète la densité de données, c.à.d. la fréquence d'apparition des données dans la région. Ainsi, une seule région liée peut entraîner une association plus significative que plusieurs régions liées si cette unique région a une densité plus importante que la densité totale de l'ensemble des autres régions liées. Nous proposons donc de prendre en compte aussi bien le nombre de régions liées que leurs densités. Pour cela, nous commençons par définir le poids d'une région, appelé aussi coefficient de pondération. Ce poids caractérise la densité de données dans cette région, ce qui revient à son support.

Pour les valeurs respectives $val_{X_1}^i$ et $val_{X_2}^j$ des attributs X_1 et X_2 , et pour un nombre total m d'exemples d'apprentissage, le coefficient de pondération w est :

$$w_{X_1 X_2}^{ij} = \frac{\text{card}(val_{X_1}^i \times val_{X_2}^j)}{m} = \text{support}(val_{X_1}^i \cup val_{X_2}^j) \quad (10)$$

où $\text{card}(val_{X_1}^i \times val_{X_2}^j)$ est le nombre d'enregistrements qui vérifient $val_{X_1}^i$ et $val_{X_2}^j$.

Nous nous inspirons ensuite du principe des RA généralisées où une taxonomie (Fig. 4) existe entre les variables. D'après (Srikant et Agrawal, 1995), les associations trouvées à un niveau donné peuvent remonter au niveau supérieur en sommant leurs supports, à condition qu'il n'y ait pas de recouvrement. Avec l'exemple de la Fig. 4, si les itemsets (*Veste, Botte*) et (*Veste, Espadrille*) sont extraits, alors il n'est pas possible de les généraliser à l'itemset de niveau supérieur (*Vêtements, Chaussure*) en sommant leurs supports, car *Veste, Botte* et *Espadrille* peuvent figurer dans une même transaction. Dans notre cas, les variables quantitatives sont partitionnées en des valeurs sous forme d'intervalles ; la présence de deux valeurs d'un seul attribut n'est donc pas possible dans le même enregistrement. En formant une taxonomie entre un attribut et ses intervalles (Fig. 5), et comme il n'y a pas de recouvrement entre les

Regroupement d'attributs par application des règles d'association

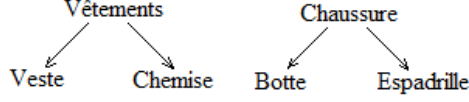


FIG. 4 – Exemple de taxonomie pris de Srikant et Agrawal (1995).

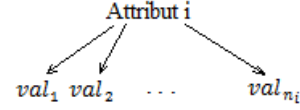


FIG. 5 – Taxonomie entre un attribut et ses valeurs.

intervalles, on peut calculer le degré d'association des attributs comme la somme des supports de leurs valeurs (intervalles) liées.

En utilisant ce principe, et en ne comptabilisant que les régions liées, nous définissons un degré d'association entre deux attributs ou plus, par la somme des coefficients de pondération de leurs régions liées. Donc, pour deux attributs X_1 et X_2 , le degré d'association β s'écrit :

$$\beta_{X_1 X_2} = \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} w_{X_1 X_2}^{ij} * r_{X_1 X_2}^{ij} \quad (11)$$

- $r_{X_1 X_2}^{ij}$ est la région formée par les valeurs $val_{X_1}^i$ et $val_{X_2}^j$; $r_{X_1 X_2}^{ij} = \begin{cases} 1 & \text{si région liée} \\ 0 & \text{sinon} \end{cases}$
- k_1 et k_2 sont respectivement les tailles des partitions des attributs X_1 et X_2 .

Dans le cas général d'un ensemble d'attributs $X = \{X_{n_1}, X_{n_2}, \dots, X_{n_l}\}$, le degré d'association de ces l attributs est :

$$\beta = \sum_{i_1=1}^{k_1} \sum_{i_2=1}^{k_2} \dots \sum_{i_l=1}^{k_l} w_X^{i_1 i_2 \dots i_l} * r_X^{i_1 i_2 \dots i_l} \quad (12)$$

$$w_X^{i_1 i_2 \dots i_l} = \frac{\text{card}(val_{X_{n_1}}^{i_1} \times val_{X_{n_2}}^{i_2} \times \dots \times val_{X_{n_l}}^{i_l})}{m} = \text{support}(val_{X_{n_1}}^{i_1} \cup \dots \cup val_{X_{n_l}}^{i_l}) \quad (13)$$

- $r_X^{i_1 i_2 \dots i_l}$ est la région formée par les intervalles $val_{X_{n_1}}^{i_1}, val_{X_{n_2}}^{i_2}, \dots, val_{X_{n_l}}^{i_l}$
- k_1, k_2, \dots, k_l sont respectivement les tailles des partitions de $X_{n_1}, X_{n_2}, \dots, X_{n_l}$.

Le degré β est compris entre 0 et 1, on peut alors définir un seuil d'association β_{\min} au delà duquel on considère que les attributs de l'ensemble X sont liés.

4.3 Choix des groupes d'attributs associés

La procédure présentée dans 4.1 et 4.2 est basée sur le principe de l'algorithme Apriori ; elle fournit donc tous les groupes d'attributs associés de différentes tailles. Il est à noter que ces groupes ne constituent pas forcément une partition de l'ensemble des attributs de départ : on peut avoir des relations d'inclusion entre deux groupes d'attributs, ou une intersection non vide. Afin de garantir une réduction du nombre de règles générées, nous nous proposons de sélectionner un ensemble de groupe d'attributs de manière à former une partition de l'ensemble des attributs de départ (voir partie 3). La sélection se base sur les deux critères suivants :

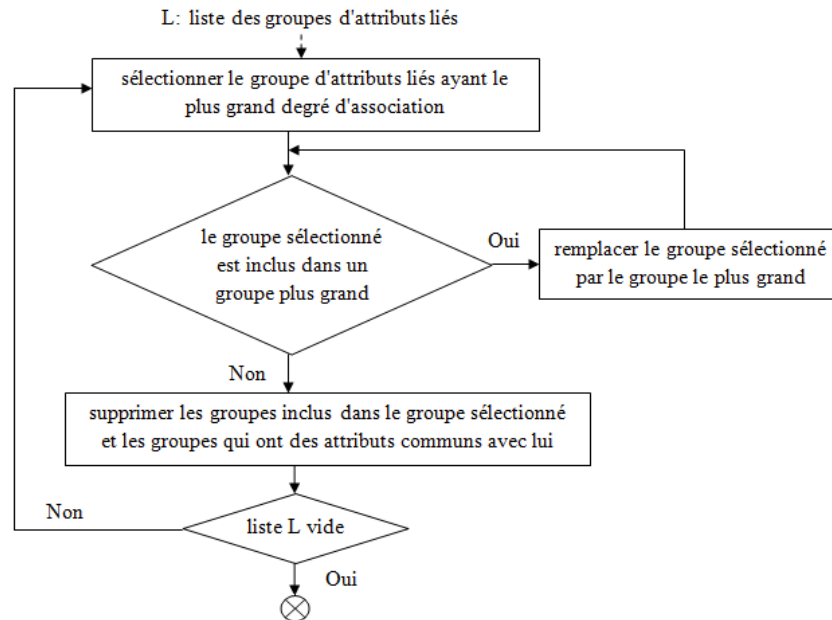


FIG. 6 – Algorithme de sélection des groupes d'attributs finaux.

1. plus le degré d'association β est grand, plus l'association est forte.
2. si un groupe d'attributs liés est inclus dans un autre groupe, alors on conserve le plus grand. On obtient ainsi des règles avec des prémisses de tailles plus longues ; Dehzangi et al. (2007) affirme que l'augmentation de la longueur des règles améliore la précision de classification.

Le diagramme de la Fig. 6 décrit l'algorithme de sélection des groupes d'attributs finaux. Les groupes d'attributs liés sélectionnés forment les groupes d'attributs finaux, auxquels on ajoute les groupes constitués d'un seul attribut qui n'ont figuré dans aucun groupe d'attributs liés et qui sont considérés comme indépendants.

5 Expérimentation

Notre système, baptisé SIFRA, utilise l'approche de regroupement des attributs dans le cadre des SIF (Ishibuchi et al., 1992) comme cela est fait dans SIFCO (Soua et al., 2012) mais avec une nouvelle méthode de regroupement des attributs, celle que nous avons proposée et qui se base sur le concept des règles d'association. Après avoir déterminé les groupes d'attributs associés, nous utilisons la démarche proposée par Ishibuchi et al. (1992) pour générer les règles floues (partie 2.1), pour chaque groupe d'attributs. La classification d'un objet inconnu se fait par la méthode de classification de Ishibuchi et al. (1992) (partie 2.2).

Nous avons testé notre système SIFRA sur des bases de données qui diffèrent par le nombre d'attributs, le nombre d'exemples et le nombre de classes (Tab. 1). Pour évaluer la capacité de

Regroupement d'attributs par application des règles d'association

généralisation de notre méthode, nous avons adopté la technique de validation croisée d'ordre 10 (Kohavi, 1995). Dans le tableau 2, nous présentons le taux de bonne classification suivi entre parenthèses du nombre de règles générées. Les meilleurs taux de classification sont présentés en gras. Le terme "imp" fait référence à l'impossibilité de générer les règles floues à cause du nombre de règles très élevé (supérieur à 10^5).

Pour la phase de regroupement des attributs, nous avons utilisé une discrétisation à intervalles égaux et avons fixé le nombre d'intervalles à 3. D'autres tailles de discrétisation ainsi que d'autres méthodes de discrétisation pourront être étudiées dans de prochains travaux. Au niveau de la phase d'apprentissage, nous avons utilisé une partition floue homogène et une partition floue supervisée. Pour la partition homogène, nous avons testé plusieurs valeurs de la taille de partition k . Comme dans SIFCO, la valeur de k qui permet d'obtenir le meilleur taux de bonne classification dépend fortement des données de la base. Pour la partition floue supervisée, nous avons adopté la méthode MDLP de Fayyad et Irani (1993).

Base de données	Nombre d'instances	Nombre d'attributs	Nombre de classes
Iris	150	4	3
Lupus	87	3	2
Wine	178	13	3
Vehicle	846	18	4
Sonar	208	60	2

TAB. 1 – Description des bases utilisées.

Nous avons testé différentes valeurs des seuils $minSupp$ et β_{min} afin de déterminer la combinaison permettant d'obtenir le meilleur taux de classification. Le choix des valeurs de $minSupp$ et β_{min} semble fortement dépendre des données utilisées. Une perspective immédiate de ce travail serait d'étudier le lien entre ces valeurs et les caractéristiques des données étudiées.

Nous présentons dans Tab. 2 une comparaison de notre méthode SIFRA avec les deux méthodes SIF et SIFCO. Chacune des 3 méthodes possède des paramètres d'entrée à définir, à savoir la taille de la partition floue k , le seuil et la méthode de corrélation pour SIFCO, les seuils $minSupp$ et β_{min} pour SIFRA. Pour comparer la performance des 3 méthodes et pour simplifier la lecture des résultats, nous présentons dans Tab. 2, pour chaque méthode, le meilleur taux de bonne classification obtenu en faisant varier ses paramètres d'entrée.

D'après Tab. 2, il est clair que notre approche SIFRA fournit des taux de bonne classification très satisfaisants comparée à la méthode SIF, et des taux similaires ou meilleurs comparée à SIFCO. Comparée aux SIF, notre approche permet d'améliorer la performance de classification et de diminuer notablement le nombre de règles (en particulier avec les bases Wine, Vehicle et Sonar pour lesquelles la génération des règles avec SIF est impossible vu l'explosion de leur nombre). Comparée à SIFCO, notre méthode donne le meilleur taux de classification pour la base Iris avec un nombre de règles plus élevé mais qui reste faible (33). Pour Lupus, Wine et Sonar, les mêmes taux ont été obtenus par SIFRA et SIFCO. Concernant la base Vehicle, notre approche améliore considérablement le taux de bonne classification (67.73% contre 54.97% avec SIFCO) mais avec un nombre de règles plus important. Ce différentiel du nombre de règles s'explique par le fait que les groupes d'attributs liés détectés par SIFRA (basé sur les RA) contiennent plus d'attributs que les groupes détectés par SIFCO (basé sur une recherche

de corrélation linéaire) (équation 7). Avec ces données, les liaisons entre attributs déterminées par notre approche semblent donc être plus pertinentes que celles trouvées avec SIFCO.

Méthode	SIF	SIFCO		SIFRA	
		Régulière	Supervisée	Régulière	Supervisée
Iris	96 (46)	96.67 (24)	97.33 (14)	98 (33)	97.33 (14)
Lupus	77.01 (11)	79.31 (12)	79.31 (4)	79.31 (9)	79.31 (5)
Wine	imp	93.26 (60)	98.88 (38)	94.94 (58)	98.88 (38)
Vehicle	imp	50.82 (54)	54.97 (83)	60.05 (217)	67.73 (217)
Sonar	imp	71.63 (240)	69.23 (111)	71.63 (231)	66.35 (78)

TAB. 2 – Comparaison de notre système SIFRA avec les systèmes SIF et SIFCO.

6 Conclusion

Nous avons proposé une méthode originale de regroupement des attributs, intitulée SIFRA, qui se base sur le concept des règles d'association. L'avantage de SIFRA par rapport à SIFCO est qu'elle peut détecter des associations autres que linéaires entre les attributs. Dans ce travail, nous avons appliqué notre approche à des variables quantitatives. Dans un cadre plus général, elle pourra être appliquée à des variables quantitatives et/ou qualitatives. Les tests réalisés sur différentes bases donnent des résultats très satisfaisants aussi bien en termes de taux de bonnes classifications qu'en termes du nombre de règles générées. Ces tests sont toutefois à compléter avec d'autres bases, en particulier celles où les données ne sont pas linéairement corrélées.

Il serait intéressant d'envisager d'utiliser d'autres méthodes de discrétisation au niveau de la phase de regroupement des attributs. Nous envisageons également d'appliquer notre approche à des grilles autre que la grille simple, comme par exemple la grille rectangulaire (Ishibuchi et al., 1994).

Références

- Agrawal, R., T. Imieliński, et A. Swami (1993). Mining association rules between sets of items in large databases. In *ACM SIGMOD Conf.*, Washington DC, USA, pp. 207–216.
- Agrawal, R. et R. Srikant (1994). Fast algorithms for mining association rules. In *20th VLDB Conf.*, Santiago, Chile, pp. 487–499.
- Borgi, A. (1999). *Apprentissage supervisé par génération de règles : le système SUCRAGE*. Thèse de doctorat, Université Paris 6.
- Dehzangi, O., M. J. Zolghadri, S. Taheri, et S. M. Fakhrahmad (2007). Efficient fuzzy rule generation : A new approach using data mining principles and rule weighting. In *4th Int. Conf. on Fuzzy Systems and Knowledge Discovery*, pp. 134–139. IEEE.
- Fayyad, U. et K. Irani (1993). Multi-interval discretization of continuousvalued attributes for classification learning. In *13th Int. Joint Conf. on Artificial Intelligence*, San Francisco, CA, pp. 1022–1027. Morgan Kaufmann.

Regroupement d'attributs par application des règles d'association

- Ishibuchi, H., K. Nozaki, et H. Tanaka (1992). Distributed representation of fuzzy rules and its application to pattern classification. *Fuzzy Set and Systems* 52, 21–32.
- Ishibuchi, H., K. Nozaki, N. Yamamoto, et H. Tanaka (1994). Construction of fuzzy classification systems with rectangular fuzzy rules using genetic algorithms. *Fuzzy Set and Systems* 65, 237–253.
- Ishibuchi, H., K. Nozaki, N. Yamamoto, et H. Tanaka (1995). Selecting fuzzy if-then rules for classification problems using genetic algorithms. *Fuzzy Systems, IEEE Trans.* 3, 260–270.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *14th Int. Joint Conf. on Artificial Intelligence*, Canada, pp. 1137–1143. Morgan Kaufmann.
- Lee, H. M., C. M. Chen, J. M. Chen, et Y. L. Jou (2001). An efficient fuzzy classifier with feature selection based on fuzzy entropy. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Trans.* 31, 426–432.
- Nozaki, K., H. Ishibuchi, et H. Tanaka (1994). Selecting fuzzy rules with forgetting in fuzzy classification systems. In *Fuzzy Systems, IEEE World Congress on Computational Intelligence, the 3rd IEEE Conf.*, pp. 618–623. IEEE.
- Savasere, A., E. R. Omiecinski, et S. B. Navathe (1995). An efficient algorithm for mining association rules in large databases. In *21st VLDB Conf.*, Zurich, Switzerland, pp. 432–444.
- Soua, B., A. Borgi, et M. Tagina (2012). An ensemble method for fuzzy rule-based classification systems. *Knowledge and Information Systems* 36, 385–410.
- Srikant, R. et R. Agrawal (1995). Mining generalized association rules. In *21st VLDB Conf.*, Volume 95, pp. 407–419.
- Srikant, R. et R. Agrawal (1996). Mining quantitative association rules in large relational tables. In *ACM SIGMOD Int. Conf. on Management of Data*, Montreal, Canada, pp. 1–12.
- Valentini, G. et F. Masulli (2002). Ensembles of learning machines. In *13th Italian workshop on neural nets*, Berlin, pp. 3–19. Springer.

Summary

In fuzzy rule-based classification systems, a large number of descriptive attributes leads to an explosion of the generated rules' number and may affect the accuracy of learning algorithms. In order to address this problem, a solution is to treat separately subgroups of attributes. This allows decomposing the learning problem into sub-problems of lower complexity, and getting more intelligible rules as they are smaller. We propose a new method to regroup attributes; it is based on the concept of association rules. These rules highlight interesting relationships between value ranges of attributes. These local associations are then aggregated at the attributes' level according to the number of found associations and to their significance. Our approach, tested on different learning bases and compared to the classical approach (SIF), allows improving the accuracy and guarantees a reduction of the rules' number.