

Génération incrémentale de concepts formels et leur utilisation dans le Classifieur Flou par Localisation de Concepts

S. Elloumi, Ch. Ben Youssef, S. Ben Yahia, H. Ounelli

Département des Sciences de l'Informatique
Faculté des Sciences de Tunis
Campus Universitaire, 1060 Tunis, Tunisie.
{samir.elloumi;sadok.benyahia;Habib.ounalli}@fst.rnu.tn

RÉSUMÉ. Le Classifieur Flou par Localisation de Concepts (CFLC) fait partie des approches conceptuelles pour l'apprentissage automatique. Contrairement aux approches pénalisées par le temps de calcul nécessaire pour la génération du treillis de concepts, le CFLC commence, moyennant les opérateurs de connexion de Galois ainsi qu'un ajustement d'un "seuil de vérification", par la localisation du concept formel le plus proche de l'objet à classer et prend ce concept comme point de départ pour la prédiction de sa classe. Ainsi, et après calcul de localisation, le concept formel est utilisé uniquement une seule fois pour un objet donné. La principale amélioration que nous proposons d'apporter à cette approche est de réutiliser les concepts trouvés pour d'autres objets en constituant une base de concepts. Par conséquent, nous traitons dans ce papier les problèmes suivants : la gestion incrémentale de cette base des concepts et la recherche de la localisation du concept adéquat à un objet donné à partir de cette dernière.

ABSTRACT. The fuzzy classifier by concepts localization (FC2L) can be considered as a conceptual approach for supervised automatic classification. It's main feature consists in avoiding the Galois lattice construction cost. Indeed, it is based on the localization of the "closest" concept by means of "verification threshold" to predict a novel object class. In this paper, we extend the aforementioned classifier by introducing the "Concepts base", and we tackle the some induced problems such as : the incremental maintenance of such Base and the efficient manner of concept localization within.

MOTS-CLÉS : Classification conceptuelle, Treillis de Galois, Maintenance incrémentale.

KEYWORDS : Conceptual classification, Galois lattice, Incremental maintenance.

1. Introduction

L'apprentissage conceptuel se situe parmi des méthodes symboliques et se base sur les notions de concept formel et treillis de Galois faisant partie du domaine de l'analyse formelle de concepts.[6, 11]

L'utilisation de l'apprentissage conceptuel, pour la prédiction, a été présentée par quelques approches qui ont eu beaucoup de succès [2, 5]. Ces dernières ont montré l'avantage de prédire une classe pour une nouvelle observation, en cherchant le concept formel qui lui correspond le mieux. Cependant, ces approches ont soulevé le problème d'application réelle, vue la complexité élevée relative aux tâches de génération et d'identification de concepts formels.

Le Classifieur Flou par Localisation de Concepts (CFLC) fait partie des approches conceptuelles pour l'apprentissage automatique. Contrairement aux approches pénalisées par le temps de calcul nécessaire pour la génération du treillis de concepts, le CFLC commence, moyennant les opérateurs de connexion de Galois ainsi qu'un ajustement d'un "seuil de vérification", par la localisation du concept formel le plus proche de l'objet à classer et prend ce concept comme point de départ pour la prédiction de sa classe. Ainsi, et après calcul de localisation, le concept formel est utilisé uniquement une seule fois pour un objet donné.

La principale amélioration que nous proposons d'apporter à cette approche est de réutiliser les concepts trouvés pour d'autres objets en constituant une base de concepts. Par conséquent, nous traitons dans ce papier les problèmes suivants : la gestion incrémentale de cette base des concepts et la recherche de la localisation du concept adéquat à un objet donné à partir de cette dernière.

Ainsi, ce papier est structuré comme suit. La Section 2 présente les fondements mathématiques sous-jacents de l'approche proposée. Dans la troisième Section, nous rappelons la méthode CFLC en expliquant à quel point le seuil de vérification est important pour une meilleure qualité du classifieur. La Section 4 introduit l'algorithme de classification automatique en se basant sur la base de concepts. Une comparaison empirique fait l'objet de la Section 5. La section 6 conclut ce papier et dresse les perspectives futures du présent travail.

2. Fondement mathématique

Cette section présente les éléments fondamentaux sur lesquels se base notre approche. Il s'agit principalement d'un rappel sur les ensembles flous ainsi que la présentation des notions sur le domaine de l'analyse de concepts formels flous.

Employé	k ₁	k ₂	k ₃	k ₄	Classe
e ₁	1.0	0.5	0.9	0.0	C ₁
e ₂	1.0	0.4	0.8	0.4	C ₁
e ₃	0.9	0.3	1.0	0.9	C ₂
e ₄	0.0	1.0	1.0	1.0	C ₃
e ₅	0.1	0.9	1.0	1.0	C ₃

Tableau 1. Relation floue \tilde{R}

2.1. Sous-ensembles flous

Definition 1 Un ensemble flou \tilde{A} de l'univers de discours $U = \{u_1, \dots, u_n\}$ est caractérisé par une fonction d'appartenance $\mu_{\tilde{A}} : U \rightarrow [0, 1]$, où $\mu_{\tilde{A}}(u)$ désigne le degré d'appartenance de u à l'ensemble flou \tilde{A} . L'ensemble flou \tilde{A} est dénoté par : $\tilde{A} = \left\{ \begin{matrix} \mu_{\tilde{A}}(u_1) \\ \dots \\ \mu_{\tilde{A}}(u_n) \end{matrix} \right\}$

Exemple 1 Soit $U = \{a, b, c\}$. L'ensemble $\tilde{A} = \left\{ \begin{matrix} 0.5 \\ a \\ 0.1 \\ b \\ 0.9 \\ c \end{matrix} \right\}$ est un exemple d'ensemble flou. Les degrés d'appartenance de a, b et c dans \tilde{A} sont, respectivement, 0.5, 0.1 et 0.9. Plus particulièrement, avec le degré d'appartenance 0.9, l'élément c possède une forte appartenance à \tilde{A} ; tandis que l'élément b , avec le degré 0.1, y appartient faiblement.

2.2. Concept formel flou

Différentes extensions floues du concept formel ont été proposées dans la littérature selon différents points de vues. Plus particulièrement, nous distinguons la proposition de Wolff [12] qui consiste à ramener le problème à un contexte non flou contrairement à Pollandt [10] et Belohlavěk [1] qui ont utilisé des opérateurs flous dans les nouvelles définitions qu'ils ont proposées. Nous rappelons dans ce qui suit la définition que nous avons déjà développée dans [4].

Definition 2 Soit G un ensemble d'objets et M un ensemble d'attributs (propriétés). Une relation floue \tilde{R} entre les ensembles G et M est un ensemble flou défini sur $G \times M$. La valeur $\mu_{\tilde{R}}(g, m) \in [0, 1]$ est interprétée comme étant le degré de vérité de la proposition "l'objet $g \in G$ possède l'attribut $m \in M$ ".

Exemple 2 La Table 1 présente une relation floue \tilde{R} décrivant à quel degré chaque employé vérifie une qualification donnée. La relation \tilde{R} comporte, également, un attribut classe qui représente la classe attribuée à chaque employé.



Definition 3 Soit triplet $\langle G, M, \tilde{R} \rangle$ appelé contexte flou. Soient A, \tilde{B} deux ensembles tels que A est un ensemble ordinaire de G , \tilde{B} est un ensemble flou défini dans M et $\delta \in [0, 1]$. On définit deux opérateurs \tilde{f} et \tilde{h}_A comme suit :

$$\tilde{f}(A) = \{m \mid \alpha = \min\{\mu_{\tilde{B}}(g, m), g \in A\}, m \in M\}$$

$$\tilde{h}_A(\tilde{B}) = \{g \in G \mid \forall m, m \in M \Rightarrow [\mu_{\tilde{B}}(m) \rightarrow_{I_L} \mu_{\tilde{B}}(g, m)] \geq \delta\}$$

Avec \rightarrow_{I_L} désigne l'implication de Lukasiewicz i.e. pour $a, b \in [0, 1], a \rightarrow_{I_L} b = \min(1, 1 - a + b)$.

Definition 4 Un concept formel flou, ou concept approximatif (CAPS) au niveau δ du contexte $\langle G, M, \tilde{R} \rangle$ est une paire (A, \tilde{B}) tel que : $\tilde{f}(A) = \tilde{B}$ et $\tilde{h}_A(\tilde{B}) = A$. L'ensemble A est appelé l'extension ou domaine du CAPS et l'ensemble \tilde{B} est appelé son intention ou son codomaine.

3. Classifieur Flou par Localisation de Concepts

Le principe de cette méthode consiste à localiser un concept formel flou (CAPS) contenant un nouvel objet α_x , sans générer le treillis de Galois [3]. Le classifieur flou par localisation de concepts est composé de deux étapes : i) recherche du concept approximatif, ii) affectation de classes.

3.1. Recherche du Concept Approximatif

La localisation du CAPS contenant α_x consiste seulement à calculer $\tilde{f} \circ \tilde{h}_A(\alpha_x)$. En effet, l'opération $\tilde{h}_A(\alpha_x)$ est utilisée afin d'extraire un ensemble d'objets $A_1 \subseteq G$ qui vérifient toutes les propriétés de α_x selon leurs degrés respectifs au niveau δ . En d'autres termes, $\forall \alpha \in A_1, m \in M, (\mu_{\tilde{B}}(m) \rightarrow_{I_L} \mu_{\tilde{B}}(\alpha, m)) \geq \delta$.

Par conséquent, $\tilde{h}_A(\alpha_x)$ permet de générer une relation, dénotée par *RelApp* avec $RelApp \subseteq_{\delta} \tilde{R}$, que l'on peut qualifier de "relation Approximative" à cause de l'utilisation du seuil de vérification δ . Le CAPS généré à partir de *RelApp* est qualifié comme étant un *concept Approximatif*.

Exemple 3 Considérons la relation floue \tilde{R} donnée par la Table 1, pour $\alpha_x = \{k_1, k_2, k_3, k_4\}$ on a $A_1 = \tilde{h}_1(\alpha_x) = \{e_1, e_2, e_3, e_4, e_5\}$. La relation *RelApp* est présentée dans la Table 2.

Pour calculer $\tilde{B}_1 = \tilde{f}(A_1)$ qui représente un ensemble flou sur des propriétés, il s'agit seulement de repérer le degré minimal pour chaque colonne " e_i ", $i = 1, \dots, n$, dans



	k_1	k_2	k_3	k_4	classe
e_1	0.9	0.2	0.3	0.8	C_1
e_2	0.9	0.2	0.9	0.8	C_2
e_3	0.9	0.2	0.3	0.8	C_1
e_4	0.6	0.1	0.3	0.8	C_2
e_5	0.5	0.1	0.3	0.5	C_3

Tableau 2. La relation $RelApp$.
 $RelApp$ et l'affecter à la propriété " k_i ". Par exemple, à partir de $RelApp$ donnée par la Table 2, on trouve $\tilde{B}_1 = \tilde{f}(A_1) = \{k_1, k_2, k_3, k_4\}$. Le couple (A_1, \tilde{B}_1) représente ainsi le concept Approximatif contenant l'objet e_5 .

3.2. Affectation de classes

Le choix de la classe d'un nouvel objet e_x est basé sur la relation $RelApp$. En effet, l'affectation d'une classe pour e_x dépend de la correspondance entre les objets du concept supérieur et leurs classes respectives. Il s'agit de repérer, dans $RelApp$, la valeur minimale par propriété et d'observer la classe qui lui est associée. Cela permet de dresser une table de décision pour chaque propriété. Ces tables de décision sont à la base de la décision finale. A partir de ces tables de décision, différentes approches peuvent être proposées pour le choix d'une seule décision.

Exemple 4 En utilisant la relation $RelApp$ (cf. Table 2), chaque ligne de la table 3 représente une table de décision. Par exemple, la table de décision pour k_1 indique qu'il existe un seul objet qui possède le degré minimal (0.5) et pour lequel est associée la classe C_3 . Par contre, ceci n'est pas le cas de la table de décision pour k_3 , où on observe quatre objets qui possèdent le degré minimal (0.3) et qui concernent des classes différentes.

Il est à noter que l'approche CFLC est une méthode simple et facile à mettre en oeuvre. Elle utilise seulement les opérateurs \tilde{f} et \tilde{h}_i pour localiser le concept formel flou (CAPS) contenant un nouvel objet e_x et elle exploite ce CAPS pour l'affectation d'une classe. Toutefois, un problème majeur peut être rencontré et peut entraîner une mauvaise classification. Il s'agit de l'obtention d'un CAPS vide. En effet, l'application de $\tilde{h}_i(e_x)$ peut donner un ensemble vide. En d'autres termes, il n'y a aucun degré supérieur ou égal à un

	C_1	C_2	C_3
k_1	0	0	1
k_2	0	1	1
k_3	2	1	1
k_4	0	0	1

Tableau 3. Tables de décision à partir de $RelApp$

	k_1	k_2	k_3	k_4	classe
e_1	0.9	0.2	0.3	0.8	C_1
e_2	0.9	0.2	0.9	0.8	C_2
e_3	0.9	0.2	0.5	0.8	C_1

Tableau 4. La relation *RelApp* pour $\delta = 0.9$

degré associé à α_i pour une propriété k_i , $i = 1, \dots, n$. Par exemple, pour $\alpha_2 = \{k_1, k_2, k_3, k_4\}$ on a : $\tilde{f} \circ \tilde{h}_1(\alpha_2) = \emptyset$. Par conséquent, aucune classe ne peut être affectée à α_2 . Toutefois, une réduction de δ , i.e. si $\delta = 0.9$ alors, la relation *RelApp* présentée dans la Table 4 n'est à présent plus vide. Le CAPS à α_2 est égal à : $\{e_1, e_2, e_3\} \times \{k_1, k_2, k_3, k_4\}$, où nous remarquons que la valeur associée à k_2 , i.e. 0.2, est inférieure au degré 0.3 associée à la même propriété dans α_1 .

4. Base de Concepts dans le CFLC

Les calculs ainsi que le temps d'exécution lors de la classification d'un objet peuvent être optimisés par la génération incrémentale d'une Base de Concepts (BC) qui permettra de conserver tous les concepts créés afin de les réutiliser et d'éviter ainsi de parcourir à chaque fois tout l'échantillon d'apprentissage.

4.1. Principe

Lors de la classification d'un nouvel objet, la recherche du CAPS ne s'effectue plus directement dans l'échantillon d'apprentissage (EA). En effet, un concept adéquat est recherché dans la BC, s'il existe on affecte sa classe à l'objet. Cette recherche est guidée par une nouvelle variable appelé *Ecart*. Cette variable permet de choisir parmi les concepts qui vérifient l'objet ceux qui sont proches de l'objet à classer. Dans le cas contraire, c'est-à-dire si aucun concept n'a été retenu dans BC, on retourne à EA. Cette amélioration peut s'avérer très utile surtout si le nombre d'objets dans EA est élevé.

4.2. Algorithme

Au départ, un seuil, un pas et un écart sont fixés. Le seuil permet dans EA de sélectionner les objets et dans BC les concepts. Le pas sert à diminuer le seuil si aucun concept n'a été trouvé pour l'objet à classer. Enfin, l'écart permet de choisir le concept le plus proche de l'objet en question. Le pseudo-code de l'algorithme INC-CON-CLASS est donné par l'Algorithme 1.

Tous les concepts de l'ensemble BC sont structurés comme suit : "La classe du concept, "La liste des objets du concepts" et "La liste des propriété du concept. BC peut être orga-

Algorithm 1: INC-CON-CLASS

Données: Objet à classer, Seuil, Pas, Ecart (EA et BC)

Résultat: Décision de classification et BC modifiée

début

V-Seuil=Seuil

Un-concept = concept

tant que aucune décision de classification n'a été prise **faire**

Un-concept = RECHERCHE-BC(Seuil, Ecart, pas, liste-des-écarts)

si Un-concept non vide **alors**

Décision de classification

sinon

Un-concept= RECHERCHE-EA(Seuil, pas)

si Un-concept non vide **alors**

Décision de classification

si Un-concept. *ecart* liste-*des-écarts* **alors** MAJ BC**sinon** Seuil=Seuil - Pas**fin**

Algorithm 2. *Generation incrémentale de concepts formels et leur utilisation dans le CFLC*
 nisée sous la forme d'une simple liste chaînée; sous la forme d'un treillis flou ou mieux encore sous la forme d'un trie flou.

A noter que la routine RECHERCHE-BC effectue une recherche du concept adéquat, i.e., celui qui vérifie l'objet à classer et qui lui soit proche. Pour cela il s'agit de parcourir BC à la recherche des concepts qui vérifient l'objet selon le seuil puis de retenir celui qui a le plus petit écart. Si ce dernier est adéquat, c'est-à-dire son écart est inférieur ou égal à Ecart alors on affecte sa classe à l'objet à classer. Afin d'éviter la redondance, un trace des différents écarts des concepts vérifiant l'objet est gardée. Cependant, la routine RECHERCHE-EA applique la méthode CFLC. La routine de Mise à jour de BC est appelée à chaque calcul d'un nouveau concept (après recherche EA). On calcule d'abord l'écart entre ce nouveau concept et l'objet à classer. Si cet écart existe dans la liste des écarts générée lors du parcours de BC alors il s'agit probablement d'un concept de BC et il n'est donc pas nécessaire de l'ajouter à BC.

5. Evaluation

Cette première évaluation a pour but de tester l'impact de l'utilisation de BC sur le Taux de Mauvaise Classification (TMC). Notre but n'est pas encore d'optimiser le

temps d'exécution. Pour cela nous avons jusqu'ici utilisé trois bases floues connues, <http://kel.otago.ac.nz/hyspeech/corpusinfo.html>, dont les caractéristiques sont données par la Table 5 :

	Nombre de propriétés	Nombre de classes	Taille EA	Taille ET
Iris	4	3	140	10
Formant	5	4	463	230
Voyelle	12	4	1200	600

Tableau 5. Liste des caractéristiques des différentes bases de test

5.1. Résultats obtenus avec IRIS

– **CFLC** : Lors de l'application de CFLC, nous avons remarqué que seule la variation du seuil de départ influençait le TMC de Iris (cf., Figure 1 (Iris(a))).

– **CFLC avec BC** : Nous avons alors choisi de faire varier l'écart pour évaluer l'impact de la classification à partir de BC et nous avons choisi 1 et 0.95 comme seuils de départ. Nous avons aussi effectué pour chaque écart un test sur BC vide au départ ensuite sur BC pleine. Nous avons alors remarqué que pour un écart nul, le TMC reste inchangé. De plus, plus l'écart augmente plus le TMC augmente. Quant à la comparaison entre BC vide et BC non Vide nous avons remarqué que le TMC est presque le même (cf., Figure 1 (Iris(b) et Iris(c))).

5.2. Résultats obtenus avec Formant

– **CFLC** : Lors de l'application de CFLC, nous avons remarqué que seule la variation du pas influençait le TMC de Formant (cf., Figure 1 (Formant(a))).

– **CFLC avec BC** : Comme avec Iris, nous remarquons que TMC reste inchangé pour un Ecart nul. Cependant, l'augmentation de l'écart a nettement amélioré le TMC. Cette amélioration est expliquée par l'utilisation de BC. En effet, plus l'écart augmente plus l'utilisation de BC est favorisée. De plus, dans BC on ne choisit pas seulement les concepts qui vérifient les objets à classer mais aussi ceux qui sont proches ce qui n'est pas le cas lors de l'application de CFLC dans EA. C'est aussi pour cette raison que le TMC reste inchangé pour un écart nul puisque la classification sera dans la majorité des cas effectuée à partir de EA. Nous remarquons aussi que une base pleine n'a jamais augmenté le TMC mais au contraire elle l'a nettement diminué (cf., Figure 1 (Formant(b) et Formant(c))).

5.3. Résultats obtenus avec Voyelle

- **CFLC** : Lors de l'application de CFLC, nous avons remarqué que seule la variation du pas influençait le TMC de Voyelle (cf., Figure 1 (Voyelle(a))).
- **CFLC avec BC** : Nous avons aussi effectué pour chaque écart un test sur BC vide au départ ensuite sur BC pleine. Comme avec les autres bases un écart nul ne change pas le TMC puisque la classification est effectuée dans la majorité des cas dans EA. (cf., Figure 1 (Voyelle(b) et Voyelle(c))).

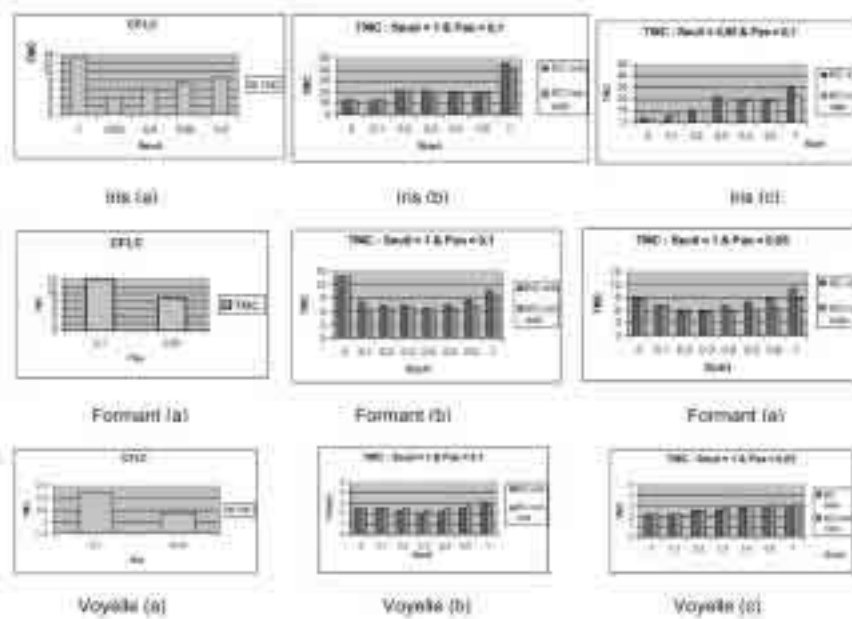


Figure 1. Résultats des expérimentations

6. Conclusion

Nous avons présenté une approche pour la classification des objets basée sur l'utilisation des concepts formels. Il s'agit de localiser le concept formel le plus proche d'un objet afin de décider de sa classe. L'étape la plus importante de cette approche c'est la localisation de concepts. Elle peut se faire ou bien sur une liste de concepts appelée base

de concepts (BC), ou bien, c'est-à-dire, dans le cas d'une recherche non fructueuse, sur la totalité de l'échantillon d'apprentissage (EA). Ainsi, une mise à jour incrémentale de BC est effectuée à la suite d'une recherche dans EA. Le principal avantage de cette approche réside dans la construction progressive de la liste des concepts et dans la génération unique d'un même concepts. L'expérimentation confirme le fait que l'utilisation de cette nouvelle approche permet d'améliorer le taux de mauvaise classification. Cependant, cette amélioration est pénalisée par les temps de recherche vue la structure séquentielle de BC. L'étape suivante serait donc d'améliorer, d'une part, la structure de la base de concepts afin d'accélérer la recherche de concepts et, d'autre part, d'effectuer correctement la mise à jour des concepts dans cette nouvelle structure.

7. Bibliographie

- [1] R. Belohlavěk. Fuzzy Galois connections. *Math. Logic Journal*, 45(4) :497–504, 1999.
- [2] C. Carpineto and G. Romano. A Lattice Conceptual Clustering System and its Application to Browsing Retrieval. *Machine Learning Journal*, 24 :95–122, 1996.
- [3] S. Elloumi and A. Jaoua. Automatic classification using fuzzy concepts. In *Proc. JCIS'2000, Atlantic City, USA*, volume 1, pages 276–279, 2000.
- [4] S. Elloumi and A. Jaoua. A multi-level conceptual learning approach using fuzzy galois connection. In *Proc. of the International Symposium on Innovation in Information and Communication Technology, Amman, Jordan*, 2001.
- [5] J.G. Ganasia. *Charade : Apprentissage de bases de connaissances, in : Induction Symbolique et Numérique à partir de Données*, pages 309–326. Cépadues-édition, 1991.
- [6] B. Ganter and R. Wille. *Formal Concept Analysis*. Springer-Verlag, Heidelberg, 1999.
- [7] R. Godin, G. W. Mineau, R. Missaoui, and H. Mili. Méthodes de Classification Conceptuelle Basées sur le Treillis de Galois et Applications. *Revue d'intelligence Artificielle*, 9(2) :105–137, 1995.
- [8] M. Liquière and E. M. Nguifo. Legal : Learning with galois lattice. In *Proc. of 5èmes Journées Françaises sur l'Apprentissage, Lannion*, 1990.
- [9] V. Novak. *Fuzzy sets and their applications*. Adam Higler, 1989.
- [10] S. Polland. *Fuzzy-Begriffe, Formale Begriffsanalyse unscharfer Daten*. Springer-Verlag Berlin Heidelberg New York, 1997.
- [11] R. Wille. Restructuring lattice theory : an approach based on hierarchies of concepts. In *Proc. of Nato Advanced Study Institute, Ed. by I. Rival, Reidel Publ. Comp*, volume 1, 1982.
- [12] K.E. Wolff. Conceptual interpretation of fuzzy theory. In *Proc. 6th European Congress on Intelligent techniques and Soft computing*, volume 1, pages 555–562, 1998.
- [13] L.A. Zadeh. *Fuzzy Sets and their Application to Pattern Classification And Clustering Analysis*. Academic Press, New York, 1977.