Classification Methods Based on Formal Concept Analysis

Olga Prokasheva, Alina Onishchenko, and Sergey Gurov

Faculty of Computational Mathematics and Cybernetics, Moscow State University

Abstract. *Formal Concept Analysis (FCA) provides mathematical models for many domains of computer science, such as classification, categorization, text mining, knowledge management, software development, bioinformatics, etc. These models are based on the mathematical properties of concept lattices. The complexity of generating a concept lattice puts a constraint to the applicability of software systems. In this paper we report on some attempts to evaluate simple FCA-based classification algorithms. We present an experimental study of several benchmark datasets using FCA-based approaches. We discuss difficulties we encountered and make some suggestions concerning concept-based classification algorithms.*

Keywords: Classification, pattern recognition, data mining, formal concept analysis, biclustering

1 Introduction

Supervised classification consists in building a classifier from a set of examples labeled by their classes or precedents (learning step) and then predicting the class of new examples by using the generated classifiers (classification step). Document classification is a sub-field of information retrieval. Documents may be classified according to their subjects or according to other attributes (such as document type, author, year of publication, etc.). Mostly, document classification algorithms are based on supervised classification. Algorithms of this kind can be used for text mining, automatical spam-filtering, language identification, genre classification, text mining. Some modern data mining methods can be naturally described in terms of lattices of closed sets, i.e., concept lattices [1], also called Galois lattices. An important feature of FCA-based classification methods do not make any assumptions regarding statistical models of a dataset. Biclustering [9, 10] is an approach related to FCA: it proposes models and methods alternative to classical clustering approaches, being based on object similarity expressed by common sets of attributes. There are several FCA-based models for data analysis and knowledge processing, including classification based on learning from positive and negative examples [1, 2].

In our previous work [15] the efficiency of a simple FCA-based binary classification algorithm was investigated. We tested this method on different problems with numerical data and found some difficulties in its application. The main purpose of this paper is to investigate critical areas of the FCA method for better understanding of its features. Several advices for developers are also provided. We test hypothesis-based classification algorithm and our modified FCA-based method on 8 benchmarks. We describe our experiments and compare the performance of FCA-based algorithms with that of SVM-classification [16].

2 Definitions

Formal Concept Analysis. In what follows we keep to standard FCA definitions from [1]. Let G and M be sets, called set of objects and set of attributes, respectively. Let $I \subseteq G \times M$ be a binary relation. The triple $K = (G, M, I)$ is called a *formal context*. For arbitrary $A \subseteq G$ and $B \subseteq M$ the mapping $(.)'$ is defined as follows:

$$A' = \{m \in M \mid \forall g \in A(gIm)\};\ B' = \{g \in G \mid \forall m \in B(gIm)\}. \tag{1}$$

This pair of mappings defines a Galois connection between the sets 2^G and 2^M partially ordered by the set-theoretic inclusion. Double application of the operation $(\cdot)'$ is a closure operator on the union of the sets 2^G and 2^M. Let a context K be given. A pair of subsets (A, B), such that $A \subseteq G$, $B \subseteq M$, $A' = B$, and $B' = A$ is called a *formal concept* of K with *formal extent* A and *formal intent* B. The extent and the intent of a formal concept are closed sets.

FCA in learning and classification. Here we keep to definitions from [2] and [3]. Let $K = (G, M, I)$ be a context and $w \notin M$ a target attribute. In FCA terms, the input data for classification may be described by three contexts w.r.t. w: the positive context $K_+ = (G_+, M, I_+)$, the negative context $K_- = (G_-, M, I_-)$, and the undefined context $K_\tau = (G_\tau, M, I_\tau)$ [2]. G_-, G_+ and G_τ are sets of positive, negative and undefined objects respectively, $I_\epsilon \subseteq G_\epsilon \times M$, where $\epsilon \in \{-, +, \tau\}$ are binary relations that define structural attributes. Operators $(\cdot)'$ in these contexts are denoted by A^+, A^-, A^τ, respectively. For short we write $g', g'', g^+, g^-, g^\tau$ instead of $\{g\}', \{g\}'', \{g\}^+, \{g\}^-, \{g\}^\tau$, respectively. A formal concept of a positive context is called a *positive concept*. Negative and undefined concepts, as well as extents and intents of the contexts K_- and K_τ, are defined similarly. A positive formal intent B_+ of $(A_+, B_+) \in K_+$ is called *a positive or (+) — prehypothesis* if is not the formal intent of any negative concept, and it is called *a positive or (minimal) (+) — hypothesis* if it is not a subset of the intent g^- for some elementary concept (g, g^-) for a negative example g; otherwise it is called a false (+)-generalization.

Negative (or (-) —) prehypotheses, hypotheses, and false generalizations are defined similarly. The definitions imply that a hypothesis is also a prehypothesis. Hypotheses are used to classify undefined examples from the set G^τ. If unclassified object g^τ contains a positive but no negative hypothesis, it is classified positively, similar for negative. No classification happens if the formal intent g^τ does not contain any subsets of either positive or negative hypotheses (insufficient data) or contains both a positive and a negative hypothesis (inconsistent data).

Biclustering. The particular case of biclustering [10–12] we have considered is a development of the FCA-based classification method. Using FCA methods, we can construct a hierarchical structure of biclusters that reflects the taxonomy of data. *Density* of bicluster (A, B) of the formal context $K = (G, M, I)$ is defined as $\rho(A, B) = |I \cap \{A \times B\}|/(|A| \cdot |B|)$. Specify some value $\rho_{min} \in [0, 1]$. The bicluster (A, B) is called dense if $\rho(A, B) \geq \rho_{min}$. *Stability index* σ of a concept (A, B) is given by $\sigma(A, B) = |\mathcal{C}(A, B)|/2^{|A|}$, where $\mathcal{C}(A, B)$ is the union of the sets $C \subseteq A$ such that $C = B'$ [13, 21]. Biclusters, as well as dense and stable formal concepts (i.e., concepts having stability above a fixed threshold), are used to generate hypotheses for clustering problems [13].

3 Basic Classification Algorithms

Several FCA-based classification methods are known [19, 15]: GRAND [31, 17], LEGAL [26], GALOIS [25], RULELEARNER [24], CIBLe [30], CLNN&CLNB [27], NAVIGALA[28], CITREC [29, 17] and classification method based on hypotheses [8, 7, 2, 3]. There are several categories of FCA-based classification methods:

1. **Hypothesis-based classification** using the general principle described in Section 2.
2. **Concept lattice based classification.** A concept lattice can be seen as a search-space in which one can easily pass from a level to another one. The navigation can e.g. start from the top concept with the least intent. Then one can progress concept by concept by taking new attributes and reducing the set of objects. Many systems use lattice-based classification, such as GRAND [31, 17], RULEARNER [24], GALOIS [25], NAVIGALA [28] and CITREC [29, 17]. The common constraint of these systems is the exponential algorithmic complexity of generating a lattice. For this reason, some systems search in a subset of the set of all concepts.
3. **Classification based on Galois sub-hierarchies.** Systems like CLNN& CLNB [27], LEGAL [26] and CIBLe [30] build Galois sub-hierarchy (ordered set of object and attribute concepts), which drastically reduces algorithmic complexity.
4. **Cover-based classification.** A concept cover is a part of the lattice containing only pertinent concepts. The construction of a cover concept is based on heuristic algorithms which reduce the complexity of learning. The concepts are extracted one by one. Each concept is given by a local optimization of a measure function that defines pertinent concepts. IPR (Induction of Product Rules) [32] was the first method generating a concept cover. Each pertinent concept induced by IPR is given by a local optimization of entropy function. The sets of pertinent generated concepts are sorted from the more pertinent to the less pertinent and each pertinent concept with the associated class gives a classification rule.

4 Classification Experiments with Benchmarks

4.1 A Hypothesis-Based Algorithm

The method for constructing concept-based hypotheses described above inspired the following binary classification algorithm [15]. The main steps of the algorithm are as follows:

1. **Data binarization.** The situation where the attributes are non-binary, but a classification method is designed for binary data brings up the problem of attribute binarization, or scaling. This problem is very difficult and a lot of papers are devoted to it. Scaling problem arises also when we use FCA for object classification. For specific tasks scaling is usually carried out empirically by repeatedly solving the problem of classification on precedents. It is clear, however, that in a couple of "scaling–recognition method" the determining factor is exactly scaling. Indeed, in the event of its successful application a 'good' transformation of the feature space will be obtained and almost any recognition algorithm will show good results in that space. So that the problem of scaling is a nonspecific for FCA-recognition methods and the current level of development of these methods unable to point the best technique of scaling focused on their use. That is why our work is not focused on this problem and we use a simple scaling, which, we believe, allowed more clearly to identify the features of FCA-classification methods. Hence, we just normalized all attributes to [0,1] interval and than applied interval-based nominal scaling. The number of intervals is fixed and equals 10. The size of intervals is also fixed and equals 0.1.
2. **Hypothesis generation and classification.** Algorithm searches common attributes for all objects from the first class (second class), which are not observed for any objects from the second class (first class). Obtained sets of attributes (hypothesis) are used to classify undefined objects.

The algorithm has been tested on numerical benchmarks. The data for the first four problems is taken from the UCI Machine Learning Repository[1]. Problem 5 (Two Norm) involving the separation of two normal 20-dimensional distributions is taken from the University of Toronto site[2]; the CART classification algorithm [22] produced for this problem an error rate of 22.1% with a training sample of 300 precedents, which is almost a factor of 10 higher than the theoretical minimum for the ideal classifier — the Fisher discriminant function. Problems 6 (Lung Cancer), 7 (Cirrhosis), and 8 (Cloud Seeding) are taken from the StatLib site[3]. The considered problems come from different specific research areas. For example, in the Liver Disorders problem, the objects are datasets obtained from the tests of six patients. The training sample consists of 345 precedents divided into positive and negative classes with respect to the

[1] http://archive.ics.uci.edu/ml
[2] http://www.cs.toronto.edu/∼delve/data/twonorm/desc.html
[3] http://lib.stat.cmu.edu/datasets, pages /veteran, /pbc, and /cloud, respectively

target attribute “presence of liver disorder”. The experiment results obtained for various problems by using leave-one-out cross-validation are presented below (Table 1). In the table heading, n is the number of attributes, l is the number of objects (the size of the training sample), err is the classification error rate and l_c is the number of classified objects ($l - l_c$ = the number of failed classifications).

Problem	n	l	l_c	err
1. Liver Disorders	6	345	20	15.00%
2. Glass identification	9	146	25	20.00%
3. Wine	13	130	47	08.50%
4. Wine quality	11	310	51	09.80%
5. Two norm	20	354	109	07.30%
6. Lung cancer	8	137	9	11.10%
7. Cirrhosis	19	276	29	34.48%
8. Cloud-seeding	5	108	6	50.00%

Table 1. Experimental results

The algorithm was updated with the following modifications in definitions of hypothesis and classification:

1. Hypothesis modification: attributes observed for ”almost” all objects of the particular class were added to the hypothesis. It was ensured that the ratio of objects which did not comply with the hypothesis in the same class did not exceed the value P (a new algorithm parameter) and obviously there was no guarantee that the hypothesis is not contained in descriptions of objects of the opposite class.
2. Introduction of an inter-object metric and modification of the classification procedure: the ”distance” between objects increases as they reveal difference in a larger number of coordinates. We compute the distance of the object being classified by positive and negative hypotheses and normalize it by the number of attributes (or 1s in binary represenation) in each hypothesis. The object is classified to the nearest class in contexts of the metric defined above.
3. Attribute weighting: an attribute is assigned a weight which increases with the number of 1s in the corresponding column.

The modified algorithm was applied to the considered problems. The experimental results for $P = 0.2$ are given in Table 2.

4.2 Classification Using Biclustering

Biclustering can be used for classification upon data scaling. For this purpose we select informative objects which are included in biclusters with density greater than threshold ρ_{min}. Hypotheses are generated using these objects. This approach avoids noise effects during learning step [15]. The difference between

Problem	n	l	l_c	err
1. Liver Disorders	6	345	79	39.2%
2. Glass identification	9	146	64	25.00%
3. Wine	13	130	87	14.9%
4. Wine quality	11	310	142	17.60%
5. Two norm	20	354	224	15.10%
6. Lung cancer	8	137	36	36.10%
7. Cirrhosis	19	276	136	33.30%
8. Cloud-seeding	5	108	37	43.20%

Table 2. Experimental results for the modified algorithm

proposed algorithm and simple FCA algorithm resides only in the second step: hypotheses are now generated using only informative objects selected by biclustering. The method has two adjustable parameters: the bicluster density ρ_{min} and the ratio P of objects which do not satisfy classical hypotheses. The parameter ρ_{min} affects the generation of hypotheses. If its value is too small, hypothesis generation is tainted by noisy attributes and outliers. If its value is too large, the hypothesis will have to meet excessively stringent requirements. It may be efficient to use a range of values for ρ_{min} and thus focus on the main objects, skipping the marginal ones. This method has been tested, but it failed to produce a significant improvement of the classification performance, which will be later explained by the specific features of the particular problem.

The parameter P affects the ratio of objects which do not satisfy the hypotheses of the same class. When the parameter P is close to zero, hypotheses are generated in accordance with the classical definitions: they include only the attributes that are observed for all the objects of the given class. The difficulty is that hypotheses may become "non-representative" for the given class. If the parameter P is taken too large, the hypotheses will require that the control object has a large number of attributes, which again may impose an excessively stringent requirement on hypotheses. In a certain sense, this is the well-known overfitting effect often observed in pattern recognition.

The experimental results assessed by leave-one-out cross-validation are presented in Table 3 and Table 4. In the table heading, n is the number of attributes, l is the number of objects (the size of the training sample). The columns present the solution results obtained with the algorithm parameters (the threshold ρ and the proportion P) optimized by two criteria: the classification error rate err (Table 3) and the number of classified objects l_c ($l - l_c$ = the number of failed classifications) (Table 4). The local optimization of the algorithm parameters was carried out by the GaussSeidel method, their optimal values ρ_{min} and P^* are shown together with err.

According to the experimental results, the lower is the error rate, the smaller is the number of classified objects. We can construct an algorithm with zero error rate, but the ratio of classified objects will be also small. We apply such an algorithm for all considered objects. The results are shown in Table 5, where

Problem	n	l	l_c	err	ρ_{min}	P^*
1. Liver Disorders	6	345	22	13.6%	0.30	0.01
2. Glass identification	9	146	28	10.00%	0.15	0.05
3. Wine	13	130	76	02.00%	0.25	0.05
4. Wine quality	11	310	83	08.40%	0.25	0.05
5. Two norm	20	354	206	12.10%	0.15	0.15
6. Lung cancer	8	137	18	05.50%	0.01	0.01
7. Cirrhosis	19	276	33	21.00%	0.05	0.05
8. Cloud-seeding	5	108	7	28.00%	0.15	0.05

Table 3. Experimental results. Classification error rate is optimized.

Problem	n	l	l_c	err	ρ_{min}	P^*
1. Liver Disorders	6	345	79	29.1%	0.30	0.20
2. Glass identification	9	146	59	16.90%	0.30	0.20
3. Wine	13	130	85	08.20%	0.30	0.20
4. Wine quality	11	310	141	13.50%	0.30	0.20
5. Two norm	20	354	233	15.20%	0.30	0.20
6. Lung cancer	8	137	98	25.50%	0.05	0.05
7. Cirrhosis	19	276	83	37.79%	0.30	0.20
8. Cloud-seeding	5	108	20	30.00%	0.15	0.15

Table 4. Experimental results. The number of classified objects is optimized.

l_c is the number of classified objects (the ratio is in brackets). The efficiency of FCA-based algorithm was compared with that of classical SVM-algorithm [16]. Each dataset was divided into training sample (80% of objects) and test sample (20% of objects). Table 5 $SVMerr$ shows the error rate of the SVM-algorithm, $SVMerr\ on\ l_c$ is the error rate on objects which were classified by the rigorous FCA-based method. Zero error rate was attained with classical hypotheses

Problem	l	l_c	ρ_{min}	P^*	$SVMerr$	$SVMerr\ on\ l_c$
1. Liver Disorders	345	18 (5.2%)	0.15	0	34.78%	22.2%
2. Glass identification	146	22 (15%)	0.15	0	31.03%	4.55%
3. Wine	130	45 (35%)	0.25	0	7.69%	2.22%
4. Wine quality	130	49 (5.8%)	0.1	0	35.48%	6.12%
5. Two norm	354	103 (29%)	0.03	0	3.85%	0%
6. Lung cancer	137	9 (6.50%)	0.01	0	40.74%	0%
7. Cirrhosis	276	24 (9.00%)	0.05	0	9.8%	12.5%
8. Cloud-seeding	108	5 (4.6%)	0.15	0	40.91%	25.00%

Table 5. Experimental results with zero error rate.

(P=0) from objects with low density ($\rho_{min} \leq 0.25$). This rigorous algorithm can be applied to problems with high error costs. It is more likely to refuse classification than make wrong decisions.

5 Conclusions

FCA provides a convenient tool for formalizing symbolic machine learning and classification models. We studied hypothesis-based classification in different areas without special modifications for each dataset, using a simple binarization (scaling) of numerical data. Our results suggest the following conclusions:

1. Application of biclustering with parameter optimization made a very slight improvement in the quality of classification compared to the updated FCA algorithm (only by 3% in problem 1).
2. In all cases there was an unacceptably high rate of classification failures.
3. In all cases there was an unacceptably high error rate.
4. Attempts to fine-tune the algorithm parameters with the objective of reducing the failure rate were generally accompanied by increasing in the number of errors, although in some cases (problem 8) the error rate increased only slightly; the number of classifiable objects in these cases increased substantially (problem 6).
5. The classical FCA-based algorithm can produce accurate classification, but it refuses to classify the majority of test sample.

The analysis of the hypotheses generated with various parameter values and different optimization criteria has shown that hypotheses of different classes are often included in one another. We can naturally assume that if the classes show less tendency to diffuse into one another, biclustering and the classical FCA method would produce more impressive results. The relative location of classes is improved in pattern recognition theory by methods that involve transformation of the attribute space. In these cases, data compactification methods may be effectively applied [14]. We can reasonably assume that another scaling algorithms with floating-size intervals and interval-length optimization may improve classification results compared to those we have obtained with the simplest scaling. Our analysis of FCA-based classification provides the following conclusions:

1. For the chosen universal scaling procedure the classification results are far from being optimal. Individual scaling for each problem may improve classification quality.
2. FCA-based classification methods without modification and/or thorough preprocessing of data are usable only for preliminary classification.
3. A well-known idea for the modification of the direct FCA approach is to develop hypothesis generation methods. It is useful to allow for the specific features of the particular subject area and to fine-tune hypotheses and the algorithms by using e.g. parameters ρ_{min} , P*, σ_{min}.
4. It is also possible to develop and apply sharper classification rules, e.g. by weighting objects, attributes, hypotheses, etc.
5. A promising approach is to use FCA-based methods to transform the attribute space, in particular using data compactness estimates.
6. concept-based methods are appropriate for classification problems with high error costs, e.g. in medical, security and military applications.

An important step in many data classification problems is the selection of a suitable similarity measure. We decided to investigate how different similarity metric described in [20, 23] affects the quality of hypothesis-based classification method. The majority of pattern recognition methods use the metric information about objects: methods based on distances, potential functions, dividing surface, the algebraic approach, etc. In these methods the amount of information about classes either is fairly used at all. The strength of FCA-based classification problems is in the identification and use of these particular data, but the metric information about the feature space is lost. Thus, in pattern recognition the classic discriminant methods and method based on FCA are at opposite poles w.r.t. metrics. In the FCA the metric information appears in a weak form as the result of scaling, which accounts for “distances” between attributes. It seems that the success in the development of FCA-based recognition methods will be related to the introduction of information about metric properties of feature spaces. Our future work will be focused on developing and applying sharper classification rules with modifications of similarity measure.

References

1. B. Ganter, R. Wille: Formal Concept Analysis: Mathematical Foundations. Springer, Berlin/Heidelberg (1999).
2. S.O. Kuznetsov: Mathematical aspects of concept analysis. In: Journal of Mathematical Science, Vol. 80, Issue 2, pp. 1654–1698, (1996).
3. S.O. Kuznetsov: Complexity of Learning in Concept Lattices from Positive and Negative Examples. Discrete Applied Mathematics, 2004, No. 142(1–3), pp. 111-125.
4. G. Birkhoff: Lattice Theory. AMS, Providence, 3rd edition (1967).
5. O. Ore: Theory of Graphs. American Mathematical Society, Providence (1962).
6. S.I. Gurov: Ordered Sets and Universal Algebra [in Russian], MGU, Moscow (2004).
7. V.K. Finn: The Synthesis of Cognitive Procedures and the Problem of Induction. Autom. Doc. Math. Linguist., 43, 149-195 (2009).
8. V.K. Finn: On machine-oriented formalization of plausible reasoning in the style of F. Bacon and D.S. Mill [in Russian], Semiotika i Informatika, 20, 35–101 (1983).
9. S.C. Madeira and A.L. Oliveira: Biclustering Algorithms for Biological Data Analysis: A Survey. IEEE/ACM Trans. Comput. Biol. Bioinformatics 1, 24–45 (2004).
10. B.G. Mirkin: Mathematical Classification and Clustering. Kluwer (1996).
11. D.I. Ignatov, S.O. Kuznetsov: Frequent Itemset Mining for Clustering Near Duplicate Web Documents. In: Proc. 17th Int. Conf. on Conceptual Structures (ICCS 2009), LNAI (Springer), Vol. 5662, 185–200, 2009.
12. D.I. Ignatov, S.O. Kuznetsov, R.A. Magizov and L.E. Zhukov: From Triconcepts to Triclusters. In: Proc. 13th Int. Conf. on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing (RSFDGrC 2011), LNCS (Springer), Vol. 6743, 257–264, 2011.
13. S.O. Kuznetsov: Stability as an Estimate of the Degree of Substantiation of Hypotheses on the Basis of Operational Similarity. In: Nauchno-Tekhnicheskaya Informatsiya, Seriya 2, Vol. 24, No. 12, pp. 21–29, 1990.
14. S.I. Gurov, N.S. Dolotova, I.N. Fatkhutdinov: Noncompact recognition problems. Circuit design according to E. Gilbert. In: Spectral and Evolution Problems: Proc.

17 th Crimean Autumn Mathematical School-Symposium, Simferopol, Crimean Scientific Center of Ukrainian Academy of Sciences, 17, pp. 37–44 (2007).
15. A.A. Onishchenko, S.I. Gurov: Classification based on Formal Concept Analysis and Biclustering: possibilities of the approach. In: Computational Mathematics and Modeling, Vol. 23, No. 3, July, pp. 329–336 (2012).
16. C. Cortes, V. Vapnik: Support–vector networks. In: Machine Learning, September 1995, Volume 20, Issue 3, pp. 273–297.
17. N. Meddouri, M. Meddouri: Classification Methods based on Formal Concept Analysis, — CLA 2008 (Posters), pp. 9–16, Palacky University, Olomouc, (2008).
18. S. I. Gurov: Boolean Algebras, Ordered Sets, Lattices: Definitions, Properties, Examples [in Russian], KRASAND, Moscow (2012).
19. S.O. Kuznetsov: Machine learning on the basis of formal concept analysis Automation and Remote Control 62 (10),pp. 1543–1564.
20. F. Alqadah, R. Bhatnagar: Similarity measures in formal concept analysis, Annals of Mathematics and Artificial Intelligence, 61(3), 245–256 (2011).
21. S.O. Kuznetsov: On Stability of a Formal Concept. Annals of Mathematics and Artificial Intelligence, Vol. 49, pp.101–115, 2007.
22. L. Breiman, J.H. Friedman, R.A. Olsen, and C.J. Stone: Classification and Regression Trees, Wadsworth Int. Group, Belmont, CA (1984).
23. F. Dau, J. Ducrou, P.W. Eklund: Concept Similarity and Related Categories in SearchSleuth. ICCS 2008: pp. 255–268
24. M. Sahami: Learning classification Rules Using Lattices. N. Lavrac and S. Wrobel eds., pp. 343–346, Proc ECML, Heraclion, Crete, Greece (Avril 1995).
25. C. Caprineto, G. Romano: GALOIS An order-theoretic approach to conceptual clustering. In proceedings of ICML93, pp. 33–40, Amherst, USA (July 1993).
26. P. Njiwoua, E.M. Nguifo: Forwarding the choice of bias LEGAL-F Using Feature Selection to Reduce the complexity of LEGAL. In Proceedings of BENELEARN-97, ILK and INFOLAB, Tilburg University, the Netherlands, pp. 89–98, 1997.
27. Zhipeng Xie, Wynne Hsu, Zongtian Liu, Mong Li Lee: Concept Lattice based Composite Classifiers for high Predictability. Artificial Intelligence, vol. 139, pp. 253–267, Wollongong, Australia (2002).
28. S. Guillas, K. Bertet, J-M. Ogier: Extension of Bordats algorithm for attributes. Concept Lattices and Their Applications: CLA07, Montpellier, France (2007).
29. B. Douar, C. Latiri, Y. Slimani: Approche hybride de classification supervisee base de treillis de Galois: application la reconnaissance de visages. In: Extraction et Gestion des Connaissances (EGC08), 309–320, Nice, France (2008).
30. P. Njiwoua, Mephu Nguifo E.: Ameliorer lapprentissage partir dinstances grace linduction de concepts: le systeme CIBLe, Revue dIntelligence Artificielle (RIA), vol. 13, 2, 1999, pp. 413–440, Hermes Science.
31. E.M. Nguifo, P. Njiwoua: Treillis de concepts et classification supervise. Technique et Science Informatiques: TSI, Volume 24, Issue 4, pp. 449–488 (2005).
32. M. Maddouri: Towards a machine learning approach based on incremental concept formation. Intelligent Data Analysis, Volume 8, Issue 3, pp. 267–280 (2004).