New taxonomy of classification methods based on Formal Concepts Analysis

Marwa Trabelsi[1], Nida Meddouri[1] and Mondher Maddouri[2]

[1] Laboratory of computing, Programming, Algorithmic and Heuristic - LIPAH,
Faculty of Science of Tunis - FST, El Manar University, Tunis, Tunisie
trabelsimarou@live.com nida.meddouri@gmail.com
[2] Management college, Jeddah university, Kingdom of Saoudi Arabia
maddourimondher@yahoo.fr

Abstract Data mining is an essential step in knowledge extraction from data. Various approaches have been proposed in supervised classification of data, among them approaches based on Formal Concept Analysis. In this paper we present a new taxonomy of classification methods based on Formal Concept Analysis.

Keywords: Data Mining, Supervised Classification, Formal Concept Analysis

1 Introduction

The volume of data stored in the web has undergone a significant and a continuous evolution. Several efforts focused therefore on knowledge retrieval (extraction) from data. In [4], the knowledge extraction from data is defined as an acquisition of new knowledge, which is potentially useful, from the hidden facts within great amounts of data. One of the main processes of the knowledge extraction is based on data mining. This operation collects several tasks such as prediction, clustering and supervised classification. The latter can be performed by methods based on neural networks, decision trees, nearest neighbor, support vector machines or Formal Concept Analysis (FCA). Several reasons lead to use the classification method based on FCA, among them the ability of formal concepts to process a significant quantities of data and to simplify the prediction of classes [17].

Supervised classification based on FCA consists in building functions or models called classifiers from data to predict classes for future data. It aims at extracting the classification rules based on the concepts generated previously from data [19,21].

The whole process is performed in two main steps: a **training step** where a classifier is built to describe a predetermined set of object classes from a training set. A **classification step** where trained classifiers are used to assign a class to each new object. In this paper, we focus only on supervised classification based on FCA. Particularly, we propose a new taxonomy of classification methods based on FCA. The paper is organized as follows : we present basic notions related to FCA in Sect. 2. Sect. 3 describes new taxonomy of supervised classification methods. In Sect. 4, we make a comparative study of such methods.

2 Formal Concept Analysis

Originally, FCA is conceived by R.Wille as a "restructuring" of abstract lattice theory, i.e., making it closer to the needs of other branches of science. FCA represents a theoretical background for many applications in computer science. It covers various fields such as linguistics, information retrieval, text mining, knowledge representation and more recently knowledge extraction.

A formal context is represented as a triplet $K = (G, M, I)$ witch relies a finite set of objects G to a finite set of attributes M using binary relation I ($I \subseteq G \times M$) [7]. A formal context can be illustrated by a two-dimensional table where objects are presented by lines and attributes are presented by columns. All possible formal concepts are extracted from a formal context $K = (G, M, I)$. In fact, a formal concept is represented by a pair (A, B) such that $A \subseteq G$ and $B \subseteq M$, $A' = B$ and $B' = A$ where $A' = \{$m $\in$ M $\mid$ (g,m) $\in$ I $\forall$ g $\in$ A$\}$, i.e. the set of attributes common to all objects in A, and $B' = \{$g $\in$ G $\mid$ (g,m) $\in$ I $\forall$ m $\in$ B$\}$, i.e. the set of objects which have all attributes in B. A and B are called, respectively, extent and intent of the concept (A, B) [7]. The set of all concepts can be organized as a complete lattice of formal concepts, called Concept Lattice [7].

3 Supervised classification based on Formal Concept Analysis

Classification methods based on FCA uses either what we call exhaustive or combinatory approaches. In this section, we describe in detail the approaches and we give an overview of existing methods for each approach.

3.1 Exhaustive classification approach

Exhaustive methods are characterized commonly by the use of one single classifier. However, they vary between them according to the criteria used on concepts selection and the size of lattices outlining formal concepts. We distinguish overall methods based on complete lattices and others based on sublattices [21].

Among the works that have focused on the classification using a complete lattice as research space, we can cite GRAND [20], RULEARNER [22], GALOIS [2], CBALattice [8], NAVIGALA [24], HMCS-FCA-SC [5] and SPFC [9].

These methods carried out the validation of the characteristics associated to each concepts in the lattices level by level. The navigation in the lattice of concepts starts from the minimal concept where all the concepts of the lattice are considered as candidates without having an idea on their validity.

GRAND[3] and GALOIS are the first methods which use complete lattices of formal concepts. They build a complete lattice using an incremental algorithm and it updates the lattices by adding new nodes and removing redundant connections. Then, GRAND chooses the more specific rules to be applied for each object [20].

---

[3] Graph-based induction

GALOIS also builds a complete lattice in an incremental and ascending way. In the classification step, the system computes the similarity between new object and each concept. The similarity between an object and a concept is the number of common attributes verified by the object [2].

Other classification systems based on FCA are developed following GRAND such as RULEARNER and NAVIGALA. NAVIGALA[4] uses an object context described by numerical vectors of fixed size. These vectors are stored in a discrete table which then becomes binary [24]. RULEARNER, in a similar setting, uses a complete concept lattice as a research space in classification rules. During the classification, it uses majority voting to determine the classes of new objects [22].

CBALATTICE builds a complete lattice and applies association rules to extract classification rules. The method is incremental and progressive. Every increase in the number of objects, attributes, and classes can be handled efficiently [8].

HMCS-FCA-SC[5] also uses a complete lattice and existing information in formal concepts. However, unlike CBALATTICE, it aims to create a model of hierarchical classification. In addition of that, HMCS-FCA-SC employs a cosine measure [6] between new example and the selected concepts for classification [5].

After the construction of the lattice, SPFC[7] assigns to each concept a score which indicates the relevance degree of the concepts. Then it looks for neighbors of relevant concepts. The unknown objects will be classified in the classes of their neighbors [9].

The limit of the methods based on complete lattice consists in the exponential complexity of their training algorithms in terms of time and memory resources.

In order to overcome this problem, other efforts such as LEGAL [15], CIBLE [19], CLNN & CLNB [25], IPR [16], CLANN [23] and CITREC [3], MCSD-FCA-PS [1] use sublattices instead of complete lattice as search space. A sublattice of concepts is a mathematical structure which represents a part of the concept lattice in a selective way [7]. LEGAL[8] is a method which relies on several training parameters to build a sublattice. During the learning step, it builds an ordered set of concepts based on the class of each instance. The positive and negative instances are the instances labeled by a positive or negative class in the formal context. During classification Legal applies the majority vote [15].

CIBLE[9] operates in two succeeding steps: it starts with the construction of a sublattice from a binary context then it uses a similarity measure for the classification of new objects [19].

CLNN & CLNB[10] methods build a sublattice in a decreasing way. They incorporate then respectively a Naive Bayes classifier and a Nearest Neighbors

---

[4] Navigation into Galois Lattice
[5] Hierarchical Multi-label Classifier System - FCA with Similarity Cosine
[6] the similarity between two vectors with n dimensions by determining the cosine of the angle between them
[7] Classification by Selecting Plausible Formal Concepts in a Concept Lattice
[8] Learning with Galois Lattice
[9] Concept Induction Based Learning
[10] Concept Lattices Nearest Neighbors and Concept Lattices Naive Bayes

classifier in each node of the sublattice built. CLNN & CLNB use the same technique of vote (majority vote) in the classification step [25].

Another classification system CITREC which uses sublattices is proposed in [3]. CITREC builds the lattice starting from a reduced context containing only a representative object of each class [3]. In the classification step CITREC uses the majority vote in the same way as the methods CLNN & CLNB.

On the other side, CLANN[11] starts by building a sublattice in the training step and processes data which have only two classes. It uses then straightforward sublattice to build a neural networks which makes the classification [23].

MCSD-FCA-PS[12] has the distinction of processing sequential data. Complex sequential data are mapped onto pattern structures whose projections are used to build a pattern concept lattice. In fact, pattern structures are a generalization of formal concept analysis designed to deal with complex objects descriptions when an object is not associated to a set of attributes. MCSD-FCA-PS select relevant patterns that can be used for the classification step [1,13,12].

IPR[13] is a method which introduces the coverage of concepts. It selects from the lattice all the relevant concepts which can help to better classification. The choice of relevant concepts is based on greedy algorithm [16]. Then to classify each new object, IPR searches rules with a premise that coincides with the object attribute. The applied rules are the most weighted rules for the involved object.

The classification methods based on sublattices proceeds in the same way as methods based on complete lattices. However, using sublattices can reduce the substantial number of generated rules and keep the most relevant among them. Such proceeding leads to reduce significantly the training time, however it causes a loss of information.

Several drawbacks are observed in exhaustive methods described above. In fact, besides to the high complexity, using one single weak classifier may not be the best solution for classification. The use of many classifiers can give better results. Subsequently, the researchers move towards the integration and the use of combinatory classification methods which are based on the ensemble methods in order to improve other operations of a low single classifier (a single learner).

3.2 Combinatory classification approach

Unlike exhaustive methods which use one classifier, combinatory methods employs many classifiers which are then combined by the techniques of votes.

In this context, various methods have been proposed among them there are methods based on sequential training like BFC [18], BNC [17] and methods based on parallel training such as DNC [17], FPS-FCA [14] and RMCS [10].

The sequential training consists in generating classifiers sequentially. In other words, a classifier is generated only after the generation of its predecessor.

---

[11] Concept Lattice-based Artificial Neural Network
[12] Mining Complex Sequential Data by means of FCA and Pattern Structures
[13] Induction of Production Rules

For example, BFC[14] builds from a formal context a cover formed only by relevant concepts. The latter is based on boosting which is an adaptive approach based on the use of several classifiers of the same model [11]. These classifiers use voting techniques in order to classify correctly the objects wrongly classified. The idea of BFC consists in assigning, initially, equal weights for the training examples among them subset is selected randomly. At this point, a relevant concept is extracted from a subset by selecting the attribute which minimizes Shannon entropy[15]. BFC generates then a classification rule deducted from the relevant concept (extracted from subset) and updates the weights of training examples. This process is rebuilt recursively to produce the final classifier [18].

The method BNC[16] proceeds similarly to BFC method in generating classifiers and processing training data. However, unlike BFC which makes the processing of binary data, BNC handle nominal data in order to avoid the loss of information following the binary data representation [17].

The parallel training is based on Dagging [11], it consists in dividing data set into many groups. Classifiers are generated from the groups. DNC[17] method handles nominal data. DNC algorithm proceeds as follow: a random data is run in order to create a disjoint groups containing stratified data. A classifier of nominal concept [17] is then created for each group. Finally, the method defines as output a combination of classifiers made by a vote technique [17].

FPS-FCA[18] is a symbolic classification method based on Pattern structures and FCA, witch deal with complex input objects like logical formulas, graphs, strings and tuples of numerical. In presence of large data sets, it offers natural approximation tools (projections) and achieves classification in a parallel way by dividing the initial data set [14,13,6,12].

RMCS[19] generates a set of classifiers parallely. It allows to create a classifier based on its neighbors. The classifier realizes correctly the object classification when it is classified correctly within its neighbors. RMCS starts with the construction of a classification table from a formal context. In this table, RMCS matches a set of classifiers to the objects set existing in the context. Then, RMCS looks for the neighbors of the test set of objects using a metric of similarity and selects classifiers that have the maximum number of neighbors found. The selected classifiers are then recommended for classification [10].

4 Discussion

As mentioned previously, methods based on FCA are gathered in two main categories: **exhaustive** methods and **combinatory** ones. Methods of each category vary from each other in several aspects and share some others. Exhaustive

---

[14] Boosting Formal Concepts
[15] The information quantity contained or supplied by a source of information
[16] Boosting Nominal Concepts
[17] Dagging Nominal Concepts
[18] Fitting Pattern Structures to Knowledge Discovery in Big Data
[19] Recommender based Multiple Classifier System

methods generate only one ordinary classifier for the classification of the objects. Table 1 presents exhaustive methods cited previously. In tables 1 and 2, n denote the number of objects and m the number of attributes. In order to identify characteristics of each method, we chose six criteria that seem to be the most distinctive.

System	GRAND	CIBLE	IPR	CITREC	CLANN	MCSD-FCA-PS
Structure concepts	Complete lattice	Sub-lattice	Coverage	Sub-lattice	Sub-lattice	Sub-lattice
Data	Binary	Numerical	Binary	Binary	Binary	Sequential data
Selection concepts	Coherence maximal	Function selection	Entropy Shannon	Support	Algorithms heuristical	Pattern structures
Combination	No	No	No	No	No	No
Classification	Majority vote	K-PPV	Weighted rules	Vote	Neural network	Projections
Theoretical complexity	$O(2^k \cdot k^4)$ $k = min(m,n)$	$O(\|L\| \cdot m^3)$ $\|L\|$=sublattice	$O(n^2 \cdot m^2 \cdot nm)$	$O(2^m \cdot n)$	$O(2^{min(m,n)})$	$O(\|L\| \cdot \|A\| \cdot m^3 \cdot n)$ m=maximium sequence size A=alphabet of sequence letters

Table 1. Theoretical comparison of exhaustive methods

As shown in Table 1, exhaustive methods have exponential complexity. It is mainly due to a navigation in the totality of the research space.

On the other side, combinatory methods share the classification process in different classifiers a combination method. The problem is thus divided into many sub-problems. Similarly, to Table 1, Table 2 provides a description of combinatory methods. We used the same criteria in two tables for comparative reasons.

Tables 1 and 2 show that GRAND, IPR, CITREC, CLANN, BFC and RMCS handle binary data, BNC and DNC manipulate nominal data while CIBLE handle numerical data. However FPS-FCA and MCSD-FCA-PS differ from the previous methods by their capacity to handle complex data like graphs and sequential data. BNC and DNC use the informational gain to select concepts, while IPR and BFC use Shannon entropy. Concerning CLANN, it utilizes heuristic algorithms for the selection.

In classification process, GRAND, CITREC and DNC use the majority voting. The weighted voting is applied in IPR, BFC and BNC. However, CLANN differ from other methods by the use of neural networks.

The combining technique (cf. section 3.2) contributed strongly in optimizing the complexity. In fact, combinatory methods generate classifiers sequentially and have a polynomial logarithmic complexity. Similarly, methods generating parallel classifiers reach a comparable complexity in the order of $nmlog(n)$ for RMCS method, nm/k for FPS-FCA method and n for DNC.

System	BFC	BNC	DNC	RMCS	FPS-FCA
Concepts structure	Sub-lattice	Sub-lattice	Sub-lattice	Complete lattice	Sub-lattice
Data	Binary	Nominal	Nominal	Binary	Graphs, formulas, strings
Concepts selection	Shannon entropy	Gain informational	Gain informational	Distance euclidean	Relevant patterns
combination	Boosting	Boosting	Dagging	Dagging	Dagging
Classification	Weighted vote	Weighted vote	Majority vote	Maximum number of neighbors	Hypotheses
Theoretical complexity	$O(n\log(n)+nm)$	$O(n\log(n)+nm)$ m=nominal attribute	$O(n')$ n'=size of stratified samples	$O(nm\log(n))$	$O(nm/k)$ k=number of processors

Table 2. Theoretical comparison of combinatory methods

5 Conclusion

In this paper, we focused on supervised classification of data based on FCA. We presented firstly exhaustive classification methods which are divided into methods based on complete lattices and methods based on sublattices. Secondly, we described combinatory classification methods which are themselves divided into methods based on sequential training and others based on parallel training.

Our future work will be based on the complexity and move towards combinatory methods that offer reasonable complexity, especially methods that generate parallel classifiers.

References

1. Buzmakov, A., Egho, E., Jay, N., Kuznetsov, S.O., Napoli, A., Raïssi, C.: On mining complex sequential data by means of fca and pattern structures. International Journal of General Systems 45(2), 135–159 (2016)
2. Carpineto, C., Romano, G.: Concept data analysis: Theory and applications. John Wiley & Sons (2004)
3. Douar, B., Latiri, C.C., Slimani, Y.: Approche hybride de classification supervisée à base de treillis de galois: application à la reconnaissance de visages. In: Extraction et Gestion des Connaissances. pp. 309–320 (2008)
4. Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R.: Advances in knowledge discovery and data mining (1996)
5. Ferrandin, M., Nievola, J.C., Enembreck, F., Scalabrin, E.E., Kredens, K.V., Ávila, B.C.: Hierarchical classification using fca and the cosine similarity function. In: Proceedings on the International Conference on Artificial Intelligence. p. 1. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp) (2013)
6. Ganter, B., Grigoriev, P.A., Kuznetsov, S.O., Samokhin, M.V.: Concept-based data mining with scaled labeled graphs. In: International Conference on Conceptual Structures. pp. 94–108. Springer (2004)

7. Ganter, B., Stumme, G., Wille, R.: Formal concept analysis: foundations and applications. springer (2005)
8. Gupta, A., Kumar, N., Bhatnagar, V.: Incremental classification rules based on association rules using formal concept analysis. In: Machine Learning and Data Mining in Pattern Recognition, pp. 11–20. Springer (2005)
9. Ikeda, M., Yamamoto, A.: Classification by selecting plausible formal concepts in a concept lattice. In: Workshop on Formal Concept Analysis meets Information Retrieval. pp. 22–35 (2013)
10. Kashnitsky, Y., Ignatov, D.I.: Can fca-based recommender system suggest a proper classifier? What can FCA do for Artificial Intelligence (2015)
11. Kotsianti, S., Kanellopoulos, D.: Combining bagging, boosting and dagging for classification problems. In: Knowledge-Based Intelligent Information and Engineering Systems. pp. 493–500. Springer (2007)
12. Kuznetsov, S.O.: Learning of simple conceptual graphs from positive and negative examples. In: European Conference on Principles of Data Mining and Knowledge Discovery. pp. 384–391. Springer (1999)
13. Kuznetsov, S.O.: Machine learning and formal concept analysis. In: International Conference on Formal Concept Analysis. pp. 287–312. Springer (2004)
14. Kuznetsov, S.O.: Fitting pattern structures to knowledge discovery in big data. In: International Conference on Formal Concept Analysis. pp. 254–266. Springer (2013)
15. Liquiere, M., Mephu Nguifo, E.: Legal: Learning with galois lattice. Journées Françaises sur l'Apprentissage, Lannion pp. 93–113 (1990)
16. Maddouri, M.: Towards a machine learning approach based on incremental concept formation. Intelligent Data Analysis 8(3), 267–280 (2004)
17. Meddouri, N., Khoufi, H., Maddouri, M.: Parallel learning and classification for rules based on formal concepts. Procedia Computer Science 35, 358–367 (2014)
18. Meddouri, N., Maddouri, M.: Boosting formal concepts to discover classification rules. In: Next-Generation Applied Intelligence, pp. 501–510. Springer (2009)
19. Njiwoua, P., Mephu Nguifo, E.: Améliorer l'apprentissage à partir d'instances grâce à l'induction de concepts: le système cible. Revue d'intelligence artificielle 13(2), 413–440 (1999)
20. Oosthuizen, G.D.: The use of a lattice in knowledge processing (1980)
21. Prokasheva, O., Onishchenko, A., Gurov, S.: Classification methods based on formal concept analysis. FCAIR 2012–Formal Concept Analysis Meets Information Retrieval p. 95 (2013)
22. Sahami, M.: Learning classification rules using lattices. In: Machine Learning: ECML-95, pp. 343–346. Springer (1995)
23. Tsopzé, N., Nguifo, E.M., Tindo, G.: Clann: Concept lattice-based artificial neural network for supervised classification. In: Concept Lattice and their applications. vol. 331. Citeseer (2007)
24. Visani, M., Bertet, K., Ogier, J.M.: Navigala: An original symbol classifier based on navigation through a galois lattice. International Journal of Pattern Recognition and Artificial Intelligence 25(04), 449–473 (2011)
25. Xie, Z., Hsu, W., Liu, Z., Lee, M.L.: Concept lattice based composite classifiers for high predictability. Journal of Experimental & Theoretical Artificial Intelligence 14(2-3), 143–156 (2002)