Machine Learning and Artificial Intelligence
A.J. Tallón-Ballesteros and C. Chen (Eds.)

doi:10.3233/FAIA200762

Survey on Formal Concept Analysis Based Supervised Classification Techniques

Hayfa AZIBI [a,1], Nida MEDDOURI [b] and Mondher MADDOURI [c]
[a] *LIPAH, Faculty of Sciences of Tunis, University of Tunis El Manar, Tunisia*
[b] *GREYC-CNRS UMR 6072, University of Caen Normandy, France*
[c] *College of Business, University of Jeddah, Kingdom of Saudi Arabia*

Abstract. Classification is a data mining task and which is a two-phase process: learning and classification. The learning phase consists of constructing a classifier or a model from a labeled set of objects. The classification phase consists classifying new objects by using the generated classifier. Different approaches have been proposed for supervised classification problems through Formal Concept Analysis, and which is a mathematical theory to build upon hierarchies of formal concepts. The proposed approaches in literature rely on the use of single classifier and ensemble methods. Single classifier methods vary between them according to different criteria especially the number of formal concepts generated. We distinguish overall complete lattice methods, sub-lattice methods and concept cover methods. Methods based on ensemble classifiers rely on the use of many classifiers. Among these methods, there are methods based on sequential training and methods based on parallel training. However, with the large volume of data generated from various sources, the process of knowledge extraction with traditional methods becomes difficult. That's why new methods based on distributed classifier have recently appeared. In this paper, we present a survey of many FCA-based approaches for classification by dividing them into methods based on a mono-classifier, methods based on ensemble classifiers and methods based on distributed classifiers. Different methods are presented and compared within this paper.

Keywords. Artificial intelligence, Data mining,, Machine learning, Supervised classification, Formal concepts analysis, Ensemble methods, Cloud

1. Introduction

The exploding volume and speed of data growth have triggered several challenges in many learning problems in real world. Classification is one of the most important tasks in Machine Learning. The classification problem aims to predict a class to which new data will fall under. In fact, the supervised classification analyzes the attributes and develops an accurate description or model for each class using descriptions submitted by attributes. Several classification algorithms were proposed in the literature and widely applied in practice. As references in the fields, we can highlight the Artificial Neural Network, Association Rule Mining, Formal Concept Analysis, Induction of Decision

[1]Corresponding Author: Hayfa Azibi, Faculty of Sciences of Tunis, University of Tunis El Manar, Tunisia. E-mail: hayfa.azibi@fst.utm.tn

Trees, Naïve Bayes and Support Vector Machine. Each supervised classification method is characterized by some features that can be fitted to some classification tasks.

Formal Concept Analysis (FCA) [1] is a popular method of Machine Learning methods [2]. It is a mathematical theory which builds upon hierarchies of formal concepts. Also, FCA is a theoretical framework which structures a group of objects and their attributes. The classification approach based on FCA is divided into two steps: learning step and classification step. In the learning step, a classifier is built by means of analysis of objects described by attributes in the training set. Each object is assumed to belong to a pre-defined class represented by a particular attribute in the training set. In the classification step, the model built in the first stage is used to classify the new objects.

This article provides a comparative study of FCA-based classification methods. In literature, several studies had carried out this comparative study. The authors in [3] carried out a theoretical and experimental comparative studies of some classification methods based on FCA. New methods have appeared in [4] which are based on a single classifier. The authors classified the classification methods based on the FCA by evoking the notions of complete lattice, sub-lattice and cover of the concept. Other algorithms are based on the taxonomy of [2] of existing supervised classification methods. This taxonomy is divided into two categories: exhaustive methods and combinatory methods. The first category is characterized by the use of a single classifier. The second contains the learning methods which exploit the paradigms of ensemble learning. Hence, this article presents a comparison of the FCA-based classification methods cited in existing work with those that have recently emerged. We introduce a new category of methods which is based on a distributed classifier. The paper is organized as follows: in section 2 we introduce methods based on mono-classifier. Section 3 introduces methods based on ensemble classifiers. Section 4 presents the methods based on distributed classifiers. In section 5, we discuss such methods. Finally, we present our conclusion.

2. Methods based on mono-classifier

Several classification methods based on FCA, were presented in literature. Many FCA-based classification algorithms that generate a complete lattice have been developed, we can cite GRAND [5,6], GALOIS [7,8], RULEARNER [9] and NAVIGALA [10].

GRAND builds a complete lattice. All concepts are presented in the lattice except the supremum or the infimum each one of them is an empty set. To update the lattice, for each new object that has attributes shared lattice nodes, nodes with common attributes will be inserted, in the meantime, all redundant connections will be removed. It induces the most accurate rules in order to be applied on each new object.

GALOIS is a system that provides an incremental aspect of lattice construction. It uses two different methods to determine classes of new objects. The system performs a similarity calculation between the new object and each concept. This similarity is the concept properties verified by the object. Finally, GALOIS assigns to the object the class of the most similar concept.

RULEARNER constructs a full-concept lattice. In the learning phase, this system builds a set of rules that overlap the object nodes of the lattice. Besides, to classify new objects with an ordered list, RULEARNER uses these rules by keeping the order. However, it uses majority voting for unordered list.

NAVIGALA is a recognition system that was developed to recognize noisy graphical objects and symbol images by navigating through the complete lattice like navigating in a classification tree [10].

In [11], the authors proposed an incremental learning method for mining sequential patterns to find different human behavioral patterns in non-stationary smart environments. The input data are labeled sequences that gradually arrive from a sensor. If a lattice is not found before, the presented method makes a lattice initialization. For each element of the new training data set, an iteration of the lattice checks whether the iterated concept should be updated, created or ignored [11].

FCA-based query expansion was discussed in [12]. This study is based on the extraction of description topics from documents. In fact, a set of retrieved documents is obtained based on a query against a set of documents to perform the expansion. The description topics defined as intrinsic concepts in a document are extracted from the recovered documents. Using the retrieved documents as objects and the description topics as attributes, a lattice is constructed as the possible expansion space. The expanded query will be generated by the selected lattice nodes.

Despite the several systems of lattice concept-based classification, their problem remains in time and space complexities. This common limit is due to the navigation in the whole space search.To solve this issue, many researches presented approaches based on sub-lattice classification. A sub-lattice is a partial part of Galois lattice. The classification process is the same for a complete lattice and sub-lattice methods but the major difference between them is about how many formal concepts are generated. LEGAL [8], CIBLe [13], CLNN & CLNB [14] and CLANN [15] build a sub-lattice. The sub-lattice generation contributed drastically to the reduction of theoretical complexities and execution time.

To build a sub-lattice, LEGAL applies two learning parameters. The objects of the initial formal context will be divided into positive objects and negative objects. For each new node, LEGAL begins by constructing its sub-nodes. Valid nodes are then retrieved using learning parameters. These valid nodes are characterized by a great number of positive objects. The algorithm ends when there are no valid nodes.

CIBLe starts with the construction of a sub-lattice. It gives a numerical re-description to the training data. In its classification step, CIBLe performs a similarity calculation to classify new objects. In practice, it uses three different measures: Manhattan distance, Mahalanobis distance and Euclidean distance.

CLNN & CLNB integrate respectively two classifiers, the Naïve Bayes classifier and the Nearest Neighbors classifier, in the lattice concept. They also use the majority vote to classify new objects.

CLANN builds a sub-lattice in the training phase and only data with two classes are handled. The obtained sub-lattice will be used to construct neural networks that perform classification.

The authors in [16] proposed a classification method based on FCA which applies the minimum description length (MDL) to select concepts. Target class code tables are used individually to get compression objects. For classification, the attributes of an object are covered by sets of elements found in code tables of classes. Finally, coding lengths are calculated for all classes. The class that has minimum coding length is chosen.

A concept cover is defined as a part of lattice which contains only relevant concepts. To build cover concepts, IPR [17] resorted to the greedy algorithm. For classification, IPR looks for rules with the premise that matches attributes of the new object. Rules that were applied represent the most weighed ones for the involved object. CITREC [18] is another cover—based classification method. The first step of CITREC is to convert numeric and nominal attributes to binary ones. Then, the creation of a new context (the reduced context) where the objects and the classes of the different objects of the training set are equal. Next, the lattice is built using the reduced context. To classify new objects, CITREC uses the majority vote.

Different supervised classification methods based on FCA were presented in this section: complete lattice methods, sub-lattice methods and concept cover methods. Concept lattice and sub-lattice proceed similarly. But, using sub-lattice is feasible due to its running time compared to the concept lattice. In fact, this feature leads to the generation of the relevant rules but this causes a loss of information. The problems in the presented methods remain in the use of a single classifier, the high complexity and the type of handled data which is binary for almost all systems. As a result, many researches in literature oriented to the combination of classification methods based on the ensemble methods the best known of which are boosting and bagging.

3. Methods based on ensemble classifiers

There is a growing realization that the use of ensemble classifiers can be more effective than the use of single classifiers. Why rely on the best single classifier, while we can obtain the most accurate and reliable result from a combination of several? This is the reasoning behind the idea of ensemble methods. Several classifiers based on ensemble methods were developed in literature. There are two categories: methods based on sequential training and methods based on parallel training. The difference between the two categories is that the first one generates classifiers sequentially but the second method generates parallel classifiers.

In this context, BFC [19] and BNC [20] are two methods based on sequential training that were proposed in literature. BFC is a method based on FCA and also benefits from boosting algorithms. The basic idea of BFC consists in selecting a group of data from the learning set after assigning equal weights to the training objects. Then, BFC extracts the relevant formal concepts within the subset. To classify the learning data, the BFC method uses the training objects weights. This process is repeated until getting the final classifier. For BNC, it proceeds in the same way as BFC. However, what differs between BFC and BNC is the data type and attribute selection. BFC makes the learning from binary data but BNC handles nominal data. For attribute selection, BFC uses Shannon's Entropy while BNC uses informational gain.

FPS-FCA [21], DNC [22], RMCS [23] and B-RCL [24] are based on parallel training. FPS-FCA divides the training set into many subsets. FPS-FCA uses the obtained subsets to generate formal contexts in order to extract classification rules. DNC builds several parallel classifiers. In this case, each classifier is constructed using the same learning algorithm. DNC creates disjoint and stratified samples. On each sample, CNC (Classifier Nominal Concept) is then constructed [22]. The classifiers' outputs are finally combined by a majority vote. RMCS is also a method based on parallel training. RMCS begins with the construction of a classification table using a formal context. Then, RMCS assigns classifiers to the objects that exist in the context. After matching, it searches the test set object neighbors by means of a similarity metric. The classifiers that are selected for classification are those which have more neighbors that were found [23]. The author of [24] proposed the fusion of Random Conceptual Coverage Learner with bagging paradigm. RCL differs from other FCA coverage methods in attribute selection. RCL performs the selection randomly from the training set. B-RCL was proposed to reduce the variance caused by RCL.

4. Methods based on distributed classifiers

In recent decades, the volume of data generated from different sources flows continuously. Hence, the extraction of knowledge from numerous data sources using mono-classifier methods and ensemble learning methods becomes a difficult task. The existing algorithms are not scalable to the huge new and larger datasets for knowledge extraction and representation. To solve this issue, frameworks for big data applications are developed [25]. However, these frameworks are based on a distributed environment like Cloud Computing [26]. In this field, several distributed data mining tools were developed. In [27], the authors introduced a cloud-based framework to implement home diagnostic service. The user submits a query which contains the disease information. A dispatcher selects nodes to determine the medical records corresponding to the request. The dispatcher then merges the search results and passes them on to a data analysis cluster. A lattice will be constructed according to the medical records retrieved and reveals the relationships between diseases with common symptoms [27].

The authors of [28] presented a Multi-Cloud Service Composition approach which based on FCA. In fact, from each lattice the requested services are extracted. Then, the use of the lattice for filtering candidate clouds according to providers that were selected [28]. Finally, there is a selection of the appropriate and optimal cloud set from which the best services are selected.

The work presented in [29] is a Distributed Classifier Nominal Concept. This method is a distributed version of CNC, which handles nominal data. Dist-CNC was implemented using Distributed Weka Spark which is a distributed framework for weka. During the learning phase, the master node divides the input data into partitions and then distributes the training task and the partitions obtained to the slave nodes. Slave nodes apply CNC on each partition and return results to master node. To evaluate the model, the master node distributes the model formed to the slave nodes. Then, each slave node uses its partitions to evaluate the model. The final results are merged and then returned.

5. Discussion

Tables 1 and 2 show a comparison between supervised classification methods based on FCA by category. The comparison criteria chosen show the characteristics of each method. As shown in tables 1 and 2, these methods handle various data types as binary, numeric and nominal data. In table 1, the methods construct complete lattices, sub-lattice or cover concepts. These methods classify the datasets which contain several classes with the exception of LEGAL, RULEARNER and CLANN. For lattice construction, these methods use algorithms to generate concept lattices. These algorithms can be incremen-

Table 1. Comparison between mono-classifier based methods

Methods	GRAND	GALOIS	RULEARNER	NAVIGALA	LEGAL	CIBLe	CLNN & CLNB	CLANN	IPR	CITREC
Data type	Binary	Nominal	Nominal	Binary	Binary	Numeric	Numeric	Binary	Binary	Binary
Number of classes	multi-class	multi-class	2 classes	multi-class	2 classes	multi-class	multi-class	2 classes	multi-class	multi-class
Construction lattice algorithm	Oosthuizen	Carpineto and Romano	Oosthuizen	Bordat extension	Bordat	Modified Bordat	Top-down approach	Modified Bordat	Heuristic approach	Godin
Concepts structure	Complete lattice	Complete lattice	Complete lattice	Complete lattice	Sub-lattice	Sub-lattice	Sub-lattice	Sub-lattice	Cover concepts	Cover concepts
Concept selection	Maximally complete concepts	Maximally complete concepts	Maximally complete concepts	Distance measure	Maximally complete concepts	Lattice level Entropy	Support Precision	Heuristic algorithms	Entropy	Support
Incremental	Yes	Yes	Yes	Yes	No	No	No	No	Yes	Yes
Learned knowledge	Rules	Relevant concepts	Ordered and unordered rules	Concepts	Relevant concepts	Relevant concepts	Rules	Relevant concepts	Rules	Rules
Classification	Vote	Similarity or vote	General rule	Navigation in a GALOIS lattice like navigation in the decision tree	Vote	k-nearest neighbors algorithm	Verified rule + vote	Neural networks	weighted rules	Vote
Complexities	O(2^l × l^4 with l is the minimum between n and m.	O(3^m × 2^n × n) < O(3^{2m} × n)	Idem to GRAND.	O($\|L\|$ × n^3) + O(nm^2) with $\|L\|$ the number of concepts	O(($\|L\|$ × n (1-α)) with $\|L\|$ the number of concepts and α the validity criteria	O ($\|L\|$ × m^3 with $\|L\|$ the number of concepts of the sub-lattice	O ($\|L\|$ × n × (1-α))	O($2^{min(n,m)}$)	O(n^2 × m^2 × (m+n))	O(2^m × n)

Table 2. Comparison between ensemble based methods

Methods	BFC	BNC	FPS-FCA	DNC	RMCS	B-RCL
Concepts structure	Cover	Cover	Sub-lattice	Cover	Complete lattice	Sub-lattice
Data type	Binary	Nominal	Nominal	Nominal	Binary	Nominal
Concept selection	Shannon entropy	Informational gain	Relevant patterns	Informational gain	Distance Euclidian	Random coverage
Learned knowledge	Rule	Rules	Graph pattern structure	Rules	Concepts	Rules
Classification	Weighted vote	Weighted vote	Hypotheses	Majority vote	Maximum number of neighbors	Majority vote
Ensemble	Sequential	Sequential	Parallel	Parallel	Parallel	Parallel
Complexity	O(nlog(n)+ nm)	O(nlog(n)+ nm) with m= nominal attribute	O(nm/k) with k is the number of processors	O(n') with n'= stratified samples	O(nmlog(n))	O(N^3) with N is the number of base classifiers

tal or non-incremental. All methods in tables 1 and 2 use the concept selection to induce rules in order to classify new instances through these rules.

However, these rely on different selection measures such as the Informational Gain for BNC and DNC, the support for CITREC and the Shannon entropy for BFC. IPR and CLNN & CLNB uses support, precision and recall to obtain concepts.

Methods based ensemble classifiers use multiple classifiers that are combined by vote techniques. These methods choose to represent the learned knowledge by relevant concepts or rules. In the classification phase, each system uses its appropriate method to determine a class for each new object. The majority vote is applied by GRAND, CITREC, LEGAL, BFC and DNC. It may also be used for GALOIS that also applies the similarity calculation. CLNN & CLNB applies voting strategy and verified rules for prediction. To predict a class for a new object, RULEARNER makes a selection of rules by respecting the order of the antecedents, CIBLe applies K-Nearest Neighbors algorithm and CLANN utilizes neural networks algorithm for classification. IPR uses weighed rules and BNC uses the weighed vote. RMCS classifies new examples by looking for the maximum number of neighbors and FPS-FCA uses hypothesis.

Tables 1 and 2 also propose a comparison of the theoretical complexities of different classification methods based on FCA where n is the number of objects and m is the number of attributes. As shown in table 1, all methods have an exponential complexity. Sub-lattice methods reduce this complexity because the build is a part of the lattice. Cover concepts methods have minimal complexity thanks to the generation of only the most relevant concepts.However, for ensemble classifiers, parallel methods like DNC, RMCS and FPS-FCA have a linear complexity, a polynomial logarithmic complexity and a polynomial complexity, respectively. For BFC and BNC, there is a complexity optimization that reaches a polynomial logarithmic order. The extraction of knowledge from large data sets still a challenge and a difficult task for traditional data mining tool. Distributed classifiers constitute a solution to answer this problem.

6. Conclusion

In this paper, we presented several FCA-based classification methods. First, we introduced methods based on mono-classifier that regroup the methods based on full lattice, sub-lattice and cover concept. Second, we presented methods based on classifiers ensemble. They rely on the use of many classifiers by parallel or sequential training. Finally, we introduced methods based on distributed classifiers to answer the problem of knowledge extraction from large data sets.

References

[1] Ganter B, Wille R. Formal Concept Analysis: Mathematical Foundations. Springer Verlag; 1999.

[2] Trabelsi M, Meddouri N, Maddouri M. New taxonomy of classification methods based on Formal Concepts Analysis. In: Proceedings of the 5th International Workshop "What can FCA do for Artificial Intelligence"? co-located with the European Conference on Artificial Intelligence. vol. 1703; 2016. p. 113–120.

[3] Fu H, Fu H, Njiwoua P, Nguifo EM. A Comparative Study of FCA-Based Supervised Classification Algorithms. In: Proceeding of Second International Conference on Formal Concept Analysis; 2004. p. 313–320.

[4] Meddouri N, Maddouri M. Classification methods based on formal concept analysis. In: Proceedings of the 6th International Conference on Concept Lattices and Their Applications; 2009. p. 9–16.

[5] Oosthuizen GD. The use of a lattice in knowledge processing. University of Strathclyde. Glasgow, Scotland, UK; 1988.

[6] Oosthuizen G. The application of concept lattice to machine learning. Dept Comput Sci, Univ Pretoria, Pretoria, South Africa, Tech Rep CSTR. 1996;94(01).

[7] Carpineto C, Romano G. GALOIS: An Order-Theoretic Approach to Conceptual Clustering. In: Proceedings of the Tenth International Conference on Machine Learning. Morgan Kaufmann; 1993. p. 33–40.

[8] Njiwoua P, Nguifo EM. Forwarding the choice of bias LEGAL-F: using feature selection to reduce the complexity of LEGAL. Proceedings of BENELEARN-97, ILK and INFOLAB. 1997;p. 89–98.

[9] Sahami M. Learning Classification Rules Using Lattices (Extended Abstract). In: Proceedings of the 8th European Conference on Machine Learning. vol. 912 of Lecture Notes in Computer Science. Springer; 1995. p. 343–346.

[10] Visani M, Bertet K, Ogier J. Navigala: an Original Symbol Classifier Based on Navigation through a Galois Lattice. International Journal of Pattern Recognition and Artificial Intelligence. 2011;25(4):449–473.

[11] Hao J, Bouzouane A, Gaboury S. An incremental learning method based on formal concept analysis for pattern recognition in nonstationary sensor-based smart environments. Pervasive and Mobile Computing. 2019;59:101045.

[12] Yu H, Shi C, Bai Y, Zhang C, Hearne R. Query Expansion Based on Formal Concept Analysis From Retrieved Documents. Journal of Internet Technology. 2019;20(2):409–421.

[13] Njiwoua P, Nguifo EM. Améliorer l'apprentissage à partir d'instances grâce à l'induction de concepts: le système CIBLe. Revue d'intelligence artificielle. 1999;13(2):413–440.

[14] Xie Z, Hsu W, Liu Z, Lee ML. Concept lattice based composite classifiers for high predictability. Journal of Experimental and Theoretical Artificial Intelligence. 2002;14(2-3):143–156.

[15] Tsopzé N, Mephu Nguifo E, Tindo G. CLANN: Concept Lattice-based Artificial Neural Network for Supervised Classification. In: Proceedings of the 5th International Conference on Concept Lattices and Their Applications. vol. 331; 2007. .

[16] Makhalova T, Kuznetsov SO, Napoli A. A First Study on What MDL Can Do for FCA. In: CLA 2018 - The 14th International Conference on Concept Lattices and Their Applications; 2018. .

[17] Maddouri M. Towards a machine learning approach based on incremental concept formation. Journal of Intelligent Data Analysis. 2004;8(3):267–280.

[18] Douar B, Latiri C, Slimani Y. Approche hybride de classification supervisée à base de treillis de Galois: application à la reconnaissance de visages. In: Actes des 8èmes Journées Francophones en Extraction et

Gestion des Connaissances. vol. E-11 of Revue des Nouvelles Technologies de l'Information. Cépaduès-Éditions; 2008. p. 309–320.

[19] Meddouri N, Maddouri M. Boosting Formal Concepts to Discover Classification Rules. In: Proceeding of the 22rd International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems. vol. 5579 of Lecture Notes in Computer Science. Springer; 2009. p. 501–510.

[20] Meddouri N, Maddouri M. Adaptive learning of nominal concepts for supervised classification. In: Proceedings of the 14th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems. vol. 6276 of Lecture Notes in Computer Science. Springer; 2010. p. 121–130.

[21] Kuznetsov SO. Fitting pattern structures to knowledge discovery in big data. In: International Conference on Formal Concept Analysis. Springer; 2013. p. 254–266.

[22] Meddouri N, Khoufi H, Maddouri M. Parallel learning and classification for rules based on formal concepts. In: Proceedings of the 18th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems. Procedia Computer Science. Elsevier; 2014. p. 358–367.

[23] Kashnitsky Y, Ignatov DI. Can FCA-based Recommender System Suggest a Proper Classifier? arXiv preprint arXiv:150405473. 2015;.

[24] Ali MAT. Bagged Randomized Conceptual Machine Learning Method. College of Engineering. Qatar; 2018.

[25] Inoubli W, Aridhi S, Mezni H, Maddouri M, Nguifo EM. An experimental survey on big data frameworks. Future Generation Computer Systems. 2018;86:546 – 564.

[26] Mezni H, Abdeljaoued T. A cloud services recommendation system based on Fuzzy Formal Concept Analysis. Data and Knowledge Engineering. 2018;116:100 – 123.

[27] Lin W, Dou W, Zhou Z, Liu C. A cloud-based framework for Home-diagnosis service over big medical data. Journal of Systems and Software. 2015;102:192 – 206.

[28] Mezni H, Sellami M. Multi-Cloud Service Composition Using Formal Concept Analysis. Journal of Systems and Software. 2017;134(C):138152.

[29] Fray R, Meddouri N, Maddouri M. Cloud Implementation of Classier Nominal Concepts using DistributedWekaSpark. In: Supplementary Proceedings of ICFCA 2019 Conference and Workshops. vol. 2378 of CEUR Workshop Proceedings. CEUR-WS.org; 2019. p. 125–136.