A New Approach to Classification by Means of Jumping Emerging Patterns

Aleksey Buzmakov[1][2], Sergei O. Kuznetsov[2], and Amedeo Napoli[1]

[1] LORIA(CNRS-Inria NGE-Université de Lorraine),Vandoeuvre les Nancy,France
[2] National Research University Higher School of Economics, Moscow, Russia

Abstract. Classification is one of the important fields in data analysis. Concept-based (JSM) hypotheses are a well-known approach to this task. Although the accuracy of this approach is quite good, the coverage is often insufficient. In this paper a new classification approach is presented. The approach is based on the similarity of an object to be classified to the current set of hypotheses: it attributes the new object to the class that minimizes the set of new hypotheses when a new object is added to the training set. The proposed approach provides a better coverage in compare with the classical approach.

Keywords: Classification, Formal Concept Analysis, JSM-Hypotheses, Jumping Emerging Patterns, Experiments

1 Introduction

Data analysis applications play important role in nowadays scientific researches. One of the possible tasks is to predict object properties, for instance, prediction of a molecule toxicity. Objects can be described in different ways, one of them is by a set of binary attributes. For example, in chemistry domain, a molecule could be characterized by a set of functional groups, belonging to the molecule. Given a set of objects, labeled with several classes (like toxic and non toxic), the prediction task is to estimate the class of some unlabeled object.

Jumping emerging patterns (JEP) is a well studied and interesting approach to the classification[1, 2]. Given a set of classes, like toxic or non toxic molecule, a JEP is a set of characteristics describing a class in a unique way (in the same way as a "monothetic" property). For example, a set of functional groups say S is a JEP when all the database molecules, including all functional groups from S, are toxic. Most of the time, JEPs can be ordered, thanks to an ordering relation, and w.r.t. domain knowledge. In particular, this can be found in [3–5] where JEPs are studied through the so-called JSM-hypotheses.

Then, a classical way to classify an object w.r.t JEPs is to search for JEPs, describing the object, and if these JEPs are of the same class say k, then the object should be classified in k. If there is no such JEP or there are JEPs of different classes, the object remains unclassified. Although for the classical approach the prediction accuracy (the probability that the prediction is correct) is quite high, its coverage (the probability that the object attributed to any class by the classifier and this attribution is correct) is rather low. So a new method is proposed with comparable accuracy and much better coverage. The method relies on the MDL (minimal length description) principle, where the outcome class for an object is the class, minimizing the number of associated JEPs.

There are two main objectives in the paper. The first is to connect JEPs with JSM-hypotheses; and the second is to suggest a new classification approach, based on JEPs, and to check it experimentally.

The paper is organized as follows. In Section 2 definitions are introduced. Then Section 3 describes the classical and the new approaches to classification. Section 4 details the computer experiments and their results. And finally, Section 5 concludes the paper.

2 Definitions

2.1 Formal Concept Analysis and Pattern Structures

This section briefly introduces the main definitions on pattern structure in formal concept analysis (see [6]) and emerging patterns (see [1, 2]).

Definition 1. *A pattern structure is a meet-semilattice $(D, \sqcap)$. Elements of a set D are called patterns.*

Definition 2. *A pattern context is a triple $(G, (D, \sqcap), \delta)$, where G is a set of objects, $(D, \sqcap)$ is a pattern structure, and $\delta : G \to D$ is a mapping function from objects to their descriptions.*

The recently studied interval patterns [7] and the pattern structure given by sets of graphs [6] are examples of pattern structures.

Usually a formal context is introduced as follows [8].

Definition 3. *A formal context is a triple (G, M, I), where G is a set of objects, M is a set of attributes and $I \subseteq G \times M$ is a binary relation between G and M.*

A 'classical' formal context (G, M, I) could be considered as a special case of pattern context $(G, (D, \sqcap), \delta)$. The set of objects remains G, $D = 2^M$, with a semilattice operation corresponding to intersection of sets, and $\delta = g \in G \to \{m \in M | (g, m) \in I\}$. For instance a particular context is shown on Table 1. A mapping function δ maps the object g_1 to the set $\{\mathtt{m}_1, \mathtt{m}_2, \mathtt{m}_5, \mathtt{m}_6, \mathtt{m}_7\}$. For the sake of simplicity, all further examples will refer to classical contexts.

Objs\Attrs	m_1	m_2	m_3	m_4	m_5	m_6	m_7
g_1	x	x			x	x	x
g_2	x	x		x		x	x
g_3	x	x			x	x	
g_4		x	x				x
g_5	x			x	x	x	
g_6		x				x	x

Table 1: Formal Context (G, M, I).

Object	Class
g_1	k_1
g_2	k_1
g_3	k_2
g_4	k_2
g_5	k_2
g_6	?

Table 2: Labeling function.

A Galois connection associated to the context $(G, (D, \sqcap), \delta)$ is defined as:

$$A^{\diamond} = \sqcap_{e \in A} \delta(e), \qquad A \subseteq G \tag{1}$$

$$d^{\diamond} = \{e \in G | d \sqsubseteq \delta(e)\}, \qquad d \in D \tag{2}$$

For $a, b \in D$, $a \sqsubseteq b \Leftrightarrow a \sqcap b = a$, and the operation $(\cdot)^{\diamond\diamond}$ is a closure operator.

Definition 4. *A pattern $d \in D$ is closed iff $d^{\diamond\diamond} = d$.*

Definition 5. *Generator of a closed pattern $d \in D$ is a pattern $x \in D$, such that $x^{\diamond\diamond} = d$.*

Definition 6. *A pattern concept is a pair (A, d) such that $A \subseteq G$, $d \in D$, $A^{\diamond} = d$, $A = d^{\diamond}$. A is called the extent of the concept and d is called the intent. The intent of a formal concept is a closed pattern (while the extent A is a closed set of objects, i.e. $A^{\diamond\diamond} = A$).*

For example $(\{\mathtt{g}_1, \mathtt{g}_2\}, \{\mathtt{m}_1, \mathtt{m}_2, \mathtt{m}_6, \mathtt{m}_7\})$ is a concept w.r.t the context shown on the Table 1. One of the possible generators of its intent is $\{\mathtt{m}_2, \mathtt{m}_6, \mathtt{m}_7\}$.

2.2 Classification Concepts

The classification operation can be carried out in FCA using so-called hypotheses. In classification there are a set of classes K and a mapping function $\xi : G \to K \cup \{?\}$, where '?' means unknown class of an object.

Definition 7. *Given a certain class $k \in K$, we note the set of objects belonging to the class k as $G_{k+} = \{g \in G | \xi(g) = k\}$ and the set of objects, which are not belong to class k as $G_{k-} = \{g \in G | \xi(g) \neq k, \xi(g) \neq ?\}$. A hypothesis for class k is a pattern $h \in D$, such that $h^{\diamond} \cap G_{k-} = \emptyset$ and $\exists A \subseteq G_{k+} : A^{\diamond} = h$.*

For example, $\{\mathtt{m}_1, \mathtt{m}_2, \mathtt{m}_6, \mathtt{m}_7\}$ is a hypothesis for class k_1 because $\{\mathtt{m}_1, \mathtt{m}_2, \mathtt{m}_6, \mathtt{m}_7\}^{\diamond} = \{\mathtt{g}_1, \mathtt{g}_2\}$ contains objects of only one class.

In itemset mining Jumping Emerging Patterns (JEP) are used for classification [1, 2]. Although the usual definition of a JEP does not involve pattern structures, it can be convenient to introduce JEP w.r.t pattern structures.

Definition 8. *A pattern $d \in D$ is a JEP for a class $k \in K$ when $d^{\diamond} \neq \emptyset$ and $\forall g \in d^{\diamond}$, $\xi(g) = k$.*

According to definitions 7 and 5, a hypothesis for a class $k \in K$ is a JEP, whereas a JEP for a class $k \in K$ is a generator of some hypothesis for the class k. For the context on Table 1 and ξ function from Table 1 $\{\mathtt{m}_6, \mathtt{m}_7\}$ is a JEP for the class k and it is a generator for $\{\mathtt{m}_1, \mathtt{m}_2, \mathtt{m}_6, \mathtt{m}_7\}$, which is a hypothesis.

3 Classification

This section introduces classification by means of Jumping Emerging Patterns (JEP) in two different ways: the classical approach and the new approach.

For some class $k \in K$, H_{k+} is the set of all JEPs for class k and H_{k-} is the union of JEPs for all other classes. The union of all JEPs is denoted as $H = H_{k+} \cup H_{k-}$.

Definition 9. *A JEP $h \in H_{k+}$ describes an object $g \in G$ if $h \sqsubseteq g^{\diamond}$.*

According to the classical approach [3], a new object g_{new} should be attributed to the class $k \in K$ iff there is a JEP for the class k, describing g_{new} $(\exists h \in H_{k+} : h \sqsubseteq \delta(g_{new}))$, and there is no JEP for other classes, describing the object $(\nexists h \in H_{k-} : h \sqsubseteq \delta(g_{new}))$. This method will be referred as Cl-method.

For example, object g_6 should be attributed to the class k_1 because there exists a JEP for the class k_1, namely $\{\mathtt{m}_6, \mathtt{m}_7\}$, and no JEP for any other class.

In contrast, it is not possible to classify the object with hypotheses, because the corresponding hypothesis would be $\{\mathtt{m}_1, \mathtt{m}_2, \mathtt{m}_6, \mathtt{m}_7\}$ which does not describe the object g_6.

The classical approach usually works well but there are a lot of objects that may not be classified [9]. Another problem is related to real-world data and interpretation of the classification: one may expect to have only one JEP attributing an object to a class. For instance, in the task of predicting toxicity of a molecule, every JEP is a set of substructures and so ideally it should be the set of those substructures which raises the toxicity of the molecule, while in practice there are a lot of JEPs describing every object and so some of them have no relation to the toxicity-specific set of substructures.

For going in this direction, one could recall a principle, widely used in natural science: among all explanation of phenomena one should select the simplest one. So a set of JEPs in our case should classify as many objects from training set as possible, whereas it should not be too complicated. The whole number of JEPs is rather arbitrary, and so it cannot be a measure of complexity. On the other hand if an object should be attributed to a class by only one JEP, then it is natural to suggest that "important JEPs" a) covers all objects and b) that these JEPs are rather general. So the complexity of a system of JEPs could be measured by the minimal number of JEPs required to describe all the objects attributed to any class.

3.1 Running Example

On Table 3a formal context is shown: real life objects, described by some properties, like color and weight. The objects are labeled whether they are natural or human-made. The given labeling is shown on Table 3b. The task is to predict labels of `Cat` and `Elephant`. Tables 3e-3d are other labeling functions used during classification procedure.

	alive	can move	metal	light	green
Tree	x				x
Fungus	x			x	
Velo		x	x	x	x
Car		x	x		x
Cat	x	x		x	
Elephant	x	x			

(a)

Object	Made by
Tree	Nature
Fungus	Nature
Velo	Human
Car	Human
Cat	'?'
Elephant	'?'

(b)

Obj	M
T	N
F	N
V	H
Car	H
Cat	**N**
El	'?'

(c)

Obj	M
T	N
F	N
V	H
Car	H
Cat	**H**
El	'?'

(d)

Obj	M
T	N
F	N
V	H
Car	H
Cat	'?'
El	**N**

(e)

Obj	M
T	N
F	N
V	H
Car	H
Cat	'?'
El	**H**

(f)

Table 3: Running Example Formal Context. Figures 3b-3f are different correspondences between objects and their classes (ξ-functions).

The JEPs for the context on Table 3a and labeling function on Table 3b are the following: $\mathtt{a}$(alive) → `N`, $\mathtt{cm}$(can move) → `H`, $\mathtt{m}$(metal) → `H`, $\mathtt{l}$(light), $\mathtt{g}$(green) → `H`. Neither `Cat` nor `Elephant` may be classified, as they both include JEPs, corresponding to different labels ($\mathtt{a} \to \mathtt{N}$ and $\mathtt{cm} \to \mathtt{H}$). But maybe we are still able to classify them? Let us assume that `Cat` (or `Elephant`) is made by `Nature` (Tables

3c, 3e) and then that they are made by Human (Tables 3d, 3f). And then as a response to the classification task we give the class of the best assumption.

Let us assume that the Cat is made by Nature, the labeling function is shown on Table 3c. The corresponding set of JEPs is as following: a → N; m → H; l,g → H; cm,g → H. We should notice that the label (or class) of every object from Table 3a can be explained by at least one JEP, i.e. for an object g there is a JEP describing object g and corresponding to the class of object g. Let now assume that object Cat is made by Human, the labeling function is shown on Table 3d. The corresponding set of JEPs is as following: a,g → N; cm → H; m → H; l,g → H. Among these JEPs, there is no JEP explaining the class of object Fungus, and so we can say that the assumption that Cat is made by Nature is better than the assumption that Cat is made by Human, and so the Cat should be classified to class Nature.

For the Elephant let us assume first that it is made by Nature, the labeling function on Table 3e. The set of JEPs are a → N; m → H; l,g → H; cm,l → H; cm,l → H. They explain classes of every object from the context. Let us assume that the Elephant is made by Human. The set of JEPs are a,g → N; a,l → N; cm → H; m → H; l,g → H. They do also explain all the objects from the context but we are still able to make a good prediction. For that we should calculate the minimal number of JEPs required to explain every object from the set. For the assumption that Elephant is made by Nature, one requires 2 JEPs to explain every object from the context (a → N; m → H). For the assumption that Elephant is made by Human, one requires 3 JEPs (a,g → N; a,l → N; cm → H). Thus we could say that although both assumptions are possible, the first one is more simple (require only 2 JEPs for explaining every object from the context) and the Elephant should be classified to class Nature.

3.2 The New Approach

We have a pattern context $(G, (D, \sqcap), \delta)$ and a set of classes K. Every object in G can either have a class from K or no class, denoted as '?'. A labeling function $\xi : G \to K \cup \{?\}$ attributes an object g to a class k. Given a context $(G, (D, \sqcap), \delta)$, a set of classes K and a labeling function ξ, one can derive a set of JEPs named H. A system of JEPs refers to a set of all JEPs, derived from a certain context, a certain set of classes, and a certain ξ function.

Definition 10. *A coverage of a system of JEPs H is the set of objects, attributed to some class and described by at least one JEP from H,*
$Coverage(H) = \{g \in G | \xi(g) \neq '?' \text{ and } \exists h \in H, h \sqsubseteq g^{\diamond}\}$.

Definition 11. *A covering set of JEPs denoted by H^* for a given system of JEPs H is such that:*

- *$H^* \subseteq H$;*
- *all objects in $Coverage(H)$ are described by at least one JEP from H^*,* $\forall g \in Coverage(H) : \exists h^* \in H^* : h^* \sqsubseteq g^{\diamond}$

Definition 12. *For a given system of JEPs H, a size of a minimal covering set of JEPs $MinCover(H)$ is the size of a covering set (for the system) with the minimal number of JEPs among all others covering sets for that system.*

Our approach is based on the above definitions. The definitions consider a JEP only w.r.t. a set of objects described by this JEP. And so any JEP among JEPs describing the same set of objects can be considered, without changing the outcome. It is more efficient to mine only closed patterns. Given a context $(G, (D, \sqcap), \delta)$, one can find a set of concepts and then derive a set of hypotheses H for a given set of classes and a given ξ function. Recall that a hypothesis $d \in D$ is associated to a concept (A, d) and every object in A is labeled by the same class or by '?'. Actually a concept (A, d) will not yield a hypothesis when A includes two objects g_1 and g_2 such that $\xi(g_1) \neq' ?'$, $\xi(g_2) \neq' ?'$ and $\xi(g_1) \neq \xi(g_2)$.

Now we can explain our classification approach. For every unclassified object $g \in G$ the method proceeds as follows:

1. For every class $k_i \in K$, one should change the ξ-function to return class k_i for the object g (instead of '?'), $\xi(g) := k_i$. It leads to changing a system of JEPs to H_i. (For instance, in section 3.1 we assume that `Cat` and `Elephant` are either made by `Nature` or by `Human`).
2. For every system of JEPs H_i one should calculate its coverage $(Coverage(H_i))$. If the assumption $\xi(g) := k_i$ is right, all the objects from $Coverage(H)$ and the object g should be covered by H_i. H_i is called complete if $Coverage(H_i) = Coverage(H) \cup \{g\}$ (In Section 3.1, only the system corresponding to the assumption that `Cat` is made by `Human` was incomplete).
   If there is only one complete system then the corresponding class is considered as a result class (as it was made for `Cat` in Section 3.1).
3. For every complete system H_i one should calculate the size of a minimal covering set of JEPs $(MinCover(H_i))$.
4. The only system minimizing the size of minimal covering set corresponds to the predicting label of the object (In Section 3.1, the assumption that `Elephant` is made by `Nature` brings to 2 JEPs in minimal covering set, and corresponds to the predicted `Elephant` class, i.e. `Nature`). If there are more then one minimizing system then the object is unclassifiable.

The full method will be referred as M1 and the method of only first 2 steps will be referred as M2. In Section 3.1 `Cat` can be classified with M1- and M2-method, contrary the `Elephant` can be classified with only M1-method.

The task of finding minimal cover is NP-complete [10]. It can be shown that difference between minimal covering sets sizes $(|MinCover(H) - MinCover(H_i)|)$ of these two systems is often equal to 1. So an approximate solution for the minimal cover set problem can significantly the classification quality.

4 Computer Experiments

Section presents computer experiment and the results.

A database 'Prediction Toxicity Challenge 2000-2001'[3] was used for the experimentation. It consists of molecules labeled by the chemical toxicity with respect to rats and mice of different sexes. Although there are some intermediate labels beside positive and negative. Only positive and negative labels were considered. In Table 4 the sizes of training and test sets are shown.

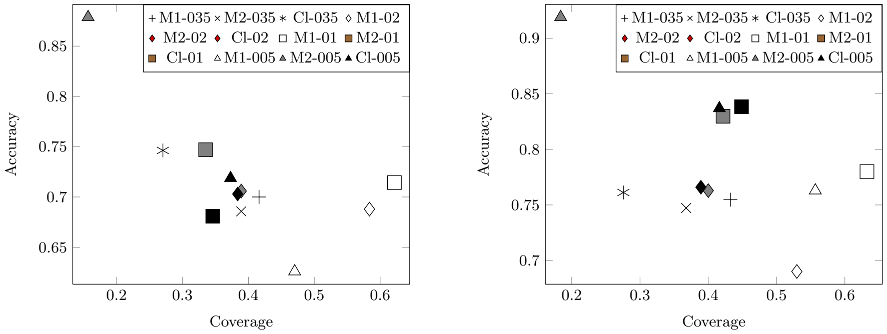
[3] http://www.predictive-toxicology.org/ptc/

	Male Rats	Female Rats	Male Mice	Female Mice
Positives Examples	69	63	68	79
Negatives Examples	192	229	207	206
Test set Positives Examples	84	63	55	66
Test set Negatives Examples	198	219	227	216

Table 4: Numbers of positives and negatives examples in the databases.

One of the way to describe a molecule for applying FCA is to consider it as a graph, where vertices are atoms and edges are bonds between atoms. Then every molecule can be considered as the set of frequent subgraphs, included into the molecule graph. Frequent subgraph means that it is at least present in a certain number of molecules graphs. After converting a set of molecules into graphs, one could use different frequent graph miners [11, 12] to find all frequent subgraphs. Further a frequency limit will be given as percent of the whole molecule set.

To realize M1-classifier one needs a solver for the minimal cover set problem. A greedy algorithm was used to solve the problem approximately. On every iteration the algorithm selects the set, covering the maximal number of uncovered elements. Algorithm stops when all elements are covered by the selected sets.



(a) Male Rats

(b) Female Mice

Fig. 1: The Classification Results.

The results for different frequency limit on the database of male rats are shown on Figure 1a and results for female mice are shown on Figure 1b, results for females rats and males mice databases are not shown for the sake of space. Every point on the plots corresponds to the accuracy and coverage of some classifier, while the molecule is considered as a set of frequent substructure. The classifier and the frequency limit are written in the legend.

The quality of M2-classifier is usually higher than the quality of Cl-classifier, whereas coverage of M2-classifier is decreasing with decreasing of frequency limit (increasing the length of description). M2-classifier refers only to the coverage of a system of hypotheses, thus the coverage is an important measure for the classification. The coverage of M1-classifier is much higher then coverage of classical classifier, but the accuracy is worse then for the classical approach, especially in the case of low frequency limit (long description). This could mean that either M1-classifier is over-learned (it became too specific to training set) or it is important for the algorithm to use an exact solution for minimal cover set problem. As it was mentioned in the step 3 of our approach we need to solve a minimum cover set problem, but for the sake of efficiency the greedy algorithm was used instead of the exact solution. With decreasing of frequency limit the size of minimal cover is increasing, and so an absolute error in defining the size of the minimal cover is increasing as well.

5 Conclusion

In the paper a new approach to classification was suggested. The quality of this approach was checked and it was shown that the number of objects covered by a system of hypothesis is an important characteristic for classification task.

Although the new approach classifies more objects than the classical approach, in some situations it has worse classification quality. One of the possible reasons is an approximate solution for the minimal cover problem. The influence of the approximate minimal cover problem solution should be checked.

References

1. Dong, G., Li, J.: Efficient mining of emerging patterns: discovering trends and differences. In: Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '99, New York, ACM (1999) 43–52
2. Poezevara, G., Cuissart, B., Crémilleux, B.: Extracting and summarizing the frequent emerging graph patterns from a dataset of graphs. Journal of Intelligent Information Systems **37** (July 2011) 333–353
3. Ganter, B., Kuznetsov, S.: Formalizing hypotheses with concepts. In Ganter, B., Mineau, G., eds.: Conceptual Structures: Logical, Linguistic, and Computational Issues. Volume 1867 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2000) 342–356 10.1007/10722280_24.
4. Blinova, V.G., Dobrynin, D.A., Finn, V.K., Kuznetsov, S.O., Pankratova, E.S.: Toxicology analysis by means of the JSM-method. Bioinformatics **19**(10) (2003) 1201–1207
5. Kuznetsov, S.O., Samokhin, M.V.: Learning closed sets of labeled graphs for chemical applications. In: ILP. (2005) 190–208
6. Ganter, B., Kuznetsov, S.O.: Pattern structures and their projections. In: ICCS. (2001) 129–142
7. Kaytoue, M., Duplessis, S., Kuznetsov, S., Napoli, A.: Two FCA-Based methods for mining gene expression data. In Ferré, S., Rudolph, S., eds.: Formal Concept Analysis. Volume 5548 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2009) 251–266 10.1007/978-3-642-01815-2_19.
8. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. 1st edn. Springer-Verlag New York, Inc., Secaucus, NJ, USA (1997)
9. Helma, C., King, R.D., Kramer, S., Srinivasan, A.: The predictive toxicology challenge 2000–2001. Bioinformatics **17**(1) (2001) 107–108
10. Cormen, T.H.: Introduction to algorithms. MIT Press, Cambridge (2009)
11. Yan, X., Han, J.: gSpan: graph-based substructure pattern mining. In: Proceedings of IEEE International Conference on Data Mining, 2002. (2002) 721 – 724
12. Nijssen, S., Kok, J.: The gaston tool for frequent subgraph mining. Electronic Notes in Theoretical Computer Science **127** (March 2005) 77–87