Discovering Knowledge from Local Patterns with Global Constraints

Bruno Crémilleux and Arnaud Soulet

GREYC, Université de Caen
Campus Côte de Nacre
F-14032 Caen Cédex France
{Forename.Surname}@info.unicaen.fr

Abstract. It is well known that local patterns are at the core of a lot of knowledge which may be discovered from data. Nevertheless, use of local patterns is limited by their huge number and computational costs. Several approaches (e.g., condensed representations, pattern set discovery) aim at grouping or synthesizing local patterns to provide a global view of the data. A global pattern is a pattern which is a set or a synthesis of local patterns coming from the data. In this paper, we propose the idea of global constraints to write queries addressing global patterns. A key point is the ability to bias the designing of global patterns according to the expectation of the user. For instance, a global pattern can be oriented towards the search of exceptions or a clustering. It requires to write queries taking into account such biases. Open issues are to design a generic framework to express powerful global constraints and solvers to mine them. We think that global constraints are a promising way to discover relevant global patterns.

Keywords: local patterns, constraint-based paradigm, global constraints

1 Introduction

The constraint-based pattern mining framework is a powerful paradigm to discover relevant information from databases [13]. Constraints provide a focus on the most promising knowledge by reducing the number of extracted patterns to those of a potential interest given by the user. There are now generic approaches to discover *local patterns* under constraints [5, 18] and this issue is rather well-mastered, at least for the itemset data. Nevertheless, even if the number of produced patterns is reduced thanks to the constraint, the output still remains too large for an individual and global analysis performed by the end-user. The most significant patterns are lost among too many trivial, noisy and redundant information. Many works propose methods to reduce the collection of patterns, like the condensed representations [3] or the compression of the dataset by exploiting Minimum Description Length Principle [16] but most of them are dedicated to frequent patterns. Two recent approaches - constraint-based pattern set mining [6] and pattern teams [10] - aim at reducing the redundancy by selecting

Dagstuhl Seminar Proceedings 07181
Parallel Universes and Local Patterns
http://drops.dagstuhl.de/opus/volltexte/2007/1259

patterns from the initial large set of patterns on the basis of their usefulness in the context of the other selected patterns. But, these approaches cannot take into account in a flexible way a bias given by the user to orient the final set of patterns towards a specific aim like for instance the search of exceptions or a clustering.

On the other hand, we think that it should be a pity to consider the synthesis of local patterns only from the point of view of the redundancy. We claim that local patterns can be fruitfully gathered to produce *global patterns* asked by the end-user. Interestingly, such global patterns can capture the joint effect of local patterns. For instance, in the context of co-classification, Pensa et al. show that the bi-clusters of the final bi-partition (i.e., the global pattern) are not necessary elements of the initial set of the local patterns. The bi-partition comes from a reconstruction of the objects and attributes of the local patterns. Furthermore, we think that local patterns are also suitable to revisit classical data exploration tasks like for example classification or clustering. For instance, let us assume that the user would like to produce a classifier based on rules. A complete and correct constraint-based data mining method ensures to produce the proper set of classification rules. It differs from the decision trees technique where the attribute selection criterion ensures to pick the best attribute at each node but it does not guarantee the best whole tree [2]. Obviously, having a proper set of classification rules does not straightforwardly provide a relevant classifier. More generally, it is clear that moving from local patterns to global patterns or models like classifiers is still a challenge. Studying carefully the complementarity of the various pattern domains (e.g., clustering and local pattern discovery, local pattern discovery as feature construction for supervised classification) is needed. In this paper, we address this general issue by using the constraint paradigm and defining global constraints. We show that a lot of useful global patterns can be achieved in a flexible and declarative way thanks to the notion of global constraints. This approach enables the user to express a bias and discover relevant global patterns.

This paper is organized as follows. Section 2 is devoted to related work and provides some examples of global patterns coming from local patterns. We will see that the current techniques are specific to the kind of targeted global patterns. In Section 3, we propose the notion of global constraints and we show how it enables to address a lot of global constraints. In Section 4, we discuss current research which is related to this framework. We give an example of discovery of local patterns in parallel universes. A further natural extension of the model is to mine global constraints in several universes in order to get more powerful global patterns and models.

2 Context and related work

In this section, we sketch some methods providing global patterns coming from local patterns. We start by methods producing sets of all patterns satisfying a property involving several local patterns. Then we give other examples of methods building global syntheses from all the local patterns, these syntheses are oriented towards a bias given by the user like the designing of a classifier or a clustering.

2.1 Global patterns as sets of local patterns

Finding characterization rules is an important issue in a lot of applications. The usefulness of the simplest rules (i.e., rules having minimal premises) for this task is shown in [4]. Basically, given a bounded number of exceptions δ, a rule has a minimal premise if any rule with a proper subset of its premise and less than δ exceptions does not enable to conclude on the same class value (Section 3 gives a formal definition highlighting the comparison between two local patterns). Such rules are inferred from a specific subset of the δ-free patterns [1]. Prototypes extracting the condensed representations of the δ-free sets can be easily updated to extract the whole collection of these patterns [4]. Unfortunately, this approach cannot extended if we are interested in other kinds of characterization rules.

The discovering of exception rules from a data set without domain-specific information is also of a great interest [20]. An exception rule is defined as a deviational pattern to a strong rule and the interest of a rule is evaluated according to another rule. Again, the comparison between rules means that these patterns are not local patterns. Suzuki proposes a method based on sound pruning and probabilistic estimation [20] to extract all exception rules. But, as previously, this approach is devoted to this kind of patterns.

Another example is the top-k patterns. These patterns are the k patterns maximizing an interestingness measure m (e.g., the frequency [8]). The collection of the top-k patterns is concise and gathers the most relevant patterns according to m. It can be seen as a global pattern. Nevertheless, patterns belonging to the a collection of top-k patterns are sometimes trivial and often redundant.

2.2 Global patterns as models

Other examples start from the whole collection of local patterns to provide a global pattern or a model. A typical example is the associative classification. This technique proposes the integration of association rule mining and classification [12, 11]. As the number of potential classification rules is very large, pruning techniques are used. Rules which are redundant from a functional point of view or may cause incorrect classification are deleted. Pruning is usually applied as a post-processing step on the extracted rules by using statistical parameters such as support, confidence and chi-square test. A more recent approach finds the best k correlated association rules for classification by using a measure which has suitable pruning properties [22]. A common characteristic of all these methods is to use heuristics techniques to select the rules from a complete collection of local patterns to produce a classifier.

Associations, like the frequent patterns, are also at the core of clustering works [21]. For instance, Ecclat [7] is based on frequent closed patterns and has the originality to enable a slight overlapping between clusters. A global pattern produced by ECCLAT is performed by a greedy method in which the interest of a cluster is evaluated according to an interestingness measure.

Finally, let us come back on co-classification, already mentioned in Section 1. This approach is a way of conceptual clustering and provides a limited collection of bi-clusters. These bi-clusters are linked for both objects and attribute-value pairs. In [15], the authors propose a generic framework for co-classification. Its great interest is that the bi-partition comes from a reconstruction of the objects and attributes of the local patterns. Nevertheless, a distance between the bi-sets which are at the origin of the bi-clusters has to be chosen.

These examples show that combinations of local patterns are ad hoc to a specific goal and often use heuristics or parameters which have to be set by the user. The next section proposes to notion of global constraints and their use to design a generic framework for producing global patterns and models.

3 Designing global constraints for mining global patterns

This section proposes the notion of *global constraint* which formalizes the building of the global patterns previously introduced. Global constraints consider simultaneously at least two patterns contrary to usual constraints, named *local*, which are checked individually on each pattern.

Definition 1 (Global constraint). *A constraint q is said global if several patterns have to be compared to check if q is satisfied or not.*

Checking a usual local constraint requires individual information about a pattern like its frequency, its cardinality, etc. On the contrary, a global constraint satisfaction is based on information coming from at least two patterns. This point is highlighted in the examples given below by the two patterns X and Y. For instance, several constraints (e.g., freeness, characterization rules, exception rules) require to check properties of the subsets Y of a X. Peak constraint compares two neighbour patterns and top-k any pair of patterns.

We present now a non-exhaustive list of global constraints for illustrating Definition 1 in the context of itemset mining. All the examples given in Section 2.1 are written with the formalism of the global constraints. Some of them are very well-known, like freeness or closedness. We also propose the peak constraint which highlights an exceptional behaviour of a pattern with respect to its neighbors. The set of items is denoted by $\mathcal{I}$ and the language of itemsets corresponds to $\mathcal{L}_{\mathcal{I}} = 2^{\mathcal{I}}$.

- **Freeness and closedness:** A free (resp. closed) pattern is a minimal (resp. the maximal) pattern of its equivalence class of frequency [1, 14]. This kind of patterns is useful to design condensed representations or to derive association rules. The frequency of X, denoted by $freq(X)$, corresponds to the number of transactions containing X. These properties require to check the frequency of X with respect to the frequencies of its subsets.

$$freeness(X) \equiv \begin{cases} true & \forall Y \in \mathcal{L_I} \text{ such that } Y \subset X \text{, one have } freq(X) < freq(Y) \\ false \text{ otherwise} \end{cases}$$

$$closedness(X) \equiv \begin{cases} true & \forall Y \in \mathcal{L_I} \text{ such that } X \subset Y \text{, one have } freq(X) > freq(Y) \\ false \text{ otherwise} \end{cases}$$

- **Characterization rules:** A formulation of characterization rules is given in [4]. It can be defined by the following global constraint:

  $$characterization(X \to c) \equiv \begin{cases} true & \forall Y \in \mathcal{L_I} \text{ such that } X \to c, Y \subset X, \neg(Y \to c) \\ false \text{ otherwise} \end{cases}$$

  where $X \to c$ denotes a rule with $freq(X \to c) \geq \gamma$ and $freq(X \to \neg c) \leq \delta$. It means that $X \to c$ is a rule with minimal premise to conclude to c with lower δ exceptions.
- **Exceptions:** Suzuki [20] defines exception rules which isolate unexpected information as the following set or rules:

  $$exception(XY \to \neg A) \equiv X \to A \wedge XY \to \neg A \wedge Y \not\to \neg A$$

  The rule $XY \to \neg A$ is an exception rule since usually if X then A and if Y then not frequently $\neg A$.
- **Top-k constraint:** Let $k > 0$ be an integer and $m : \mathcal{L_I} \to \Re$ be a measure (e.g., frequency [8]). The top-k constraint w.r.t m are the k best patterns according to m and it equals to:

  $$top_{k,m}(X) \equiv |\{Y \in \mathcal{L_I} | Y \neq X \wedge m(Y) > m(X)\}| < k$$

- **Peak constraint:** A peak pattern is a pattern whose surrounders have a value for a measure m lower than a threshold. It means that a peak pattern has an exceptional behaviour compared to its neighbours. Let $m : \mathcal{L_I} \to \Re$ be a measure, d be a distance (e.g., $d(X,Y) = |X \backslash Y| + |Y \backslash X|$), δ be an integer and ρ be a real, one have:

  $$peak(X) \equiv \forall Y \in \mathcal{L_I} \text{ such that } d(X,Y) < \delta \text{, one have } m(X) \geq \rho \times m(Y)$$

All examples of global patterns given in Section 2.1 can be straightforwardly expressed with global constraints. We think that this unified approach brought by the global constraints improves the understandability of the global patterns. Furthermore, this approach enables to easily design new global patterns like peaks, but many other examples can be addressed.

Global constraints provide a reduced collection of patterns forming a coverage or gathering the best ones according to a criterion. The *covering constraints* synthetize the whole set of local patterns by deleting redudancies: freeness, closedness, characterization rules. The *optimizing constraints* maximize a given criterion over some patterns (i.e., exceptions, peak constraint) or over all the patterns (i.e., top-k constraint). The border between these two categories is fuzzy: for instance, a characterization rule also satisfies a criterion, i.e., a maximal number of exceptions.

4 Research open problems

In this section, we discuss current research open problems related to the design and the use of global constraints for discovering global patterns and models.

Mining patterns under global constraints This task is harder than the usual constraint-based pattern mining. A naive post-processing method is slow or intractable due to the very high number of comparisons between patterns required by the global constraints. Specific global constraints like closedness [14] or freeness [1] have dedicated algorithms mainly relying on anti-monotone pruning. A branch-and-bound algorithm for top-k patterns is proposed in [6], but only few measures are addressed. To the best of our knowledge, there is no general method of pattern mining under global constraints. Recently, the new *Approximate-and-Push* approach [17] enables to mine different global constraints like freeness, closedness or top-k constraint by dynamically relaxing global constraints into local ones. Finally, approaches coming from the field of Constraint Satisfaction Problem (CSP) would be probably useful for the pattern mining under global constraints [9]. CSP has the ability to define constraints on several variables. These constraints are able to synthesize sets of elementary constraints and they provide more powerful pruning conditions for handling global constraints.

Global constraints for global models In Section 2.2, we saw that a whole collection of local patterns is used to infer global patterns and models. We think that global constraints enable to go further in this direction. Indeed, by selecting patterns revealing a global structure emerging from the database, global constraints can provide richer patterns than the local patterns currently used. For this purpose, local patterns can be compared and then restrain to a smaller collection gathering the best ones or forming a coverage (for instance, discovering the k best classification rules according to interestingness and coverage criteria). Open avenue is to design high-level constraints to directly build global models, it would avoid the choice of heuristics, as it is the case in the current methods (see Section 2.2). We think that it is a promising and powerful approach.

Global constraints in parallel universes Parallel universes is a way to build global models from a set of connected universes, each object may have a different (and partial) representation in each universe. Local patterns can be of multiple types. We briefly give now an example of discovering local patterns from several data resources which can be seen as distinct universes. A constraint-based approach links the information scattered in the different sources. A constraint selects the most relevant patterns by evaluating their interestingness in relation with the different sources. Previous work [19], shows that the constraints based on a set of primitives enable to link the information disseminated in various knowledge sources. For instance, in the domain of genomics, biologists are interested in constraints both on synexpression groups and common characteristics of the genes and/or biological situations involved in these groups. Figure 1 illustrates this constraint-based task with a typical query. Each part of the constraint q

refers to a different universe indicated on the right: synexpression groups/boolean matrix, description of genes/textual data and similarity matrices.

Boolean matrix $\mathcal{D}$

Situations	Genes
s_1	A E F
s_2	B C D
s_3	A B C D E F
s_4	A B C D

Similarity matrix

	A	B	C	D	E	F
A		.07	?	?	.2	0
B			.06	?	?	0
C				.07	.05	.04
D					.03	.1
E						?

Textual data

A	'metal ion binding' 'transcription factor'
B	'serine-type peptidase activity' 'proteolysis'
C	'DNA binding' 'metal ion binding'
D	'ATP binding' 'nucleotide binding'
E	'proteolysis'
F	'ATP binding' 'metal ion binding'

freq, length,... *sumsim, svmsim,...* *regexp*

$$
\begin{aligned}
q(X) \equiv\ & freq(X) \times count(X) \geq 24 && \text{Boolean matrix}\\
& \wedge count(regexp(X, '*ribosom*', \texttt{TEXT\_terms})) \leq 1 && \text{Textual data}\\
& \wedge svsim(X, \texttt{TEXT})/(svsim(X, \texttt{TEXT}) + mvsim(X, \texttt{TEXT})) \geq 0.7 && \text{Similarity matrix}\\
& \wedge sumsim(X, \texttt{TEXT})/svsim(X, \texttt{TEXT}) \geq 0.025 && \text{Similarity matrix}
\end{aligned}
$$

Fig. 1. Example of a genomic mining context and a constraint through parallel universes.

A further step is clearly to extend the notion of global constraints to parallel universes. The global constraints would benefit from comparisons between patterns coming from different universes (e.g., synexpression groups and gene descriptions in litterature) in order to reveal stronger correlations. For instance, exception rules through several universes may reveal very unexpected knowledge, e.g., a synexpression group X concluding on `cancer` while connected gene descriptions concluding on `no cancer`. One global constraint formalizing such rules could be (the function Φ connects the different parallel universes by returning the patterns describing X in the other universes):

$$exception(X) \equiv X \rightarrow \texttt{cancer} \wedge (\forall Y \in \Phi(X) \wedge Y \rightarrow \texttt{nocancer})$$

Consequently, such global constraints over several universes need to mine in parallel local patterns in each universe for comparing them. We think that the context of parallel universes emphasizes the interest of the global constraint approach for discovering knowledge from data and local patterns.

References

[1] J.-F. Boulicaut, A. Bykowski, and C. Rigotti. Free-sets: A condensed representation of boolean data for the approximation of frequency queries. *Data Min. Knowl. Discov.*, 7(1):5–22, 2003. Kluwer Academics Publishers.

[2] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and regression trees.* Statistics probability series. Wadsworth, Belmont, 1984.

[3] T. Calders, C. Rigotti, and J.-F. Boulicaut. A survey on condensed representations for frequent sets. In J.-F. Boulicaut, L. De Raedt, and H. Mannila, editors,

Constraint-Based Mining and Inductive Databases, volume 3848 of *Lecture Notes in Computer Science*, pages 64–80. Springer, 2005.

[4] B. Crémilleux and J.-F. Boulicaut. Simplest rules characterizing classes generated by delta-free sets. In *22nd Int. Conf. on Knowledge Based Systems and Applied Artificial Intelligence (ES'02)*, pages 33–46, Cambridge, UK, December 2002.

[5] L. De Raedt, M. Jäger, S. D. Lee, and H. Mannila. A theory of inductive query answering. In *proceedings of the IEEE Conference on Data Mining (ICDM'02)*, pages 123–130, Maebashi, Japan, 2002.

[6] L. De Raedt and A. Zimmermann. Constraint-based pattern set mining. In *Proceedings of the Seventh SIAM International Conference on Data Mining*. SIAM, 2007.

[7] N. Durand and B. Crémilleux. ECCLAT: a New Approach of Clusters Discovery in Categorical Data. In *22nd Int. Conf. on Knowledge Based Systems and Applied Artificial Intelligence (ES'02)*, pages 177–190, Cambridge, UK, December 2002.

[8] A. Fu, R. W., W. Kwong, and J. Tang. Mining n-most interesting itemsets. In *proceedings of the 12th International Symposium ISMIS*, volume 1932 of *Lecture Notes in Computer Science*, pages 59–67, Charlotte, NC, USA, October 2000. Springer.

[9] C. Gervet and P. Van Hentenryck. Length-lex ordering for set csps. In *AAAI*, 2006.

[10] A. Knobbe and E. Ho. Pattern teams. In J. Fürnkranz, T. Scheffer, and M. Spiliopoulou, editors, *proceedings of the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'06)*, volume 4213, pages 577–584, Berlin, Germany, September 2006. Springer-Verlag.

[11] W. Li, J. Han, and J. Pei. CMAR: accurate and efficient classification based on multiple class-association rules. In *proceedings of the IEEE International Conference on Data Mining, ICDM 01*, pages 369–376, San Jose, California, 2001. IEEE Computer Society.

[12] B. Liu, W. Hsu, and Y. Ma. Integrating classification and association rules mining. In *proceedings of Fourth International Conference on Knowledge Discovery & Data Mining (KDD'98)*, pages 80–86, New York, August 1998. AAAI Press.

[13] R. T. Ng, V. S. Lakshmanan, J. Han, and A. Pang. Exploratory mining and pruning optimizations of constrained associations rules. In *proceedings of ACM SIGMOD'98*, pages 13–24. ACM Press, 1998.

[14] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal. Discovering frequent closed itemsets for association rules. *Lecture Notes in Computer Science*, 1540:398–416, 1999.

[15] R. Pensa, C. Robardet, and J.-F. Boulicaut. A bi-clustering framework for categorical data. In A. Jorge, L. Torgo, P. Brazdil, R. Camacho, and J. Gama, editors, *proceedings of the 9th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'05)*, pages 643–650, Porto, Portugal, October 2005.

[16] A. Siebes, J. Vreeken, and M. van Leeuwen. Item sets that compress. In *Proceedings of the Sixth SIAM International Conference on Data Mining, April 20-22, 2006, Bethesda, MD, USA*. SIAM, 2006.

[17] A. Soulet. *Un cadre générique de découverte de motifs sous contraintes fondées sur des primitives.* PhD thesis, Université de Caen Basse-Normandie, France., 2006.

[18] A. Soulet and B. Crémilleux. An efficient framework for mining flexible constraints. In H. T. Bao, D. Cheung, and H. Liu, editors, *9th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'05)*, volume 3518 of

Lecture Notes in Artificial Intelligence, pages 661–671, Hanoi, Vietnam, May 2005. Springer.

[19] A. Soulet, J. Klema, and B. Crémilleux. *Post-proceedings of the 5th International Workshop on Knowledge Discovery in Inductive Databases in conjunction with ECML/PKDD 2006 (KDID'06)*, chapter Efficient Mining under Rich Constraints Derived from Various Datasets. Lecture Notes in Computer Science. Springer, 2007.

[20] E. Suzuki. Undirected discovery of interesting exception rules. *International Journal of Pattern Recognition and Artificial Intelligence*, 16(8):1065–1086, 2002.

[21] K. Wang, X. Chu, and B. Liu. Clustering transactions using large items. In *proceedings of ACM CIKM 99*, 1999.

[22] A. Zimmermann and L. De Raedt. Corclass: correlated association rule mining for classification. In E. Suzuki and S. Arikawa, editors, *proceedings of the 7th International Conference on Discovery Science (DS'04)*, Lecture notes in artificial intelligence, pages 60–72. Springer-Verlag, 2004.