Available online at www.sciencedirect.com
ScienceDirect
Information Sciences 177 (2007) 5675–5690
INFORMATION SCIENCES
AN INTERNATIONAL JOURNAL
www.elsevier.com/locate/ins

# Jumping emerging patterns with negation in transaction databases – Classification and discovery

Pawel Terlecki, Krzysztof Walczak *

*Institute of Computer Science, Faculty of Electronics and Information Technology, Warsaw University of Technology, Nowowiejska 15/19, 00-665 Warsaw, Poland*

Received 17 October 2006; received in revised form 16 July 2007; accepted 17 July 2007

**Abstract**

This paper examines jumping emerging patterns with negation (JEPNs), i.e. JEPs that can contain negated items. We analyze the basic relations between these patterns and classical JEPs in transaction databases and local reducts from the rough set theory. JEPNs provide an interesting type of knowledge and can be successfully used for classification purposes. By analogy to JEP-Classifier, we consider negJEP-Classifier and JEPN-Classifier and compare their accuracy. The results are contrasted with changes in rule set complexity. In connection with the problem of JEPN discovery, JEP-Producer and rough set methods are examined.


*Keywords:* Jumping emerging pattern; Negation; Contradictory database; Extended database; Rough set; Local reduct; Transaction database

## 1. Introduction

Patterns play an important role in knowledge discovery. In the field of transaction data analysis, they have been successfully applied to various practical problems, like association mining, classification or clustering. In contrast to many statistical tools or neural networks, they can be easily understood and interpreted by humans. Moreover, their simple structure has led to the recent development of efficient concise representations that help to manage large pattern collections.

Positive patterns, that is, specific sets of items supported by transactions of a given dataset, are a very intuitive concept. Our paper considers one of its possible extensions, i.e. patterns with negation, which combines positive and negative aspects of data. Originally, negative relationships were introduced in [22], where a chi-square model was applied in order to estimate independence between two variables. As far as data mining is concerned, the vast majority of publications employ the idea of negation to formulate new interesting association rules. Therefore, many algorithms include variants of frequent itemset mining [11]. Since extending the

* Corresponding author. Tel.: +48 22 6607432; fax: +48 22 8251635.
*E-mail addresses:* P.Terlecki@ii.pw.edu.pl (P. Terlecki), K.Walczak@ii.pw.edu.pl (K. Walczak).

0020-0255/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2007.07.018

definition of a pattern results in search space enlargement, special algorithmic strategies should be developed, e.g. more efficient pruning can be achieved by supplementing the support-confidence framework with additional measures of interestingness [28,1]. Another approach is to search for rules of a constrained form that often correspond to a specific interpretation. For example, in negative association rules [28] and confined association rules [1], only a complete antecedent or consequent can be negated, whereas in unexpected rules [16] and exceptional rules [12] negative items in antecedents are used to represent exceptions to the regular associations. Some other solutions make use of domain knowledge to formulate valuable rules [21,29]. Last but not least, the problem of mining frequent itemsets with negation can be addressed by means of concise data representations [4,13].

Our discussion focuses on a specific type of patterns called jumping emerging patterns (JEPs, [6,14]). Compared to many other similar propositions, JEPs seem to be a very accurate tool for classification problems. A JEP is an itemset which is supported in one database and absent from others. Let us consider the example in Table 1. Note that the patterns: *cef*, *de* are JEPs in class 0. Moreover, the latter is minimal, because neither of its proper subsets is a JEP. For the sake of brevity, we omit brackets and commas in the itemset notation, e.g. the pattern *ab* should be understood as $\{a, b\}$.

A JEP is an example of positive information. It indicates that a certain pattern is specific to some class. Negative knowledge can be modeled by negated items, which are treated as items that do not co-exist with regular ones in a given transaction. In terms of the formal data mining apparatus, let us represent the fact that a transaction does not contain an item $a$ by the statement that it contains the respective negated item, denoted by $\bar{a}$. In other words, we express negation in terms of existence. Now, let us consider the patterns: $\bar{b}d\bar{f}$, $\bar{a}e$. The first one is supported by only one transaction $id2$, while the second one by all the three transactions of class 0: $id1$, $id2$, $id3$. We call these patterns jumping emerging patterns with negation (JEPNs). Note that a JEP is a specific type of a JEPN that contains only positive items, a positive JEPN (posJEPN). Similarly, we can introduce a dual concept, a negative JEPN (negJEPN) that is built only out of negated items, e.g. $\bar{a}\bar{d}$.

Both negJEPNs and JEPNs are similar to JEPs in their nature, as we demonstrate that these new types of patterns are still JEPs in specifically modified databases. This observation transfers many important features of JEPs to negJEPNs/JEPNs and allows us to use the same pattern discovery methods. Unfortunately, this task is often much harder than finding JEPs in the original database. In particular, JEPN mining requires analysis of a database with a doubled number of possible items and longer transactions. Thus, this problem has more dimensions, is more time consuming and, for many datasets, this kind of analysis proves unfeasible. As an alternative mining solution, we propose a method based on the notion of a local reduct in a decision table [3]. This concept derives from the rough set theory, which provides one of the most elegant descriptions of relational data. Basic ideas, proposed by Zdzislaw Pawlak in 1982 [17,18], triggered intensive research, which has resulted in the creation of many convenient tools for real data analysis, e.g reduct-based classifiers or discretizers [3,23,20,19]. It is worth mentioning that several relations between the rough set theory and JEPs have already been observed [26,25]. Here, we demonstrate a correspondence between JEPNs and local reducts in a binary relation database constructed for a transaction database. Taking advantage of this correspondence, we compare the efficiency of JEPN discovery by means of JEP and local reduct finding methods.

In general, in various works that consider the idea of negation in transaction databases, classification issues are rarely raised, and tackled by means of associations [15,1]. Therefore, we believe that the role of negation in the field of emerging patterns is worthy of a comprehensive study. Also, the experimental results presented in

Table 1
Sample transaction database

| TID | Itemset | Class |
|---|---|---|
| $id1$ | $\{cef\}$ | 0 |
| $id2$ | $\{de\}$ | 0 |
| $id3$ | $\{e\}$ | 0 |
| $id4$ | $\{bcd\}$ | 1 |
| $id5$ | $\{df\}$ | 1 |
| $id6$ | $\{ace\}$ | 1 |