APPLICATION OF ASSOCIATION RULES FOR EFFICIENT LEARNING WORK-FLOW

Sylvia Encheva
Stord/Haugesund University College
Bjornsonsg. 45, 5528 Haugesund, Norway
sbe@hsh.no

Sharil Tumin
University of Bergen
IT-Dept., P. O. Box 7800, 5020 Bergen, Norway
edpst@it.uib.no

Abstract This paper describes application of association rules in education. To make everything more clearly visible a graphic display of objects and attributes in a lattice structure is provided. Vectors and matrices are used to reduce the computational complexity while searching for association rules in the case when at least one attribute included in the 'if' part of the statement is known.

Keywords: E-learning, data mining

Introduction

Association-rule mining is a technique for finding association and/or correlation relationships among data items in large databases. Association rules are probabilistic in nature and show attribute value conditions that occur frequently together in a given dataset. The information association rules provide is a statement in an antecedent/consequent format. The probabilistic approach deals with statements of the form 'the presence of attributes $\aleph$ and $\Re$ often also involves attribute $\Im$'. This approach has an application in different fields such as market basket analysis [5], medical research [9] and census data [12].

In this paper we show an application of association rules in education. A graphic display of objects and attributes in a lattice structure is also provided.

The rest of the paper is organized as follows. Related work is described in Section 1. Some definitions and statements from formal concept analysis and rule mining may be found in Section 1. The main results of the paper are placed in Section 2. The paper ends with a conclusion in Section 3.

1. Related Work

Formal concept analysis [10], [16] started as an attempt of promoting better communication between lattice theorists and users of lattice theory. Since 1980's formal concept analysis has been growing as a research field with a broad spectrum of applications. Various applications of formal concept analysis are presented in [11]. An excellent introduction to ordered sets and lattices and to their contemporary applications can be found in [8].

The complexity of mining frequent itemsets is exponential and algorithms for finding such sets have been developed by many authors such as [4], [6], [15] and [18].

Mining association rules is addressed in [1]. Algorithms for fast discovery of association rules have been presented in [2], [3], and [14].

Preliminaries

A *concept* is considered by its *extent* and its *intent*: the *extent* consists of all objects belonging to the concept while the *intent* is the collection of all attributes shared by the objects [8].

A *context* is a triple (G, M, I) where G and M are sets and $I \subset G \times M$. The elements of G and M are called *objects* and *attributes* respectively.

For $A \subseteq G$ and $B \subseteq M$, define $A' = \{m \in M \mid (\forall g \in A) \ \ gIm\}$ and $B' = \{g \in G \mid (\forall m \in B) \ \ gIm\}$; so A' is the set of attributes common to all the objects in A and B' is the set of objects possessing the attributes in B. Then a *concept* of the context (G, M, I) is defined to be a pair (A, B) where $A \subseteq G$, $B \subseteq M$, $A' = B$ and $B' = A$. The *extent* of the concept (A, B) is A while its intent is B. A subset A of G is the extent of some concept if and only if $A'' = A$ in which case the unique concept of the which A is an extent is (A, A'). The corresponding statement applies to those subsets B of M which are the intent of some concept.

The set of all concepts of the context (G, M, I) is denoted by $\mathfrak{B}(G, M, I)$. $\langle \mathfrak{B}(G, M, I); \leq \rangle$ is a complete lattice and it is known as the *concept lattice* of the context (G, M, I).

An association rule $Q \rightarrow R$ holds if there are sufficient objects possesing both Q and R and if there are sufficient objects among those with Q which also possess R [7].

A context (G, M, I) satisfies the association rule $Q \rightarrow R_{minsup,minconf}$, with $Q, R \in M$, if $sup(Q \rightarrow R) = \frac{|(Q \cup R)'|}{|G|} \geq minsup$, and $conf(Q \rightarrow R) = \frac{|(Q \cup R)'|}{|Q'|} \geq minconf$ provided *minsup* $\in [0, 1]$ and *minconf* $\in [0, 1]$.

The ratios $\frac{|(Q \cup R)'|}{|G|}$ and $\frac{|(Q \cup R)'|}{|Q'|}$ are called, respectively, the *support* and the *confidence* of the rule $Q \rightarrow R$. In other words the rule $Q \rightarrow R$ has support $\sigma\%$ in the transaction set $\mathcal{T}$ if $\sigma\%$ of the transactions in $\mathcal{T}$ contain $Q \cup R$. The rule

Table 1. Context for students groups

	Preliminary knowledge			Chapter 1		Chapter 2		Chapter 3	
	sufficient	some	none	can	cannot	can	cannot	can	cannot
Gr. 1	×			×		×		×	
Gr. 2	×			×			×	×	
Gr. 3		×			×	×		×	
Gr. 4			×		×	×			×
Gr. 5	×			×		×		×	
Gr. 6	×			×			×	×	
Gr. 7		×			×	×		×	
Gr. 8			×		×	×			×
Gr. 9	×				×	×			×

has confidence $\psi\%$ if $\psi\%$ of the transactions in $\mathcal{T}$ that contain Q also contain R.

2. Application of Association Rules

Students taking a course are divided in groups according to gender and results from a test. The goal is to find the association rules that relate attributes, chosen by a lecturer, to students' results from the test.

Group 1 (Gr. 1) - male students with score above 80% on the test
Group 2 (Gr. 2) - male students with score between 60% and 80% on the test
Group 3 (Gr. 3) - male students with score between 40% and 59% on the test
Group 4 (Gr. 4) - male students with score between 20% and 39% on the test
Group 5 (Gr. 5) - female students with score above 80% on the test
Group 6 (Gr. 6) - female students with score between 60% and 80% on the test
Group 7 (Gr. 7) - female students with score between 40% and 59% on the test
Group 8 (Gr. 8) - female students with score between 20% and 39% on the test
Group 9 (Gr. 9) - students with score less than 20% on the test.

The corresponding Hasse diagram is shown in Fig. 1.

Based on this context a binary matrix G with rows and columns corresponding to the rows and columns in Table 1 is generated. The rows in the context are denoted by $g_i, i = 1, ..., 9$. An element in the matrix is set equal to 1 if the corresponding entry in the context is marked and to 0 otherwise.

Suppose we are interested in the association rules that involve the attribute 'has sufficient preliminary knowledge'. We then choose all vectors that have 1 as their first coordinate. The obtained matrix is denoted by $G_{i,j}$ (in this

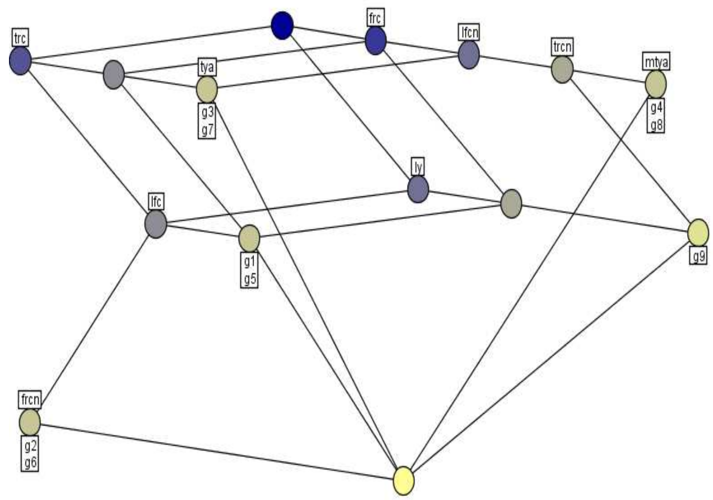

Figure 1. Hasse diagram for students' results from an implicit differentiation test

particular case $G_{6,1}$), where the value of i shows the number of rows in $G_{i,j}$ and the value of j shows which attribute has been chosen.

This way we considerably reduce the number of any other following operations that involve the attribute 'has sufficient preliminary knowledge' in the 'if' part of our search. We then add g_1 to each of the remaining vectors in $G_{6,1}$ and thus obtain $G'_{6,1}$.

Advantage: At this point we can again reduce the number of rows in $G'_{6,1}$ by deleting all rows with 0's and 2's only. They are a repetition of the first vector in $G_{6,1}$ and only their number is of importance for our further work. Such a row reduction applied in our example leads to a matrix $G'_{5,1}$ with five rows only.

The next step is to look for positions in $G'_{6,1}$ (or in its redused version if applicable, in this case $G'_{5,1}$) with value 2, beginning with the first row. Those positions indicate association rules.

The association rules that have the attribute 'has sufficient preliminary knowledge' as an antecedent are

- If a student has sufficient preliminary knowledge then he/she can work with the material in Chapter 1 and Chapter 3 with a probability 79%.

- If a student has sufficient preliminary knowledge and can work with the material in Chapter 2 fractions then he/she can work with the material in Chapter 1 and Chapter 3 with a probability 66%.
- If a student has sufficient preliminary knowledge then he/she can work with the material in Chapter 2 with a probability 59%.
- If a student has sufficient preliminary knowledge and can work with the material in Chapter 1 and Chapter 3 then he/she can work with the material in Chapter 2 with a probability 50%.

Similar inquires will show any other association rule.

3. Conclusion

In this paper association rules in education have been used for finding correlations among students' preliminary knowledge in a course and their abilities to work with the material in three chapters from the curriculum.

References

[1] R. Agrawal, T. Imielinski and A. Swami. Mining association rules between sets of items in large databases. *Proceedings of ACM SIGMOD international conference on management of data*, Washington, DC, USA, 207-216, 1993.

[2] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen and A.I. Verkamo. Fast discovery of association rules. (Uthurusamy, F., Piatetsky-Shapiro, G., Smyth, P. eds) *Advances in Knowledge discovery of association rules*, MIT Press, 307–328, 1996.

[3] R. Agrawal, and R. Srikant. Fast algorithm for mining association rules. *Proceedings of the 20th very large data base conference*, Santiago, Chile, 487-489, 1994.

[4] T. Bastide,R. Taouil, N. Pasquier, G. Stumme and L. Lakhal. Mining frequent patterns with counting inference. *SIGKDD explorations*, Special issue on scalable algorithms, 2(2):71–80, 2000.

[5] S. Brin, R. Motwani, J.D. Ullmann and S. Tsur. Dynamic itemset counting and implication rules for market basket data. *Proceedings of the ACM SIGKDD international conference on management of data*, Tuscon, AZ, USA, 255–264, 1997.

[6] D. Burdick, M. Calimlim and J. Gehrke. MAFIA: a maximal frequent itemset algorithm for transactional databases. *Proceedings of the 7th international conference on data engineering*, IEEE Computer Society, Heidelberg, Germany, 443-452, 2001.

[7] C. Carpineto and G. Romano. *Concept Data Analysis: Theory and Applications*. John Wiley and Sons, Ltd., 2004.

[8] B. A. Davey and H. A. Priestley. *Introduction to lattices and order*. Cambridge University Press, Cambridge, 2005.

[9] M. Delgado,D. Sanchez, M.J. Martin-Bautista and M.A. Vila. Mining association rules with improved semantics in medical databases. *Artificial Intelligence in Medicine*, 21(1-3):241-5, 2001.

[10] B. Ganter and R. Wille. *Formal Concept Analysis - Mathematical foundations*. Springer Verlag, Berlin Heidelberg, 1999.

[11] B. Ganter, G. Stumme and R. Wille. *Formal Concept Analysis - Foundations and Applications*. Springer Verlag, LNCS, 3626, 2005.

[12] D. Malerba, F.A. Lisi, A. Appice and F. Sblendorio. Mining spatial association rules in census data: a relational approach. *Proceedings of the ECML/PKDD'02 workshop on mining official data*, University Printing House, Helsinki, 80–93, 2002.

[13] N. Pasquier, T. Bastide, R. Taouil and L. Lakhal. Discovering frequent closed itemsets for association rules. *Proceedings of the 7th international conference on database theory*, Jerusalem, Israel, 398–416, 1999.

[14] N. Pasquier, T. Bastide, R. Taouil and L. Lakhal. Efficient mining of association rules using closed itemset lattices. *Journal of Information Systems*, 24(1):25-46, 1999.

[15] J. Pei, J. Han, and R. Mao. Closet: An efficient algorithm for mining frequent closed itemsets. *Proceedings of the ACM SIGKDD international workshop on research issues in data mining and knowledge discovery*, Dallas, USA, 21–31, 2000.

[16] R. Wille. Concept lattices and conceptual knowledge systems. *Computers Math. Applic.* 23(6-9):493–515, 1992.

[17] M.J. Zaki. Generating non-redundant association rules. *Proceedings of the 6th ACM SIGKDD international conference on knowledge discovery and data mining*, Boston, USA, 34–43, 2000.

[18] M.J. Zaki and C.-J Hsiao. CHARM: An efficient algorithm for closed itemset mining. *Proceedings of the 2nd SIAM international conference on data mining*, Arlington, VA, USA, 34–43, 2002.