Mining Bases for Association Rules Using Closed Sets

Rafik Taouil   Nicolas Pasquier   Yves Bastide   Lotfi Lakhal
LIMOS - Université Clermont-Ferrand II - 24 av. des Landais, 63177 Aubière France
{pasquier,bastide,taouil,lakhal}@libd2.univ-bpclermont.fr

Abstract

We address the problem of the usefulness and the relevance of the set of discovered association rules. Using the frequent closed itemset groundwork, we propose to generate bases for association rules, that are non-redundant generating sets for all association rules.

1. Introduction

Association rules are conditional implications between frequent itemsets. The problem of the usefulness and the relevance of the set of discovered association rules is related to the huge number of rules extracted and the presence of many redundancies among these rules for many datasets. We address this important problem using the Galois connection framework and we show that we can generate bases for association rules using the frequent closed itemsets extracted by the Close [4] or the A-Close [5] algorithms.

2. Galois connection framework

The association rule extraction is performed from a data mining context that is a triplet $\mathcal{D} = (\mathcal{O}, \mathcal{I}, \mathcal{R})$, where $\mathcal{O}$ and $\mathcal{I}$ are finite sets of objects and items respectively, and $\mathcal{R} \subseteq \mathcal{O} \times \mathcal{I}$ is a binary relation. Each couple $(o, i) \in \mathcal{R}$ denotes the fact that the object $o \in \mathcal{O}$ is related to the item $i \in \mathcal{I}$. The closure operator h of the Galois connection [1] is the composition of the application f, that associates with $O \subseteq \mathcal{O}$ the items common to all objects $o \in O$, and the application g, that associates with an itemset $I \subseteq \mathcal{I}$ the objects related to all items $i \in I$. The closure operator $h = f \circ g$ associates with I the maximal set of items common to all the objects containing I, i.e. the intersection of these objects. Using this closure operator, we define the frequent closed itemsets that constitute a minimal non-redundant generating set for all frequent itemsets and their support, and thus for all association rules, their support and their confidence. This property comes from the fact that the support of a frequent itemset is equal to the support of its closure and that the maximal frequent itemsets are maximal frequent closed itemsets [4, 5].

Definition 1 (Frequent closed itemsets) *A frequent itemset I is a frequent closed itemset iff $h(I) = I$. The smallest (minimal) closed itemset containing an itemset I is $h(I)$, i.e. the closure of I. We denote FC the set of frequent closed itemsets in $\mathcal{D}$.*

3. Bases for association rules

We adapt the Duquenne-Guigues basis for global implications [2, 1] and the Luxenburger basis for partial implications [3] to the exact association rules (100% confidence rules) and the approximate association rules respectively.

Theorem 1 (Duquenne-Guigues basis) *A frequent pseudo-closed itemset is a frequent itemset that is not closed and that contains the closures of all its subsets that are frequent pseudo-closed itemsets. Let FP be the set of frequent pseudo-closed itemsets in $\mathcal{D}$. The Duquenne-Guigues basis for exact association rules contains all rules of the form $I_1 \rightarrow (I_2 \setminus I_1)$ for $I_1 \in FP$, $I_2 \in FC$ and $h(I_1) = I_2$.*

Theorem 2 (Luxenburger basis) *The Luxenburger basis for approximate association rules contains all rules of the form $I_1 \rightarrow (I_2 \setminus I_1)$ for $I_1, I_2 \in FC$ and $I_1 \subset I_2$. The transitive reduction of this basis, i.e. for $I_1 \subset I_2$ and $\nexists I_3 \in FC$ such as $I_1 \subset I_3 \subset I_2$, is also a basis for all approximate association rules.*

All approximate association rules, their support and their confidence can be deduced from the Luxenburger basis, or its reduction, and all exact association rules can be deduced from the Duquenne-Guigues basis. They are minimal non redundant sets of association rules.

References

[1] B. Ganter and R. Wille, *Formal Concept Analysis: Mathematical foundations*, Springer, 1999.

[2] V. Duquenne and J.-L. Guigues, "Famille minimale d'implications informatives résultant d'un tableau de données binaires", *Math. et Sci. Hum.*, 24(95):5–18, 1986.

[3] M. Luxenburger, "Implications partielles dans un contexte", *Math., Inf. et Sci. Hum.*, 29(113):35-55, 1991.

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

[5] N. Pasquier, Y. Bastide, R. Taouil and L. Lakhal, "Discovering frequent closed itemsets for association rules", *Proc. ICDT Conf.*, pp 398–416, January 1999.