Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 127 (2018) 481–488

Procedia Computer Science

www.elsevier.com/locate/procedia

The First International Conference On Intelligent Computing in Data Sciences

Mining Negatives Association Rules Using Constraints

Said Jabbour[a], Fatima Ezzahra El Mazouri[b], Lakhdar Sais[a]

[a] *CRIL-CNRS, Université d'Artois, F-62307 Lens Cedex, France*
[b] *Faculty of Sciences and Technology, P.O. Box 30000, Fès, Morocco*

Abstract

Pattern discovery techniques, such as association rule discovery is one of the fundamental problem in data mining. Usually the task is limited to positive rules of the form $X \rightarrow Y$ when X and Y are susbets of items. To enlarge the knowledge discovery from data. Many works pointed out that other rules can be mined linking the present items in transactions with the missing ones designed as negative rules. To mine the most relevant negative rules, the mining task of negative association rules is often coupled with new measure such as lift or conviction to limit the set of extracted association rules.

In this work we address the problem of mining strong negative rules by extending the SAT-Based approach proposed in [1]. We show that the conviction constraint leads to a non-linear constraints that have to be managed efficiently to prune the search space. Experiments results explore the efficiency of our new approach.

Keywords: Data Mining; Association rules; Satisfiability

1. Introduction

Extracting association rules from transactional databases have received intensive research since its introduction by Rakesh Agrawal et al. in [2]. Initially referring to data analysis, it have been successfully extended to other application domains have been, including bioinformatics, medical diagnosis, networks intrusion detection, web mining, documents analysis, and scientific data analysis. This broad spectrum of applications enabled association analysis to be applied to a variety of datasets, including sequential, spatial, and graph-based data. Interestingly, association patterns are now considered as a building block of several other learning problems such as classification, regression, and clustering. To tackle many classical task of data mining such as itemset mining or associations rules mining, recent works deal with the formulation of such problems into constraint programming (CP), propositional satisfiability (SAT), and answer sat programming (ASP) [3, 4, 5, 6, 7, 8]. In [1], the authors proposed a new framework for mining

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000.
E-mail address: jabbour@cril.fr

1877-0509 © 2018 The Authors. Published by Elsevier B.V.

10.1016/j.procs.2018.01.146

association rules in one step using propositional satisfiability leading to a competitive approach compared to specialized techniques. Positive frequent association rules expressing causal relation between items present in transactions of a database have been extensively explored. Other works reveled that mining infrequent associations rules can bring up new knowledge. This is encouraged the look for new hidden associations rules considered as informative as classical ones. Henceforth, new approaches have been proposed to mine other association rules such as dissociation rules [9], rare rules [10], exception rules, and negatives rules. Negative association rules consider the same sets of items as positive association rules but, in addition, may also include negated items within the antecedent $X \rightarrow \neg Y$, $\neg X \rightarrow Y$, and $\neg X \rightarrow \neg Y$. In order to mine negative association rules, two main problem have to be faced. The first is the effective search for interesting itemsets and second how to effectively identify interesting negative association rules. In recent years, some researchers have proposed methods for mining positive and negative association [11, 12, 13, 14]. Most approach are based on Apriori-Like procedure to mine such rules. Usually, the confidence have is a standard measure for the interest of an association rules. Thereafter, several interestingness have been introduced. The Lift [15], called also interest, is defined as follows:

$$lift(X, Y) = \frac{freq(X \cup Y)}{freq(X) \times freq(Y)}$$

The lift refers to dependencies between the antecedent and consequent of an association rule. This measure have been criticized due to its symmetrical behavior. and a new measure called conviction [16] is introduced to measure the actual implication as opposed to co-occurrence. It is defined as:

$$conviction(X \rightarrow Y) = \frac{freq(X).freq(\neg Y)}{freq(X \cup \neg Y)}$$

where $\neg Y$ means the absence of Y in a transaction. Values in $[1, \infty[$ show that there is a positive dependence between X and Y.

Encouraged by the SAT based approach, In this work, we propose an extension to enumerate relevant negative association rules by considering conviction as interestingness measure. In particular, we show that conviction measure leads to a non-linear constraint that can be managed into a SAT solver while considering a particular branching heuristic. Experiments show a comparison between the number and the time needed to exact positive and negative association rules.

2. Preliminaries

2.1. Propositional Logic

We here define the syntax and the semantics of propositional logic. Let Prop be a countably set of propositional variables. We use the letters p, q, r, etc to range over Prop. The set of *propositional formulas*, denoted Form, is defined inductively started from Prop, the constant $\perp$ denoting false, the constant $\top$ denoting true, and using the logical connectives $\neg$, $\land$, $\lor$, $\rightarrow$. We use $Var(\phi)$ to denote the set of propositional variables appearing in the formula ϕ. The equivalence connective $\leftrightarrow$ is defined by $\phi \leftrightarrow \psi \equiv (\phi \rightarrow \psi) \land (\psi \rightarrow \phi)$.

A formula ϕ in *conjunctive normal form (CNF)* is a conjunction of clauses, where a *clause* is a disjunction of literals. A *literal* is a positive (p) or negated ($\neg p$) propositional variable. The two literals p and $\neg p$ are called *complementary*. A CNF formula can also be seen as a set of clauses, and a clause as a set of literals.

An *interpretation* $\mathcal{I}$ of a propositional formula ϕ is a function which associates a value $\mathcal{I}(p) \in \{0, 1\}$ (0 corresponds to $false$ and 1 to *true*) to the variables $p \in Var(\phi)$. A *model or an implicant* of a formula Φ is an interpretation $\mathcal{I}$ that satisfies the formula in the usual truth-functional way.