

# Re-mining Positive and Negative Association Mining Results

Ayhan Demiriz<sup>1</sup>, Gurdal Ertek<sup>2</sup>, Tankut Atan<sup>3</sup>, and Ufuk Kula<sup>1</sup>

<sup>1</sup> Sakarya University  
Sakarya, Turkey  
{ademiriz,ufukkula}@gmail.com

<sup>2</sup> Sabanci University  
Istanbul, Turkey  
ertekg@sabanciuniv.edu

<sup>3</sup> Isik University  
Istanbul, Turkey  
tatan@isikun.edu.tr

**Abstract.** Positive and negative association mining are well-known and extensively studied data mining techniques to analyze market basket data. Efficient algorithms exist to find both types of association, separately or simultaneously. Association mining is performed by operating on the transaction data. Despite being an integral part of the transaction data, the pricing and time information has not been incorporated into market basket analysis so far, and additional attributes have been handled using quantitative association mining. In this paper, a new approach is proposed to incorporate price, time and domain related attributes into data mining by re-mining the association mining results. The underlying factors behind positive and negative relationships, as indicated by the association rules, are characterized and described through the second data mining stage *re-mining*. The applicability of the methodology is demonstrated by analyzing data coming from apparel retailing industry, where price markdown is an essential tool for promoting sales and generating increased revenue.

## 1 Introduction

Association mining is a data mining technique in which the goal is to find rules in the form of  $X \Rightarrow Y$ , where  $X$  and  $Y$  are two non-overlapping sets of items or events, depending on the domain. A rule is considered as significant if it is satisfied by at least a percentage of cases specified beforehand (**minimum support**) and its confidence is above a certain threshold (**minimum confidence**). Conventional association mining considers “positive” relations as in the rule  $X \Rightarrow Y$ . However negative associations such as  $X \Rightarrow \neg Y$ , where  $\neg Y$  represents the negation (absence) of  $Y$ , might also be discovered through association mining.

Association mining has contributed to many developments in a multitude of data mining problems. Recent developments have positioned the association mining as one of the most popular tools in retail analytics, as well [1]. Traditionally,

association mining generates positive association rules that reveal complementary effects. In other words, the rules suggest that purchasing an item can generate sales of other items. Association mining can also be used to find so-called “Halo effects”, where reducing the price of an item can entice and increase the sales of another item. Although positive associations are an integral part of retail analytics, negative associations are not. However negative associations are highly useful to find out the substitution effects in a retail environment. Substitution means that a product is purchased instead of another one.

There have been numerous algorithms introduced to find positive and negative associations since the pioneering work of Agrawal et al. [2]. Market basket analysis is considered as a motivation and a test bed for these algorithms. Since the price data are readily available in the market basket data, one might expect to observe the usage of price data in various applications. Conceptually quantitative association mining [3,4] can handle pricing data and other attribute data. However pricing data have not been utilized before as a quantitative attribute except in [5], which explores a solution with the help of singular value decomposition. Quantitative association mining is not the only answer to analyze the attribute data by conventional association mining. Multidimensional association mining [4] is also a methodology that can be adapted in analyzing such data. Inevitably, the complexity of association mining will increase with the usage of additional attribute data where there might be both categorical and quantitative attributes in addition to the transaction data [6]. Even worse, the attribute data might be less sparse compared to transaction data.

The main objective of this paper is to develop an efficient methodology that enables incorporation of attribute data (e.g. price, category, sales timeline) to explain both positive and negative item associations. Positive and negative item associations indicate the complementarity and substitution effects respectively. To the best of our knowledge, there exists no methodological research in data mining literature that enables such a multi-faceted analysis to be executed efficiently and is proven on real world attribute data. A practical and effective methodology is developed to discover how price, item, domain and time related attributes affect both positive and negative associations by introducing a new data mining process.

As a novel and broadly applicable concept, we define *data re-mining* as mining a newly formed data from the results of an original data mining process. Aforementioned newly formed data will contain additional attributes on top of the original data mining results. These attributes in our case will be related to price, item, domain and time. Our methodology combines pricing as well as other information with the original association mining results within the framework of a new mining process. We thereby generate new rules to characterize, describe and explain the underlying factors behind positive and negative associations. Re-mining is a different process from post-mining where the latter only summarizes the data mining results. For example visualizing the association mining results [7] could be regarded as a post-mining activity. Our methodology extends and generalizes post-mining process.