Abstract

The itemsets discovered by traditional High Utility Itemsets Mining (HUIM) methods are more useful than frequent itemset mining outcomes; however, they are usually disordered and not actionable, and sometime accidental, because the utility is the only judgement and no relations among itemsets are considered. In this paper, we introduce the concept of combined mining to select combined itemsets that are not only high utility and high frequency, but also involving relations between itemsets. An effective method for mining such actionable combined high utility itemsets is proposed. The experimental results are promising, compared to those from traditional HUIM algorithm (UP-Growth).

Introduction

A retail or marketing manager often keeps an eye on which brand sells well or what commodity has high utility, and a promotion mixture is proved to be more profitable than a single item. Here, a good promotion mixture is actionable if it is not only profitable (high utility) but also well-selling (high frequency). In this paper, we call such promotion mixtures actionable itemsets (Adomavicius and Tuzhilin 1997). Unfortunately, this kind of patterns cannot be easily captured by the existing methods for reasons below.

Typical Frequent Itemsets Mining (FIM) methods only discover frequent itemsets. High Utility Itemsets Mining (HUIM) is more useful to discover those goods with high profits. In UP-growth (Tseng et al. 2010), which is a state-of-the-art algorithm in the HUIM area, the results are disordered, even though a UP-Tree with four strategies is proposed to make it much efficient and lossless.

However, to the best of our knowledge, there is no method to extract patterns with both above features simultaneously. More importantly, no HUIM methods consider relations between itemsets, which addresses the disordered issue in outcomes. In Combined Pattern Mining (Cao 2011), the concept of combined mining was proposed for solving this sort of problems by considering the interactions between relevant individual itemsets. Involving this concept in actionable HUIM, this paper proposes a combination of two structures, called Utility-Increasing Incremental Itemsets (UIII) and Frequency-Maximum Incremental Itemsets (FMII), for mining combined actionable high utility itemsets. Figure 1 shows an example of UIII structure: a scatterplot of HUIs, mined by traditional HUIM algorithm (UP-Growth), is firstly built with item-length increasing. The X axis is the length of each itemset, and the Y axis is the utility of each itemset. For example, $X_0$ is a length-one itemset (only one item included) with utility of 624879. $X_a$ is a length-two itemset (composed of $X_0$ and $\Delta X_1$) with utility of 876934, and $X_b$ is another length-two itemset with utility of 257377. The dotted lines are linked to each itemset composing several incremental itemset paths. In this figure, only $X_a$ (from $X_0$) and $X_c$ (from $X_b$) are UIIIs.

Figure 1: Example of a 4-item itemset with utility dynamics

Combined High Utility Itemsets (CHUI) consist of different itemsets that are extracted from the same itemset component in terms of their utility and frequency perspectives.

\[
\begin{align*}
X_0 + \Delta X_1 &= X_a \\
X_0 + \Delta X_2 &= X_b
\end{align*}
\]

(1)

In the CHUI, we call $X_a$, $X_b$ Derivative Itemset (DI), which is denoted as $X$ in the rest of the paper for global view, $X_0$ Underlying Itemset (UI) and the two variant itemsets $\Delta X_1$ and $\Delta X_2$ Additional Itemset (AI) (also denoted as $\Delta X$). Furthermore, $\Delta X_1 \cap \Delta X_2 = \emptyset$ and thus $X_a \neq X_b$. 
In the rest of the paper, we propose two functions to mine the range of utility increase (called \textit{contribution}) and the representative degree by frequency (called \textit{weight}) and evaluate the interestingness of each itemset by mean value of the two functions.

\textbf{Definition Declaration}

Notations used in this paper are defined as follows.

- \( D \): The database including all original transactions.
- \( X, \Delta X, X_0 \): The itemset composed of one or more non-repetitive and non-ordered items.
- \( \text{Supp}(X) \): The support of an itemset \( X \).
- \( u(X_0) \): The utility of \( X_0 \) in \( D \).

\textbf{Proposed Method: CHUIM}

Two parameters \textit{contribution} and \textit{weight} are calculated to extract the actionable combined high utility itemsets.

\textbf{The Contribution of UI to DI}

The contribution of \( X_0 \) transferring to \( X \) is defined as:

\[
C(X) = \frac{2}{1 + e^{-R}} - 1 \ u(X) > u(X_0)
\]  

By default, the itemsets are UIIs. In Eq. (2), \( R = \frac{u(X)}{u(X_0)} \).

\textbf{The Weight of DI}

The weight of DI \( X_a \) is to measure the co-occurrence extent (in another word, the impact of one on another) between UI and AI, which is defined as:

\[
W(X) = \frac{\text{Supp}(X)}{\text{Supp}(X_0 \cup \Delta X)}
\]  

\textbf{The Interestingness Measure}

The \textit{Quadratic Mean (QM)} (also known as \textit{Root-Mean Square}) of \( C(X) \) and \( W(X) \) is used to measure the significance of an itemset \( X \) in terms of both utility and frequency perspectives, which is defined as:

\[
QM(X) = \sqrt{\frac{C^2(X) + W^2(X)}{2}}
\]  

\textbf{Preliminary Outcomes}

To evaluate our proposed approach, we conduct experiments on the real datasets downloaded from (Brijs et al. 1999). The utilities of the items are randomly generated as (Tseng et al. 2010). The database is split into 10 parts randomly. The first part contains 10% transactions in the database and each later part contains 10% more than the former part.

The top 100 experimental results are selected and shown in figure 2. The figure on the left shows the comparison between UP-Growth and QM, while the figure on the right shows the result of QM and another mean function Harmonic Mean (HM) on \( C(X) \) and \( W(X) \). The X axis is the \( k_{th} \) part of the database, and the Y axis is the match ratio, which means the ratio of the exact patterns found in the \( k_{th} \) part and the \( (k+1)_{th} \) part. As can be seen from the figures, the QM method outperforms both HM and UP-Growth.

\textbf{Conclusion and Future Work}

In this paper, we have proposed a novel actionable combined high utility itemset mining method, which integrates the value of frequent pattern mining with high utility pattern mining and involve relations between itemsets. Preliminary experimental results show that the outcomes are more actionable and stable, and avoid accidentally high utility findings that often appear in the classic pattern-growth approaches. We’ll further explore more datasets with different characteristics, and a variety of functions and measurements to improve the stability of the resultant actionable combined high utility itemsets.

\textbf{References}


