Preserving Privacy in Mining Quantitative Associations Rules

Madhu V. Ahluwalia, University of Maryland Baltimore County, USA
Aryya Gangopadhyay, University of Maryland Baltimore County, USA
Zhiyuan Chen, University of Maryland Baltimore County, USA

ABSTRACT

Association rule mining is an important data mining method that has been studied extensively by the academic community and has been applied in practice. In the context of association rule mining, the state-of-the-art in privacy preserving data mining provides solutions for categorical and Boolean association rules but not for quantitative association rules. This article fills this gap by describing a method based on discrete wavelet transform (DWT) to protect input data privacy while preserving data mining patterns for association rules. A comparison with an existing kd-tree based transform shows that the DWT-based method fares better in terms of efficiency, preserving patterns, and privacy.

Keywords: Association Rule Mining, Discrete Wavelet Transform, Privacy Preserving Data Mining

INTRODUCTION

Association rule mining is an important knowledge discovery technique that is used in many real-life applications. As a motivating example, we use the retail business where data collected at a central site is routinely accessed by vendors to better plan and execute their logistics processes. The most commonly used data-mining task in the retail industry is association rule mining. In the simplest cases where transactions consist of market basket data, association rules reflect buying habits of customers. By counting the different items that customers place in their shopping baskets, association rules indicate items that are frequently purchased together by customers.

In addition to the categorical association rules (over items), association rules can be also defined over quantitative values. For example, a retailer’s data may hold information on quantities, discounts, and prices. A hypothetical sample of this data is shown in Table 1. Let \( Q \) be quantity, \( P \) be price, and \( D \) be discount. Figure 1 shows some quantitative association rules. A retailer may benefit from sharing such data with a wholesaler because such association rules may be utilized to improved supply-chain efficiency resulting in decreased pricing from the wholesaler. However, retailers may not want to reveal the exact price/unit of an item due to concerns over market competition. Thus this ar-
The following association rules can be derived from the sample data in Table 1.
1. If Q > 20 and P > 100 then D in (20%, 30%) (confidence 50%, support 20%). Rows 7 and 9 support this rule.
2. If Q in (10, 20] and P in (50, 100] then D in (10%, 20%) (confidence 100%, support 30%). This rule is obtained from rows 2, 5, and 8.
3. If Q in [0,10] and P in [0, 50] then D in [0-10%] (confidence 100%, support 30%). Rows 3, 6 and 10 generate this rule.

Table 1. Sample data to illustrate quantitative association rules

<table>
<thead>
<tr>
<th>Row_No.</th>
<th>Quantity</th>
<th>Price</th>
<th>Discount</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25.00</td>
<td>125.99</td>
<td>0.16</td>
</tr>
<tr>
<td>2</td>
<td>19.00</td>
<td>76.95</td>
<td>0.12</td>
</tr>
<tr>
<td>3</td>
<td>8.00</td>
<td>49.99</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>27.00</td>
<td>119.49</td>
<td>0.17</td>
</tr>
<tr>
<td>5</td>
<td>15.00</td>
<td>51.99</td>
<td>0.15</td>
</tr>
<tr>
<td>6</td>
<td>6.00</td>
<td>32.45</td>
<td>0.05</td>
</tr>
<tr>
<td>7</td>
<td>47.00</td>
<td>150.05</td>
<td>0.21</td>
</tr>
<tr>
<td>8</td>
<td>18.00</td>
<td>64.25</td>
<td>0.13</td>
</tr>
<tr>
<td>9</td>
<td>35.00</td>
<td>105.87</td>
<td>0.30</td>
</tr>
<tr>
<td>10</td>
<td>5.00</td>
<td>15.25</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Figure 1. Quantitative association rules derived from Table 1

Privacy preserving association rule mining has been studied for categorical data by (Evfimevski et al., July 2002; Lin et al., 2007; Rizvi et al., 2002). In all these cases a randomization technique is applied to distort the original data and enforce privacy. Evfimevski et al. (2002) and Rizvi et al. (2002) (Evfimevski et al., 2002; Rizvi et al., 2002) conduct randomization on a per-transaction basis, i.e. each original transaction is perturbed by inserting items into it or deleting items from it. Lin et al. (Lin et al., 2007) add whole new transactions to the set of original transactions. However, it is unclear how these techniques may be applied to quantitative data. For example, we cannot insert or delete items for quantitative data. Further, as pointed out in Zhang (2004), these techniques may reveal several actual items to an adversary, if a transaction consists of 10 or more items.

For transactions consisting of numerical items, Chen et al. (2005) proposed a solution that converts quantitative attributes to Boolean attributes. However, mining Boolean association rules creates a disclosure risk because input values of correlated items are restricted to 0 and 1. Also, large data sets such as point-of-sale data are not suitable for generating Boolean association rules.

One may try to use a random perturbation method (Agrawal & Srikant, 2000; Agrawal & Aggarwal, 2001) to add a random noise to the quantitative values. However, such techniques may not preserve the correlations between different attributes (e.g., between price, quantity, and discount). Thus it is unclear whether such
Related Content

PAKE on the Web
www.irma-international.org/chapter/pake-web/49511/

www.irma-international.org/article/a-generalized-approach-to-measure-market-timing-skills-of-fund-managers/111124/

Privacy Preserving and Efficient Outsourcing Algorithm to Public Cloud: A Case of Statistical Analysis
www.irma-international.org/article/privacy-preserving-and-efficient-outsourcing-algorithm-to-public-cloud/201507/

Ethics, Risk, and Media Intervention: Women’s Breast Cancer in Venezuela
www.irma-international.org/article/ethics-risk-and-media-intervention/133547/

Software Standards, Reliability, Safety, and Risk
www.irma-international.org/chapter/software-standards-reliability-safety-risk/28499/