ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 8, No. 1, 2013, *pp*. 063-074

A Model for Vague Association Rule Mining in Temporal Databases

Anjana Pandey [1], K.R.Pardasani [2]

[1] Deptt of Information Technology, University Institute of Technology,RGPV, Bhopal,India

2 Deptt of Mathematics, Maulana Azad National Institute of Technology, Bhopal,India

(*Received October15, 2012, accepted December29, 2012*)

Abstract. There are different university offering different types of courses over several years, and the biggest issue with that is how to get information to make course more effective. In real life these types of database usually contain temporal coherences, which cannot be captured by means of standard association rule mining. Here temporal Association rule mining can be used to evaluate the course effectiveness and helps to look for in regards to changes in performance of the course from time to time. For Example there is a course offering different topics. We can say that the topics having full attendance are totally effective and carry no hesitation information. While there are some topics which are almost fully attendant carry some hesitation information. This hesitation information is valuable and can be used to make the course more effective and interesting. Thus there is need for developing temporal vague association rule algorithms that reveal such hesitation information and temporal coherences within this data.

Keywords: ***Hesitation Information, Vague Association Rule, AH pair,Temporal database.***

1. Introduction

The problem of the discovery of association rules comes from the need to discover patterns in transaction data in a supermarket. But transaction data are temporal. For example, when gathering data about products purchased in a supermarket, the time of the purchase is registered in the transaction. This is called *transaction time*, in temporal databases jargon, which matches the *valid time*, corresponding to the time of the business transaction confirmation at the register[1] In large data volumes, as used for data mining purposes, we may find information related to products that did not necessarily exist throughout the data gathering period. So we can find some products that, at the moment of performing that mining, have already been discontinued. There may be also new products that were introduced after the beginning of the gathering. Some of these new products should participate in the associations, but may not be included in any rule because of support restrictions. For example, if the total number of transactions is 30,000,000 and we fix as minimum support 0.5 %, then a particular product must appear in, at least, 150,000 transactions to be considered frequent. Moreover, suppose that these transactions were recorded during the last 30 months, at 1,000,000 per month. Now, take a product that has been sold during the 30 months and has just the minimum support: it appears on average in 5,000 transactions per month. Consider now another product that was incorporated in the last 6 months and that appears in 20,000 transactions per month. The total number of transactions in which it occurs is 120,000; for that reason, it is not frequent, even though it is four times as popular as the first. However, if we consider just the transactions generated since the product appeared in the market, its support might be above the stipulated minimum. In this example, the support for the new product would be 2%, relative to its lifetime, since in 6 months the total of transactions would be about 6,000,000 and this product appears in 120,000 of them. Therefore, these new products would appear in interesting and potentially useful association rules. Each item, itemsets and rule has now an associated lifespan [2] which comes from the explicitly defined time in database transactions. The concept of temporal support is employed to mine association rules.

There are many items that are not bought but customers may have considered buying them. We call such information on a customer's consideration to buy an item the hesitation information [3] of the item,

since the customer hesitates to buy it. The hesitation information of an item is useful knowledge for boosting the sales of the item within given time period.

However, such information has not been considered in traditional association rule mining due to the difficulty to collect the relevant data in the past. Nevertheless, with the advances in technology of data dissemination, it is now much easier for such data collection.

A typical example is an online shopping scenario, such as "Amazon.com", for which it is possible to collect huge amount of data from the Web log that can be modeled to mine hesitation information. From Web logs, we can infer a customer's browsing pattern in a trail, say how many times and how much time s/he spends on a Web page, at which steps s/he quits the browsing, what and how many items are put in the basket when a trail ends, and so on. Therefore, we can further identify and Categorize different browsing patterns into different hesitation information with respect to different applications. The hesitation information can then be used to design and implement selling strategies that can potentially turn those "interesting" items into "under consideration" items and "under consideration" items into "sold" items.

From the literature [4], it is evident that very little attention has been paid for mining hesitation information .In this paper an attempt has been made to develop a vague set model for mining hesitation information within given time period .It is illustrated with the help of problem of choosing a course in an educational institute.

There are many different type of status of a piece of hesitation information (called hesitation status (HS)) [5]. Let us consider an example of class scenario that involves following type of status: (s1) attended class between 0 - 20%; (s2) Attended class between 0-40% (s3) Attended class between 0-60%.All of the above-mentioned types of HS are the hesitation information of those classes. Some of the types of HS are comparable based on some criterion, which means we can define an order on these types of HSs. For example, given a criterion as the possibility that the student attended the classes, we have $s_1 \leq s_2 \leq s_3$.

Here we are employ the vague set theory [3,4,5] to model the hesitation status of the course attended by the students. The main benefit of this approach is that the theory addresses the drawback of a single membership value in fuzzy set theory [6] by using interval-based membership that captures three types of evidence with respect to an object in a universe of discourse: support, against and hesitation. Thus, we naturally model the hesitation information of a course in the mining context as the evidence of hesitation.

The information of the "attended the class" and the "not attended the class" (without any hesitation information) in the traditional setting of association rule mining correspond to the evidence of support and against with respect to the class.

To study the relationship between the support evidence and the hesitation evidence with respect to topics, the concepts of attractiveness and hesitation are used, which are derived from the vague membership in vague sets. A topic with high attractiveness means that the topic is well attended and has a high possibility to be attended again next time. A topic with high hesitation means that the student is always hesitating to attend the topic due to some reason but has a high possibility to attend it next time if the reason is identified and resolved. For example, given the vague membership value, [0.5, 0.7], of a topic, the attractiveness is 0.6 (the median of 0.5 and 0.7) and the hesitation is 0.2 (the difference between 0.7 and 0.5), which implies that the student may attend the topic next time with a possibility of 60% and hesitate to attend the topic with a possibility of 20%.Using the attractiveness and hesitation of topics, we model a database with hesitation information as an AH-pair[4] database that consists of AH-pair transactions, where A stands for attractiveness and H stands for hesitation. Based on the AH-pair database, we then employed the notion of Vague Association Rules, which capture four types of relationships between two sets of items: the implication of the attractiveness/ hesitation of one set of items on the attractiveness/hesitation of the other set of items. For example, if we find an AH-rule like "People always buy quilts and pillows (A) but quit the process of buying beds at the step of choosing delivery method (H)". Thus, there might be something wrong with the delivery method for beds (for example, no home delivery service provided) which causes people hesitate to buy beds. To evaluate the quality of the different types of Vague Association Rule, four types of support and confidence are defined. We also investigate the properties of the support and confidence of Vague Association Rule, which can be used to speed up the mining process. To show the incidence of time in the amount and quality of the obtained rules we extend, vague association rule algorithm that generates the frequent itemsets.Proposed Algorithm is based on the items' period of life or lifespan.

This paper is organized as follows. Section 2 gives some preliminaries on vague set and temporal association rules. Section 3 discusses the algorithm that mines vague association rules. Section 4 illustrates the example. Section 5 reports the experimental results. Section 6 concludes the paper.