A Survey on Fuzzy Association Rule Mining

Harihar Kalia, Department of Computer Science and Engineering, Seemanta Engineering College, Jharpokharia, Mayurbhanj, Odisha, India

Satchidananda Dehuri, Department of Systems Engineering, Ajou University, Suwon, South Korea

Ashish Ghosh, Center for Soft Computing Research, Indian Statistical Institute, Kolkata, India

ABSTRACT

Association rule mining is one of the fundamental tasks of data mining. The conventional association rule mining algorithms, using crisp set, are meant for handling Boolean data. However, in real life quantitative data are voluminous and need careful attention for discovering knowledge. Therefore, to extract association rules from quantitative data, the dataset at hand must be partitioned into intervals, and then converted into Boolean type. In the sequel, it may suffer with the problem of sharp boundary. Hence, fuzzy association rules are developed as a sharp knife to solve the aforesaid problem by handling quantitative data using fuzzy set. In this paper, the authors present an updated survey of fuzzy association rule mining procedures along with a discussion and relevant pointers for further research.

Keywords: Association Rule, Confidence, Data Mining, Fuzzy Association Rule, Fuzzy Set, Itemset, Support

INTRODUCTION

Data mining (Cios, Pedrycz, Swiniarski, & Kurgan, 2012; Cao & Zhang, 2006) is an influential tool for knowledge mining. Association rule mining (Liu, Zhai, & Pedrycz, 2012) is a well-known data mining technique. Association permits to incarcerate all possible rules, which signify the presence of a number of items in accordance with the presence of a few other items in the same transaction.

The inventive motivation behind association rule mining was market basket analysis to study the buying habits of customers (Agrawal, Imielinski, & Swami, 1993). In recent days, association rule mining has been widened to areas like medical diagnosis (Rajendran, & Madheswaran, 2010; Xing & Pei, 2010), network security (Wang & Bridges, 2000; Mao & Zhu, 2002; Sheikhan & Jadidi, 2009), geographical database (Koperski, & Han, 1995), biological database (Gupta, Mangal, Tiwari, & Mitra, 2006; Martinez, Pasquier, & Pasquier, 2008), stock market databases (Saradhi, Ram Prakash, Pavan Kumar, Rao, & Vijay, 2012), web mining (Chai & Li, 2010), misuse detection (Sheikhan & Jadidi, 2009), manufacturing (Wantanabe, 2010) and electronic commerce (Natarajan & Sheka, 2005).
An association rule (Agrawal, Imielinski, & Swami, 1993) is an implication, \( P \Rightarrow Q \), where \( P \cap Q = \emptyset \), \( P \) and \( Q \) are set of items. Support and confidence are two primitive measures used for validating an association rule. The percentage of transaction containing both \( P \) and \( Q \) is defined as the support of the rule, whereas the ratio of the support of \( P \cup Q \) and support of \( P \) is defined as the confidence of the rule. Thus the association rule mining problem is defined as “the discovery of all association rules satisfying user defined support and confidence.”

A good number of algorithms and methods (Brin, Motwani, & Silverstein, 1997; Han, Pei, & Yin, 2000; Park, Chen, & Yu, 1997; Srikant & Agrawal, 2000; Zhang & Zhang, 2001; Ayubi, Muyeba, Baraani-Dastjerdi, & Keane, 2009) are designed/developed for association rule mining. Majority of these algorithms are meant for handling Boolean data.

Usually, transactional data in real-world applications consists of quantitative data. As an approach to handle those quantitative data, partitioning into intervals and treating each interval as a Boolean attribute is proposed (Srikant & Agrawal, 1995, 1996). In mining process, this discrete interval method would either discard or overemphasize the data points close to the boundary of the interval, called “sharp boundary problem.”

As a remedy, fuzzy association rules are proposed (Kouk, Fu, & Wong, 1998; Hong, Kuo, & Chi, 2001; Chen & Wei, 2002; Kaya, Alhajj, Polat, & Arslan, 2002; Muyeba, Khan, & Coenen, 2008) and became popular, as fuzzy set gives a soft transition between membership and non-membership of an item and hence very less boundary elements are excluded. Additionally, the linguistic variables like, “poor,” “moderate,” “rich,” that are used as fuzzy set make the association rule more interpretable. A comparison between fuzzy association rule and quantitative association rule can be found in Verlinde, De Cock, and Boute (2006) Hullermeier and Yi (2007).

A fuzzy association rule is of the form:

\[
\text{If } P \text{ is } F \text{ then } Q \text{ is } G, 
\]

where \( F \) and \( G \) are set of fuzzy sets used to describe the set of items \( P \) and \( Q \) respectively.

The fuzzy association rule is a rule which satisfy the user specified fuzzy support and fuzzy confidence thresholds.

Several studies emphasize on the problem of mining association rules from large databases. The focuses are on the issues like developing faster algorithms for existing classical methods and familiarizing/fitting the algorithms into diverse situations. Some instances are: association rule mining in data warehouses (Tjioe, & Taniar, Psaila, & Lanzi, 2000), multidimensional database mining (Lu, Feng, & Han, 2000; Liu, Wang, & Zhou, 2011; AL-Zawaidah, Jbara, & Abu-Tanoan, 2011), exceptional rule mining (Daly & Taniar, 2004) distributed algorithm for association rule mining (Ashrafi, Taniar, & Smith, 2004; Kantarcioglu & Clifton, 2004), redundant analysis (Ashrafi, Taniar, & Smith, 2007), ontology based rule mining (Wu, Lin, Tseng, & Wu, 2007), association rule mining from OLAP data cube (Bogdanova & Georgieva, 2005; Jigna & Mahesh, 2012; Messaoud, Loudcher, Boussaid, & Missaoui, 2006) and parallel algorithms for association rule mining (Zhang, Xu, Sheu, & Yamaguchi, 2011).

Rest of the paper is organized as follows. First we give an overview of association rule mining. Afterwards, we describe the framework of fuzzy rule mining. Brief descriptions of the fuzzy rule mining algorithms proposed in the literature are outlined following that. Conclusion and discussion are given in the last section.

ASSOCIATION RULE MINING

In data mining, association rule mining is a dynamic research area (Agrawal & Srikant, 1994). This section presents a summarized introduction of the association rule mining problem.
A New Similarity Metric for Sequential Data
[www.igi-global.com/article/new-similarity-metric-sequential-data/46941?camid=4v1a](www.igi-global.com/article/new-similarity-metric-sequential-data/46941?camid=4v1a)

Incremental Algorithm for Discovering Frequent Subsequences in Multiple Data Streams
[www.igi-global.com/article/incremental-algorithm-discovering-frequent-subsequences/58635?camid=4v1a](www.igi-global.com/article/incremental-algorithm-discovering-frequent-subsequences/58635?camid=4v1a)