ORIGINAL ARTICLE

Efficient mining fuzzy association rules from ubiquitous data streams

CrossMark

Amal Moustafa *, **Badr Abuelnasr, Mohamed Said Abougabal**

Computer and Systems Engineering Department, Faculty of Engineering, Alexandria University, Egypt

Abstract   Due to the development in technology, a number of applications such as smart mobile phone, sensor networks and GPS devices produce huge amount of ubiquitous data in the form of streams. Different from data in traditional static databases, ubiquitous data streams typically arrive continuously in high speed with huge amount, and changing data distribution. Dealing with and extracting useful information from that data is a real challenge. This raises new issues, that need to be considered when developing association rule mining techniques for these data. It should be noted, that data, in the real world, are not represented in binary and numeric forms only, but it may be represented in quantitative values. Thus, using fuzzy sets will be very suitable to handle these values.

In this paper the problem of mining fuzzy association rules from ubiquitous data streams is studied, and a novel technique FFP_USTREAM (Fuzzy Frequent Pattern Ubiquitous Streams) is developed. This technique integrates fuzzy concepts with ubiquitous data streams, employing sliding window approach, to mine fuzzy association rules. In addition, the complexity and the efficiency of this technique are discussed. Examples of real data sets are used **to test** the proposed technique. Further research issues are also suggested.

1. Introduction

Recent emerging applications, such as network traffic monitoring, sensor network data analysis, web click stream mining, power consumption measurement, and dynamic tracing of stock market fluctuations, call for studying a new kind of data. This is called stream data, which can be continuous, potentially infinite flow of information, as opposed to finite, statically stored data sets. Stream Data Mining is the process of

extracting knowledge structures from continuous, and rapid data records.

The dissemination of data streams systems, wireless networks and mobile/handheld devices motivates the need for an efficient data analysis tool capable of gaining insights about these continuous data streams [1]. Ubiquitous data streams mining (UDM) is the process of pattern discovery on mobile, embedded and ubiquitous devices. It represents the next generation of data mining systems, that will support the intelligent and time-critical information needs of mobile users, and will facilitate "anytime, anywhere" data mining.

Fuzzy logic is a type of logic used in artificial intelligence. It is referred to as a multi-valued logic. Instead of having two

values (true and false), there are a continuum of possible truth values [2]. In fuzzy logic, every proposition is a statement that is assigned a number between 0 (false) and 1 (true), such a statement is called a fuzzy proposition. Fuzzy logic provides a powerful tool to categorize a concept in an abstract way by introducing vagueness.

Many data streams applications exist, that require association rule mining, such as network traffic monitoring and web click streams analysis. These applications' goal is to discover important associations among items as the presence of some items will imply the presence of others.

Fuzzy association rule approach could combine data mining results with human expertise and background knowledge, in the form of rules, to attain labeled classes for classification of data streams. Another advantage of the fuzzy logic approach is that it gives classification results, which include a degree of probability.

This paper demonstrates the effectiveness of Fuzzy Association Rules Mining from Ubiquitous Data Streams. This will be revealed in the coming sections. For this purpose, the remaining part of the paper is organized as follows: the work related to Fuzzy Association rules mining and ubiquitous data streams mining are reviewed and summarized in Section 2. An efficient fuzzy association rules mining technique from ubiquitous data streams is proposed in Section 3, and its complexity is analyzed in Section 4. Moreover, experimental results are discussed in Section 5. The paper is concluded and future research issues are presented, in Section 6.

2. Related work

This paper belongs to different inter-related research fields. The two main related topics of this work are presented: Ubiquitous Data Streams Mining Techniques that can effectively analyze continuously streaming data and Fuzzy association rules mining algorithms. These two fields will be surveyed.

2.1. Ubiquitous data streams mining

The approach, based on finite statically stored data sets, is not satisfactory in several applications. These include wireless network analysis, intrusion detection, stock market analysis, sensor network data analysis, and, in general, any setting in which every information available should be used to make an immediate decision. Such situations demand new algorithms, that are able to cope with evolutions of data as shown in Table 1 [3].

Ubiquitous Data Mining (UDM) is the time-critical process of pattern discovery in data streams in a wireless environment [4]. The widespread use of mobile devices, with increasing computational capacity, is leading to the emergence of the ubiquitous computing paradigm. This paradigm facilitates continuous access to data and information by mobile users with handheld devices [5]. UDM is the process of analyzing data from distributed and heterogeneous sources with mobile devices or within sensor networks, where the data is continuously streamed to the device, and where there are temporal constraints, that necessitate analysis "anytime, anywhere" [6].

In the following subsections preprocessing required for data streams is presented.

Table 1 Taxonomy of data mining environment.

Data localization	
Centralized	A single entity can access every data
Distributed	Each node can access just a part of the data
Homogeneous	... data related to the same entity (e.g. people) are owned by just one node
Heterogeneous	... data related to the same entity (e.g. people) may be spread among several nodes
Data evolution	
Statical	Data are definitively stored and invariable (e.g. related to some past and concluded event)
Incremental	New data are inserted and access to past data is possible (e.g. related to an ongoing event)
Evolving	The dataset is modified with either updates, insertions or deletions, and access to past data is possible
Streaming	Data arrives continuously and for an indefinite time. Access to past data is restricted to a limited part of them or summaries

2.1.1. Data streams techniques

Research problems and challenges that appeared in mining data streams have their solutions using well established statistical and computational approaches. These solutions could be categorized to data-based and task-based ones. This classification is depicted in Fig. 1 [7]. In data-based solutions, the idea is to examine only a subset of the whole dataset or to transform the data vertically or horizontally to an approximate smaller size data representation. On the other hand, in task-based solutions, techniques from computational theory have been adopted to achieve time and space efficient solutions.

2.1.1.1. Data-based techniques. Data-based techniques refer to summarizing the whole dataset or choosing a subset of the
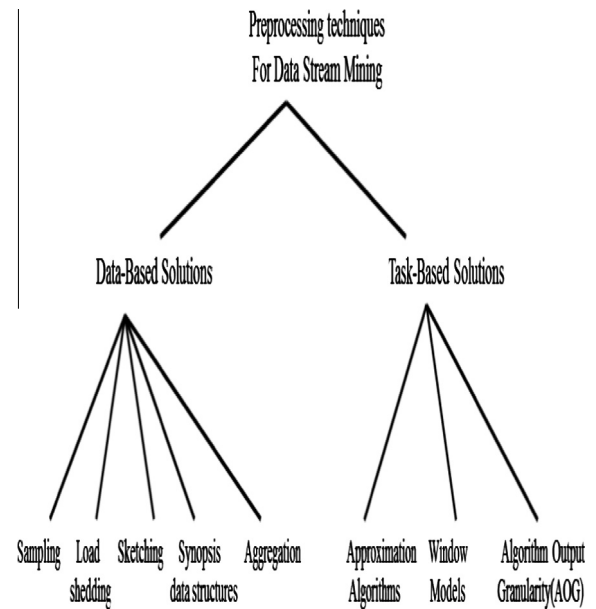

Figure 1 Classification of data streams preprocessing methods [7].