Multidimensional On-line Mining

Ching-Yao Wang[1], Tzung-Pei Hong[2], and Shian-Shyong Tseng[1]

[1] Institute of Computer and Information Science, National Chiao-Tung University, Taiwan, R.O.C.
[2] Department of Electrical Engineering, National University of Kaohsiung, Taiwan, R.O.C.

Abstract. In the past, incremental mining approaches usually considered getting the newest set of knowledge consistent with the entire set of data inserted so far. Users can not, however, use them to obtain rules or patterns only from their interesting portion of the data. In addition, these approaches only focused on finding frequent patterns in a specified part of a database. That is, although the data records are collected in under certain time, place and category, such contexts (circumstances) have been ignored in conventional mining algorithms. It will cause the lack of patterns or rules to help users solve problems at different aspects and with diverse considerations. In this paper, we thus attempt to extend incremental mining to online decision support under multidimensional context considerations. We first propose the *multidimensional pattern relation* to structurally and systematically retain the additional context information and mining information for each inserted dataset into a database. We then develop an algorithm based on the proposed multidimensional pattern relation to correctly and efficiently fulfill diverse on-line mining requests.

***Keywords*:** data mining, association rule, incremental mining, multidimensional mining, constraint-based mining, data warehouse.

1 Introduction

Knowledge discovery and data mining technology has recently gotten much attention in the field of large databases and data warehouses. It attempts to discovery non-trivial, implicit, previously unknown and potentially useful knowledge from databases or warehouses [2][7][13], and thus aids managers to make correct decisions. Among the various types of databases and mined knowledge, mining association rules from transaction databases is most commonly seen [1][3]. It discovers relationships among items such that the presence of certain items in a transaction tends to imply the presence of certain other items. Since this mining process is rather time-consuming, many approaches

have been proposed to reduce the computation time and improve the performance. Some famous mining algorithms are Apriori [3], DHP [21], Partition [24], DIC [5], Sampling [19], GSP [4][25], and among others. These algorithms process data in a batch way and reprocess the entire database whenever either the data stored in a database or the thresholds (i.e. the minimum support or the minimum confidence) set by users are changed. They do not utilize previously mined patterns for later maintenance, and will require considerable computation time when a database is massive in size.

Incremental mining algorithms were thus proposed, which utilized previously mining information (such as large itemsets) to reduce the cost of re-computation [8][9][14][15][18][23][26][27][28]. These algorithms usually consider getting the set of knowledge consistent with the entire set of data inserted so far. Users can not, however, easily use them to obtain rules or patterns only from their interesting portion of data.

In addition, data records are usually collected in blocks in different contexts (circumstances). The context attributes such as time, place and category have been ignored in conventional mining algorithms [12]. If mining algorithms can consider related context information, they could help users solve problems at different aspects and with diverse considerations. Constraint-based and multidimensional mining techniques were thus proposed to achieve this purpose [11][12][13][17][20][22].

In this paper, we attempt to extend the concept of effectively utilizing previously discovered patterns in incremental mining to support online generation of association rules under multidimensional considerations. We first propose the *multidimensional pattern relation* to structurally and systematically store the additional context information and mining information for each inserted dataset. This idea is conceptually similar to the construction of a data warehouse for OLAP [6][16][29]. Both of them preprocess the underlying data in advance, integrate related information, and consequently store the results in a centralized structural repository for later use and analysis. The mining information in a multidimensional pattern relation is, however, unlike the summarized information of *fact attributes* in a data warehouse. They can not be easily aggregated to fulfill users' mining requests. We thus develop an on-line mining algorithm based on the proposed multidimensional pattern relation to correctly and efficiently fulfill different mining requests.

2 Some Related Works

As mentioned above, most mining algorithms process data in a batch way and must reprocess the entire database whenever either the data stored in a database or the thresholds set by users are changed. They do not use previously mining information and may need considerable computation time to get the newest set of rules or patterns [8]. In real-word applications, new records may be inserted and old records may be deleted or modified along with time.