Chapter III

Maintenance of Association Rules Using Pre-Large Itemsets

Tzung-Pei Hong, National University of Kaohsiung, Taiwan
Ching-Yao Wang, National Chiao-Tung University, Taiwan

Abstract

Developing an efficient mining algorithm that can incrementally maintain discovered information as a database grows is quite important in the field of data mining. In the past, we proposed an incremental mining algorithm for maintenance of association rules as new transactions were inserted. Deletion of records in databases is, however, commonly seen in real-world applications. In this chapter, we first review the maintenance of association rules from data insertion and then attempt to extend it to solve the data deletion issue. The concept of pre-large itemsets is used to reduce the need for rescanning the original database and to save maintenance costs. A novel algorithm is proposed to maintain discovered association rules for deletion of records. The proposed algorithm doesn’t need to rescan the original database until a number of records have been deleted. If the database is large, then the number of deleted records allowed will be large too. Therefore, as the database grows, our proposed approach becomes increasingly efficient. This characteristic is especially useful for real-world applications.
Due to the increasing use of very large databases and data warehouses, mining useful information and helpful knowledge from transactions is evolving into an important research area. In the past, researchers usually assumed databases were static to simplify data-mining problems. Thus, most proposed algorithms focus on batch mining (Agrawal, Imielinski, & Swami, 1993; Agrawal & Srikant, 1994; Agrawal & Srikant, 1995; Agrawal, Srikant, & Vu, 1997; Han & Fu, 1995; Mannila, Toivonen, & Verkamo, 1994; Park, Chen, & Yu, 1997; Srikant & Agrawal, 1995; Srikant & Agrawal, 1996) and do not utilize previously mined patterns for later maintenance. This may require considerable computation time to obtain the updated set of association rules or patterns (Cheung, Han, Ng, & Wong, 1996).

Researchers have recently developed efficient mining algorithms for maintaining association rules and avoiding the above-mentioned shortcomings when new transactions are inserted. Examples include the FUP algorithm (Cheung et al., 1996), the adaptive algorithm (Sarda & Srinivas, 1998), and the incremental updating technique based on the concept of negative border (Feldman, Aumann, Amir, & Mannila, 1997; Thomas, Bodagala, Alsabti, & Ranka, 1997). The common idea among these approaches is that previously mined patterns are stored in advance for later use. When new transactions are inserted, a large part of the final results can be obtained by comparing the patterns mined from the newly inserted transactions with the pre-stored mined knowledge. Only a small portion of the patterns need to be re-processed against the entire database, thus saving much computation time.

Among these approaches, the FUP algorithms (Cheung et al., 1996) store the previously mined large itemsets for later maintenance. Some approaches utilize negative borders (Feldman et al., 1997; Thomas et al., 1997) to enlarge the amount of pre-stored mined information, thus improving maintenance performance at the expense of storage space. Furthermore, we proposed an incremental mining algorithm based on the concept of pre-large itemsets for data insertion (Hong, Wang, & Tao, 2001). The concept of pre-large itemsets is denoted as the set of itemsets having support between a lower support threshold, which is smaller than the given minimum support, and an upper support threshold, which is equal to the given minimum support. Therefore, using the pre-large itemsets to enlarge the amount of pre-stored mined information can avoid rescanning the original database until the accumulative amount of new transactions exceeds the safety bound at the expense of storage spaces. This is because they act as a buffer to avoid the movements of itemset directly from small to large and vice-versa during the incremental mining process.

In addition to record insertion, record deletion is also commonly seen in real-world applications. For example, the records which were generated some years ago may be moved to a magnetic tape. Developing efficient maintenance algorithms for deletion of records is practical and necessary. In this chapter, we first review the
Related Content

An Integrated Query Relaxation Approach Adopting Data Abstraction and Fuzzy Relation
www.igi-global.com/article/integrated-query-relaxation-approach-adopting/47419?camid=4v1a

Evaluation of MDE Tools from a Metamodeling Perspective
www.igi-global.com/article/evaluation-mde-tools-metamodeling-perspective/3393?camid=4v1a

Differential Learning Expert System in Data Management
www.igi-global.com/chapter/differential-learning-expert-system-data/20745?camid=4v1a
Enhancing Database Access Control by Facilitating Non-Key Related Cover Stories
www.igi-global.com/article/enhancing-database-access-control-facilitating/3334?camid=4v1a