Abstract

This paper presents a novel system for Incremental Association Rules Mining from Medical Documents (IARMMMD). The system concerns with maintenance of the discovered association rules and avoids redoing the mining process on whole documents during the updating process. The design of the system is based on concepts representation. It designed to develop our previous D-EART system. The IARMMMD improves the updating process, and will lead to decrease the number of scanning and the execution time. The system consists of three phases that are Text Preprocessing, Incremental Association Rule Mining, and Visualization phase. Hash-based Incremental Association Rule Mining Algorithm (HIARM) is introduced in the mining phase. The algorithm employs the power of data structure called Hash Table. The performance of the algorithm is compared with both Apriori and FUP algorithms for the execution time and the evaluation of the extracted association rules. The results reveal that the number of extracted association rules in the IARMMMD system is always less than that in Apriori-based and FUP-based systems. Furthermore, the execution time of HIARM algorithm is much better than Apriori and FUP algorithms in the updating process in all experimental cases.

References

of items in large databases. In Proceedings of the ACMSIGMOD Int. Conf. on Management of Data, Washington, D. C.
Applications (CSA-08), Australia.

- J. Han, J. Pei, Y. Yin and R. Mao, "Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree", Data Mining and Knowledge Discovery, pp. 53–87, IEEE 2004.
- Han, J., Pei, J. and Yin, Y. 2000. Mining frequent patterns without candidate generation. The ACM SIGMOD Int. Conference on Management of Data.

**Index Terms**

Computer Science

Information Sciences
Keywords
Knowledge Engineering  Text mining  Data mining  Knowledge Mining  Incremental Association Rules Mining