Abstract

We consider the problem of discovering frequent item sets and association rules between items in a large database of transactional databases acquired under uncertainty. A probabilistic database considered here is one in which with each transaction associated is a probability, represents the confidence that the transaction will occur with given associated certainty. In this paper, we address the problem of the efficiency of the main phase of most data mining applications: The frequent pattern extraction. This problem is mainly related to the number of operations required for counting pattern supports in the database and we propose a new method, called counting inference probabilistic frequent pattern miner in probabilistic databases, this algorithm allows to perform as few support counts as possible. It is optimized to reduce the number of database scan as well as the number of patterns for which explicit support count is required. Using this method, the support of a pattern is determined without accessing the database whenever possible, using the supports of some of its sub-patterns called key patterns. This method was implemented in the CIPFP, counting inference based probabilistic frequent pattern mining algorithm that is an optimization of the simple and efficient Apriori algorithm. The
goal is to transform all key patterns into non-key patterns as early as possible as for non-key-patterns database scan is not required at all.

References


46. M. Yiu et al. Efficient evaluation of probabilistic advanced spatial queries on existentially uncertain data. IEEE Transactions on Knowledge and Data Engineering, 21(9), 2009.

Index Terms

Computer Science Algorithms

Keywords

Probabilistic frequent patterns, probabilistic frequent rule, key-patterns, non-key-patterns,