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


indices

Table of Contents

1. Introduction
2. Background
3. Proposed Methodology
4. Experimental Results
5. Conclusion

1. Introduction

This chapter introduces a novel approach to mining frequent patterns in probabilistic databases. The method leverages counting and inference techniques to identify patterns that are both frequent and representative of future possibilities. The approach is designed to be scalable and adaptable to different types of probabilistic data.

2. Background

The field of probabilistic databases has seen significant advancements in recent years. Techniques such as probabilistic inference and counting have been used to handle uncertainty in data. This chapter builds upon these foundations, incorporating future possibilities into the pattern mining process.

3. Proposed Methodology

The proposed methodology involves a two-step process. First, a probabilistic inference step is used to estimate the likelihood of future events based on historical data. Second, a counting step is employed to identify patterns that are frequent and have high predictive power. This approach is particularly useful in scenarios where data has inherent uncertainty.

4. Experimental Results

The effectiveness of the proposed methodology was evaluated through a series of experiments. The results demonstrated that the approach could accurately identify key-patterns and non-key-patterns, outperforming existing methods in terms of both precision and recall.

5. Conclusion

In conclusion, the novel algorithms CIPFP for mining frequent patterns using counting inference from probabilistic databases represent a significant advancement in the field. The approach offers a robust framework for handling uncertainty and predicting future trends, making it applicable in various domains such as finance, healthcare, and environmental monitoring.

Keywords

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