

## **AN EFFECTIVE DATA SANITIZATION APPROACH FOR PRIVACY PRESERVING UTILITY ITEMSET MINING**

**R. R. RAJALAXMI AND A. M. NATARAJAN**

### **Abstract**

The current trend in business collaboration demands the need to share data or mined results to gain mutual benefit. But it has also raised a potential threat of revealing sensitive information while releasing such data to promote business growth. Hence data sanitization is used to conceal the sensitive itemsets present in the source database with appropriate modifications and the modified database is released. The problem of finding an optimal solution for the sanitization process which minimizes the non-sensitive patterns lost is NP-hard. Recent researches in data sanitization approaches hide the sensitive itemsets by reducing the support of the itemsets which considers only the presence or absence of itemsets. But in real world, the transactions made by customers contain the items with purchased quantities along with their unit price. So it is essential to consider the utility of itemsets in the source database. Utility mining model was introduced to find high utility itemsets. In this paper, the focus is primarily on protecting privacy in utility mining which minimizes the number of non-sensitive patterns lost. This paper proposes an approach for sanitization such that minimal changes are made to the database with minimum number of non-sensitive itemsets removed from the database.

-----  
**Keywords:** Data mining, Frequent Itemset mining, Association Rule Mining, Utility mining, Data sanitization