Abstract

FP-Growth is one of the most effective and widely used association rules mining algorithm for discovering interesting relations between items in large datasets. Unfortunately, classical FP-Growth mines frequent patterns by using single user-defined minimum support threshold. This is not adequate for real life applications such as crime patterns mining. On one side, if minimum support is set too low, huge amount of crime patterns (including uninteresting patterns) may be generated, and on the other side, if it is set too high lots of interesting patterns (including seasonal patterns) may be lost. This paper proposes the use of Multiple Item Support (MIS) thresholds instead of single minimum support to tackle the challenge. We employ Shannon entropy method to develop an algorithm that obtains MIS values from crime datasets. The proposed approach is tested on different sizes of input data via a developed working prototype. Experimental results show that our suggested approach outperforms classical FP-Growth in terms of running time and memory use.
Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, Pages 931-940.


Index Terms

Computer Science

Information Systems

Keywords

FP-Growth, Crime Pattern, Multiple Minimum Supports, Shannon Entropy.