

CSCI 6405, Winter 2017

Project Proposal: Improvised apriori algorithm using frequent pattern tree for real time applications in data mining

Submission date: 24/03/2017

Student ID: B00759300

Student ID: B00762770

Student Name: Tushar Bhalla

Student Name: Arun Athisamy

INTRODUCTION

Apriori Algorithm is one of the most important algorithm which is used to extract frequent itemsets from large database and get the association rule for discovering the knowledge. It basically requires two important things: minimum support and minimum confidence. First, we check whether the items are greater than or equal to the minimum support and we find the frequent itemsets respectively. Secondly, the minimum confidence constraint is used to form association rules. Based on this algorithm, proposed paper indicates the limitation of the original Apriori algorithm of wasting time and space for scanning the whole database searching on the frequent itemsets, and present an improvement on Apriori.

Project Goal

The aim of the project is to implement the proposed improved apriori algorithm in an real time intrusion detection dataset and compare the execution time with the normal apriori algorithm.

Language : C#.Net

Tools to be used : Weka

Sample Output:

Table 6. The Time Reducing Rate of Improved Apriori according to the number of transactions.

T	Original Apriori(s)	Improved Apriori (s)	Time reducing rate(%)
T1	1.776	0.654	63.17%
T2	8.221	3.982	51.56%
T3	6.871	2.302	66.49%
T4	11.940	2.446	79.51%
T5	82.558	17.639	78.63%

Table 7. Time Reducing Rate of Improved Apriori according to the value of minimum support.

Min_Sup	Original Apriori (s)	Improved Apriori (s)	Time reducing rate (%)
0.02	6.638	1.047	84.22%
0.04	1.855	0.398	78.54%
0.06	1.158	0.28	75.82%
0.08	0.424	0.183	56.83%
0.10	0.382	0.149	60.99%

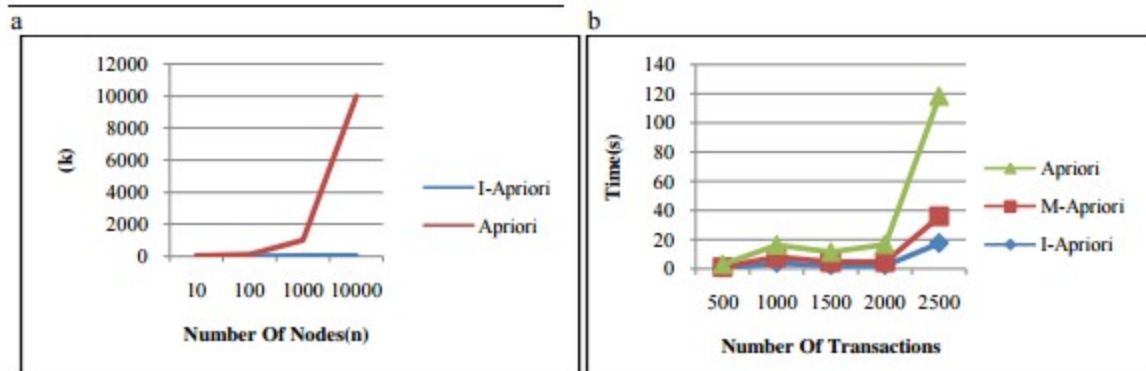


Fig. 1. (a) Graph between number of nodes and k; (b) Time consuming comparison for different groups of transactions.

References:

[1] Akshita Bhandaria , Ashutosh Gupta and Debasis Dasa. Improvised apriori algorithm using frequent pattern tree for real time applications in data mining. International Conference on Information and Communication Technologies (ICICT 2014) Procedia Computer Science 46 (2015) 644 – 651.

Research Paper Link:

<http://www.sciencedirect.com.ezproxy.library.dal.ca/science/article/pii/S1877050915001799>

Dataset Link :

<https://archive.ics.uci.edu/ml/machine-learning-databases/kddcup99-mld/>