

Question #2 Report

Implementation and results of common association rule mining algorithms



Abhinav Khanna(2017CSB1061)

CS524: Data mining

INTRODUCTION

In this part of the assignment, I have generated my own dataset, and run the various algorithms mentioned in the previous question.

THEORETICAL FOUNDATION

I have covered the theoretical aspects of the three algorithms in the report of Question 1. Thus, I will not be covering the theoretical concepts in this report.

IMPLEMENTATION

I have generated a dataset consisting of 100 items, distributed in a 1000 transactions. The

support of each item is fixed. In this particular implementation, I have followed the steps that item '1' occurs in only 1 transaction, chosen randomly. Similarly item '2' occurs in only 2 randomly chosen transactions, and so on upto item '100'. This creates a vertical dataset, with a list of transactions associated with each item. I have then converted this into a horizontal dataset, and then run the three algorithms on this dataset. The implementation of these algorithms has been covered in the previous report.

TIME AND SPACE COMPLEXITY RESULTS

Algorithm	Performance
Apriori	0.74 sec 9.88 KB Min_sup = 0.01
FP tree	0.4 sec 38 KB Min_sup = 0.01
ECLAT	145 sec 268 KB Min_sup = 0.01

Table 1

Note: Space mentioned is the memory occupied by the respective jupyter notebook, which is indicative of the actual space complexity.

● Dataset Information

Properties	Results
# of transactions	1000
Avg. width of transaction	4.95
Maximal frequent itemset size	2
# of maximal frequent itemsets	121

Total # of items	100
------------------	-----

Table 2

CONCLUSION

In this part of the assignment, I have concluded that we can generate datasets in accordance to our requirement of total number of items, total number of transactions, width of transactions etc, and that we do not necessarily need to rely on real life datasets to perform our experiments.

REFERENCES

1. Data Mining and Analysis: Fundamental Concepts and Algorithms: Textbook by Mohammed J. Zaki and Wagner Meira
2. Introduction to Data Mining: Book by Michael Steinbach, Pang-Ning Tan, and Vipin Kumar
3. Class Notes