Implementing Frequent Item set Mining by Overcoming Over-Scan Problems

Srinivasa Rao Divvela, Research Scholar, V Sucharita

Abstract: This paper overcomes the problem of over-scan problems which are occurring frequently in the frequent itemset mining. In case of improved apriori we improve the efficiency and decrease the time lapse but it is unable to solve the over-scan problem efficiently. We now propose a systematic approach for the immediate solving of the over-scan problems by implementing the RP Tree on spark framework. The adaptation of this approach will be very useful in the formation of the tree of frequent patterns and also for the visualization of the frequent-1-itemsets. This is mainly to overcome the over-scan problems in the previous improved algorithm.

Keywords: Frequent itemset mining, RP Tree, Spark Framework, Visualization, Over-scan problems.

I. INTRODUCTION

Extraction of data from databases has been critical in the modern day for which the using of the associate rule mining algorithms. Events that are frequently occurring are of less importance when compared to those which occur rarely.

II. PROPOSED SYSTEM

We propose a new systematic approach by utilizing the RP Tree on the spark framework. The adaptation of this approach will be very useful in the formation of the tree of frequent patterns and also for the visualization of the frequent-1-itemsets. This is mainly to overcome the over-scan problems in the previous improved algorithm. We use two transactional datasets for the implementation.

Keywords: Frequent itemset mining, RP Tree, Spark Framework, Visualization, Over-scan problems.
Implementing Frequent Item set Mining by Overcoming Over-Scan Problems

### III. IMPLEMENTATION

By building a spark cluster distributing environment with 5 nodes. We need an active Ethernet connection for the implementation of the procedure in the compilation using the python language. We implement the procedure and the algorithms by using of the algorithms on two specific datasets, namely

- **T1014D100k**
- **T40I10D100k**

<table>
<thead>
<tr>
<th>data set</th>
<th>The number of items</th>
<th>The number of transactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1014D100k</td>
<td>870</td>
<td>10000</td>
</tr>
<tr>
<td>T40I10D100k</td>
<td>942</td>
<td>10000</td>
</tr>
</tbody>
</table>

Tab 1: Given dataset Attributes
IV. RESULTS

The following graphs depict the differential curves between both the implemented datasets under stipulated conditions which are to be suited for the application procedure. The number of iterations of each dataset used is also given as a graph with respect to the RP Tree and Spark implementations that are done in this process.

![Graph 2: Min_rare_Sup difference](image)

![Graph 3: Cluster node number difference](image)

![Graph 4: T1014D100k](image)

![Graph 5: T40I10D100k](image)

V. CONCLUSION

By this implementation we overcome the over scan problems which are very turbulent in the frequent item set mining process. The implementing of the systematic approach of the algorithm of RP Tree on the spark framework is an added feature for the overcoming of the problems of the previous algorithms and procedures. The total time of execution for the process is minimized due to the process by the reduction of the combination of traversals in the time usage. Thus with the proposed approach we can overcome the over scan problems as well as increase the time efficiency of the system.

REFERENCES

3. “Mining association rules with multiple minimum supports” by “Liu B.” in Knowledge
Implementing Frequent Item set Mining by Overcoming Over-Scan Problems

Discovery and Data Mining. 1999:337-341.


AUTHORS PROFILE

Divvela Srinivasa Rao is currently working as Sr. Assistant Professor in Laki Reddy Balireddy College of Engineering, Mylavaram. He is a Research scholar in KL Universiry. His area of interest is Data Mining and Knowledge Engineering. He has total 13 years of teaching experience in various Engineering colleges in Andhra Pradesh. He has published 10 papers in various national and international Journals, conference proceedings and Conferences.

Dr. V. Sucharita is currently working as Professor in the Department of Computer Science and Engineering in Narayana Engineering College, Gudur. She has obtained her Ph.D degree in the faculty of Computer Science with Data Mining and Neural Networks as specialization from Sri Padmavati Mahila Visvavidyalayam, Tirupathi. She has 17 years of Teaching Experience in various Engineering colleges in Andhra Pradesh. Prior to joining Narayana Engineering College, She worked as Professor in K.L University, Guntur. She has published 60 papers in various national and international Journals, conference proceedings and Conferences. She is an Editorial Board member in reputed journals.