CHAPTER 7

COMPARISON OF APRIORI, APRIORIALLHYBRID AND PARTITION ALGORITHMS

7.1 Introduction

A frequent pattern mining plays a very important and essential role in mining frequently occurring data itemsets, and also for mining association among the data items. There are many algorithms to mine the frequent data items. In this chapter, Apriori algorithm, AprioriAllHybrid and Partition algorithm have employed for large amount of data sets and the results are compared.

Apriori algorithm is used to find the frequent patterns available in the database. The disadvantage of this method is, large number of candidates has been generated and scan time have increased because, for each time whole database have to be scanned. So AprioriAllHybrid algorithm has been chosen to improve the performance of Apriori algorithm.
In this algorithm, original database is read only once, afterwards new temporary intermediate database is created at each step of iteration, as a result database size is reduced and no of candidate sequences generated is also reduced. To make it more efficient, partition algorithm concept has been included. In this, separate partitions will be created for each 1 itemsets, 2 itemsets, 3 itemsets etc. so that scan time is decreased and as a result performance is increased.

7.2 Comparison of Algorithms

Apriori algorithm, AprioriAllHybrid algorithm and Partition algorithm have been compared and the performance evaluation is shown in the table 7.1

Table 7.1 Performance Evaluation of Apriori, AprioriAllHybrid and Partition Algorithm

<table>
<thead>
<tr>
<th>S.No</th>
<th>ALGORITHMS</th>
<th>TOTAL NO OF RECORDS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>30000</td>
</tr>
<tr>
<td>1</td>
<td>Apriori</td>
<td>264 min</td>
</tr>
<tr>
<td>2</td>
<td>AprioriAllHybrid</td>
<td>135 min</td>
</tr>
<tr>
<td>3</td>
<td>Partition</td>
<td>19 min</td>
</tr>
</tbody>
</table>
7.3 Performance Analysis

The total number of records taken for evaluation is 30000, 20000 and 10000. For 30000 records, Apriori algorithm takes 264 minutes, AprioriAllHybrid takes 135 min and partition algorithm takes 19 minutes only. For 20000 records, Apriori algorithm has taken 178 minutes, AprioriAllHybrid algorithm taken 82 minutes and Partition algorithm has taken 13 minutes. For 10000 records, Apriori algorithm takes 86 minutes, AprioriAllHybrid algorithm takes 41 minutes and Partition algorithm takes 6 minutes.

From the above table 7.1, the execution time of the algorithm to find sequential pattern depends on total no of candidates generated at each level and the time taken to scan the database.

Figure 7.1 Performance Graph of Apriori, AprioriAllHybrid and Partition Algorithm
The performance of the algorithms has shown in graphical representation in the figure 7.1. The performance of partition algorithm will be for better than AprioriAllHybrid and Apriori algorithm.

### 7.4 Result and Discussion

The execution time is increasing with increase of customer transactions in the database. When compared to other algorithms, partition algorithm has taken less execution time for finding the frequent itemsets, and association rule among the items, is shown in the table 7.1. The partition algorithm is faster than any other algorithm like AprioriAllHybrid and Apriori Algorithm.

Partition algorithm works very much faster for all datasets that ranges from giga bytes to tera bytes. The performance study shows that the proposed method has a better performance over the existing methods for massive datasets.