A STUDY ON THE CONTRIBUTION OF APRIORI ALGORITHM IN CUSTOMER BEHAVIOR PREDICTION

Sujith Baby¹, Shamanth Rai², Manish Poojary³ and Manimozhi R⁴

Abstract - Apriori algorithm is one among the most discussed algorithm among association rules. Even though there exists many algorithm for association rule, Apriori stands first because of it’s efficiency in mining association rules. In large databases, mining association rule takes lot of time, but Apriori algorithm can make it little more easier with it’s reduced number of data base scanning. Since the data mining methods have advantages and disadvantages it is important to find out which is the appropriate techniques to mine data bases. The objective of this paper is to know how suitable is Apriori algorithm for customer behavior prediction.

I. INTRODUCTION

To study how a customer wants to purchase certain product from a shop and how randomly they are buying the items, this is very interesting concept of study. To understand the psychological mind-set and converting it into statistical format so that it can be helpful in analyze buying behavior. These all things help an organization to improve their marketing strategies. The strategies such as how a person select between different choices, how a living environment affects his way of purchasing and how effective is a marketing campaign are analyze through different association rules.

II. ALGORITHM

A. Apriori algorithm –

All frequent itemset can be generated using Apriori Algorithm. The itemset whose support is greater than some user-specified minimum support is considered as frequent itemset. Candidate itemset is potentially frequent itemset. To make multiple passes over the database is the key idea of Apriori algorithm. An iterative approach breadth-first search is used through the search space, where k-itemsets are used to explore (k+1)-itemsets. At First, the set of frequent 1-itemsets is found. The set of this 1-itemsets contains one item, which satisfy the support threshold is denoted

¹ AIMIT, Mangaluru, Karnataka, India
² AIMIT, Mangaluru, Karnataka, India
³ AIMIT, Mangaluru, Karnataka, India
⁴ AIMIT, Mangaluru, Karnataka, India
by L1. We begin with a seed set of itemsets found to be large in the previous pass in each subsequent pass. This seed set is used for generating new potentially large itemsets. A new potentially large itemsets called candidate itemsets is generated using this seed set, and count the actual support for these candidate itemsets during the pass over the data. At the end of the pass, the candidate itemsets which are actually large (frequent) is determined, and they become the seed for the next pass. Therefore, first L1 is used to find L2, the set of frequent 2-itemsets, then L2 is used to find L3, and so on, until no more frequent k-itemsets can be found. Then, Apriori property is employed to reduce the search space, where the Apriori property is described as —“All nonempty subsets of a large itemset must also be large” or —“If a set is not large, then its superset can’t be large either”. This property belongs to a special category of properties called antimonotone which means that if a set cannot pass a test, all of its supersets will fail the same test as well.

1. **Pass 1**
   1. Generate the candidate itemsets in C1
   2. Save the frequent itemsets in L1

2. **Pass k**
   (i). Generate the candidate itemsets in Ck from the frequent itemsets in Lk-1
   Join Lk-1p with Lk-1q, as follows:
   insert into Ck
   select p.item1, p.item2, . . . , p.itemk-1, q.itemk-1
   from Lk-1p, Lk-1q
   where, p.item1 = q.item1, . . . p.itemk-2 = q.itemk-2, p.itemk-1<q.
   itemk-1
   • Generate all (k-1)-subsets from the candidate itemsets in Ck
   • Prune all candidate itemsets from Ck where, some (k-1)-subset of the candidate itemset is not in the frequent itemset Lk-1
   (ii). Scan the transaction database to determine the support for each candidate itemset in Ck
   (iii). Save the frequent itemsets in Lk.

**II. EXPERIMENT AND RESULT**

The test set for this evaluation experiment chess data randomly selected from the internet. NetBeans IDE 8.1 software platform is used to perform the experiment. The PC for experiment is equipped with an core i3 2.10GHz CPU with speed and 4GB RAM in personal computer.

From the simulation of the experiment results, we can draw to the conclusion that the algorithm consumes less time and minimum database scans to generate the results.
A chess data set has been used, which includes 76 items in it for analysis. A set of association rules have been obtained by applying A priori algorithm for it. During the analysis we found that A priori algorithm is efficient and consumed less time to generate the result. All these results are collected from Intel core i3 2.10GHz CPU with speed and 4GB RAM.

REFERENCES


[3] Bo Wu; Defu Zhang; Qihua Lan; Jiemin Zheng “An Efficient Frequent Patterns Mining Algorithm based on Apriori Algorithm and the FP-tree Structure” Xiamen 361005, China 2008.


[5] Agrawal R; Srikant R; Fast algorithms for mining association rules [A]; In; Proceedings of the 20th International Conference on Very Large Databases[C]; Santiago ; Chile ; 1994 487-499.

