Market Basket Analysis of Sports Store using Association Rules

Harpreet Kaur[1] and Kawaljeet Singh[2]
[1]Department of Computer Science, Punjabi University, Patiala, India.
[2]University Computer Centre, Punjabi University, Patiala, India.

ABSTRACT
This paper focuses on the market basket analysis of sports equipment, which is relatively less explored area in the literature. In the present investigation, market basket analysis of sports equipment would be done by collecting primary data directly from retail and wholesale sports basket. FP-Growth algorithm is an efficient algorithm for mining frequent patterns. In particular, we apply FP-Growth algorithm on the data collected directly from the sports vendors to trace various associations leading to a sports equipment basket. One particular advantage is that it circumvents the need for producing candidate sets, which is typically computationally intensive. The findings of this paper can be applied to optimize marketing and sale of sports equipment. In particular, knowing different dynamics of sports equipment basket, sports vender can optimize the display of sports equipment on their premises to maximize sales.

KEYWORDS: *Sports store, Market Basket Analysis, Marketing and sales strategies.*

I. INTRODUCTION

Business intelligence is a broad category of application and technology for gathering, storing, analyzing and providing access to data to help clients make better business decisions [1]. In this research, business intelligence will help the sports store how to increase profits, product recommendations, promotion and product placement by using tools from data mining.
Essential components of proactive business intelligence are as follows:

1) Real time data warehousing
2) Data Mining
3) Automatic learning and refinement
4) Data visualization and etc.

Data mining is the process of analyzing data from different perspectives and summarizing it into useful information- which can be used to increase profits for the sports store.
Some common uses of data mining are as follows:

1) Segmentation
2) Market Basket Analysis
3) Fraud Detection
4) Direct Marketing and etc.

Market Basket Analysis is a technique that assists us in determining which sports items are purchased together according to the association rules, primarily with the aim of identifying cross-selling opportunities. Its primary objective is to improve the effectiveness of marketing and sales tactics using the customer data that is accumulated with the enterprise during sales transaction [2]. Suppose a manager of a sports store wants to learn about the buying habit of the customers so that he/she can determine - "which groups or sets of items customers are likely to buy from the store. One plausible option to determine this is by performing market basket analysis on the retail history, i.e. the retail data collected from previous transactions. The main application of this analysis is in optimizing the store layout, where the items that are likely to be sold together can be placed closer to each other. For example, if tennis balls are placed in the same aisle as the tennis rackets, there is a higher probability that the customers buying tennis rackets will also buy tennis balls. Such planning can help to boost sales.
Each Data mining technique can perform following types of data modeling:

1) Association
2) Classification
3) Clustering
4) Regression
5) Sequence Discovery
6) Visualization [3]

Association Rules aims to discover items that co-occur frequently within a database. For example: a store manager may wish to know how likely a customer would buy tennis ball if he buys tennis

rackets. Such knowledge can help the store manager in products arrangement, shelf space planning and effective implementation of product promotion strategies. The items that co-occur frequently can be discovered from a customer transactions database using association rules. This paper represents each of the items with one of the two Boolean variables i.e. 0 and 1, where 0 represents the absence and 1 represents the presence. So each item can be represented in the basket by the Boolean vectors. The Boolean vectors then analyzed for buying patterns which reflect the items that are frequently associated with each other. These patterns can be represented in the form of association rules. Support, confidence and lift are the interestingness of this research. For example in the sports store, association rules among different sports equipments are well shown in the table 1.

II. TERMINOLOGY

Rule Body: One or more item in the transaction, which implies the presence of another.

Body → Head [support, confidence].

Rule Head: One item whose presence in the transaction is implied by the presence of items in the rule body. For e.g.

$X \rightarrow Y$.

Support: Percentage of all transaction containing both the body items and the head items.

Support (X->Y) = $P(X \cup Y)$

Confidence: That the head item will be in the transaction, given the presence of body items. The confidence of $X \rightarrow Y$ in a database D is the ratio of the number transaction containing $X \cup Y$ to the number of transactions that contain X.

$$\text{conf}(X \rightarrow Y) = \frac{\frac{\text{numTrans}(X \cup Y)}{|D|}}{\frac{\text{numTrans}(X)}{|D|}} = \frac{p(X \wedge Y)}{p(X)} = p(Y \mid X)$$

Lift: Degree to which the confidence is greater (or less) than expected (threshold). The Lift indicate the departure from independence of X and Y. The lift of $X \rightarrow Y$ is: [4]

$$\text{lift}(X \rightarrow Y) = \frac{\text{conf}(X \rightarrow Y)}{p(Y)} = \frac{\frac{p(X \wedge Y)}{p(X)}}{p(Y)} = \frac{p(X \wedge Y)}{p(X)p(Y)}$$

III. ALGORITHM

One of the popular algorithms for frequent itemset mining is the FP-Growth. It is based on a prefix tree representation of the given database of transactions (called FP-tree), which can save considerable amount of memory for storing the transactions. The basic idea of the FP-Growth algorithm can be described as recursive elimination scheme [9].

So, FP-Growth allows frequent itemset generation. This is only two step approach:

Step 1: Build a compact data structure called the FP-Tree

Step 2: Extracts frequent itemsets directly from FP-Tree [5].

FP-Tree construction

Input: A transaction database DB and a minimum support threshold.

Output: Its frequent pattern tree, FP-Tree.

Method: The FP-Tree is constructed in the following steps:

1. Scan the transaction database DB once. Collect the set of frequent items F and their supports. Sort F in support descending order as L, the list of frequent items.
2. Create the root of an FP-Tree, T, and label it as "null". For each transaction Trans in DB do the following.

   Select and sort the frequent items in Trans according to the order of L. Let the sorted frequent item list in Trans be [p|P], where p is the first element and P is the remaining list. Call insert_tree ([p|P],T).

   The function insert_tree ([p|P],T) is performed as follows:

   If T has a child N such that N.item-name = p.item-name, then increment N's count by 1; else create a new node N, and let its count be 1, its parent link be linked to T, and its node-Link be

linked to T, and its node-link be linked to the nodes with the same item-name via the node-link structure. If P is non empty, call insert_tree(P,N)recursively [10].

IV. EXPERIMENTS

Researcher now describes the experiments which were conducted in order to assess the practical feasibility of using samples for finding frequent sets. It present two new variants of algorithms using sampling and give experimental results. The objectives which were set by researcher are achieved as follows which will let the organization how to increase profits and how the sports industries will be promoted. By collecting data from various sports markets on various sports commonly played in India and it has been observed that some common items have been generated which were purchased frequently with many sports item. The sports which were discussed in this paper are cricket, hockey, badminton, table tennis and tennis.

Overall Association rules generated using sports equipments

Rules generated using min support and min confidence of 1 i.e. 100%

Table 1 Association rule matching for sports items

S.No.	Premises	Conclusion	Support	Confidence	Lift
1	Shuttle Cocks	Rackets	0.527	1	1.897
2	Rackets	Shuttle Cocks	0.527	1	1.897
3	Bat	Cball	0.412	1	2.242
4	VShoes	VBall	0.243	1	3.289
5	Tumbler, Cball	Towel	0.378	1	1.542
6	Tumbler, Bat	Towel	0.372	1	1.542
7	Tumbler, Small Bat	Towel	0.230	1	1.542
8	Towel, Shuttle Cocks	Rackets	0.311	1	1.897
9	Towel, Rackets	Shuttle Cocks	0.311	1	1.897
10	Towel, Bat	Cball	0.385	1	2.242
11	Towel, Small Bat	Cball	0.236	1	2.242
12	Tumbler, Shuttle Cocks	Rackets	0.318	1	1.897
13	Tumbler, Rackets	Shuttle Cocks	0.318	1	1.897
14	Tumbler, Bat	Cball	0.372	1	2.242
15	Tumbler, Small bat	Cball	0.230	1	2.242
16	Shuttle Cock, Back Pack	Rackets	0.257	1	1.897
17	Rackets, Back Pack	Shuttle Cocks	0.257	1	1.897
18	Bat, Small Bat	Cball	0.243	1	2.242
19	Playing Dress, VShoes	VBall	0.230	1	3.289
20	Towel, Tumbler, Shuttle Cocks	Rackets	0.297	1	1.897
21	Towel, Tumbler, Rackets	Shuttle Cocks	0.297	1	1.897
22	Tumbler, Bat	Towel, Cball	0.372	1	2.552
23	Towel, Tumbler, Bat	Cball	0.372	1	2.242
24	Tumbler, Cball, Bat	Towel	0.372	1	1.542
25	Tumbler, Small Bat	Towel, Cball	0.230	1	2.552

26	Towel, Tumbler, Small Bat	Cball	0.230	1	2.242
27	Tumbler, Cball, Small Bat	Towel	0.230	1	1.542
28	Towel, Bat, Small Bat	Cball	0.230	1	2.242

Market Basket Analysis is a critical tool to help improve the efficiency of marketing and sales strategies by analyzing customer's data collected from past sales transactions.

- From the above calculated association rules it is clear that which products should be kept on the shelves together which would increase the profits and the products which are frequently purchased with every sport items are common items like tumbler, towel, running shoes, socks and back pack.

Now following written is explained how above calculated association rules are helpful for the organization to improve themselves:

1) Product recommendations and promotion

Product promotion begins with offer. Product promotion is any initiative by an organization to promote an increase in sales. Products can be promoted by a) Buy one get one free. b) Customer relationship management. c) Free Gifts. d) Discount prices. e) Free Samples.
So the products which are recommended with every sport items are tumbler, towel, running shoes, socks and back pack. So these products should be promoted like by giving one tumbler as a gift with the tennis ball and tennis racket or by giving discount on tumbler or towel with the tennis ball and racket and this should be given to increase profit. These products will definitely help the retailers to increase profits.

2) Product placement

Plan your store layout visual merchandising displays. Layout of the store should be in such a manner that with every sport equipment there should be display of other items like towel, tumbler, running shoes, socks and skipping rope which will definitely attract more customers to the store.

3) Next best offer strategies

Next Best Offer is the use of predictive analytics solution to identify the products or services your customers are most likely to be interested in for their next purchase.
It follows these three steps:

1) **Predict** -> Apply predictive analysis to customer data to identify trends in purchasing behavior and product affinities.
2) **Act ->** Use product affinities to recommend other products or services customer may be interested in.
3) **Perform ->** Apply customer response to future targeting and messaging criteria to provide relevant content that addresses customers [8].

Next Best offers are targeted to the customers of products or services that they are likely to buy. Difficulty about sport store is that there are so many variables that company can employ to improve them. One of the important is what product attributes do you find appealing, and which of our products have them?
Discounts should be given on products with the sale of sports products which they find appealing like tumbler, towel, socks and etc.

V. CONCLUSION

In this research Frequent Pattern Growth algorithm is used, for efficient mining of frequent pattern in large database because there are several advantages of FP-Growth over other approaches. FP-Growth is applied on dataset and generated the association rules for the sports store which care for profit

maximization at wholesale and retailer level. The results generated above will help the sports store's how to increase the sales and earn profits.

REFERENCES

[1]. McCarthy Jonathan "Business Intelligence", UCC march 9, 2011.
[2]. Gouda Karam & J. Zaki Mohammed, "Efficiently Mining Maximal Frequent Itemsets", IEEE, 2001, pp.163-170.
[3]. Ngai E.W.T., Xiu Li, Chau D.C.K. "Application of Data Mining Technique in customer relationship Management: A literature review and classification", Expert Systems with Application 36(2009) p.p. 2592-2602.
[4]. Smirnov Evgueni "Association rules."
[5]. Yang Don-Lin, Pan Ching-Ting and Chung Yeh-Ching, "An Efficient Hash based Method for discovering the Maximal Frequent Set".
[6]. Gouda Karam & J. Zaki Mohammed,"Efficiently Mining Maximal Frequent Itemsets", IEEE, 2001, pp.163-170.
[7]. Cavique Luis, "Next-Item Discovery in the Market Basket Analysis" IEEE 2005, pp.198-199.
[8]. www.angoss.com
[9]. Borgelt Christian "An implementation of the FP-Growth Algorithm" Otto-von-Guericke-University of Magdeburg.
[10]. Jiwan Han and Micheline Kamber, Data mining, concepts and techniques, Morgan Kaufman 2009.

AUTHORS

Harpreet kaur is a student in Department of Computer Science, Punjabi University, Patiala.
Kawaljeet Singh is working as Director in University Computer Centre, Punjabi University, Patiala.