

Evaluating Techniques for Mining Customer Purchase Behavior and Product Recommendation- A Survey

Ashmin Kaul
Student
Comp Science Department
NM University, Mumbai.

Mansi Virani
Student
Comp Science Department
NM University, Mumbai.

Chaitanya Kaul
Student
Comp Science Department
NM University, Mumbai.

Teja Gummalla
Student
Comp Science Department
NM University, Mumbai.

ABSTRACT

Companies these days rely strongly upon their previous data history for predicting future trends in their business operations and strategies. Whether forecasting inventory or estimating sales, data mining has emerged as a new vibrant tool for extracting potential knowledge hidden inside the purchase behavior of the customer. Mining has made it possible to span over a large dataset in no time and come up with useful prediction. Building on the notion that customer's purchase pattern plays a vital role in planning future strategies, in this paper, we focus upon several mining techniques used for understanding the customer's purchase behavior. We analyze the purchase behavior from three different mining aspects i.e. classification, association rules, clustering technique and compare their accuracies. Thereafter, the study builds upon a scope whereby these techniques can be used to achieve a much filtered and unique dataset for analysis and also can be used for product recommendation for new customers. Finally, the study lays the scope of getting these techniques to be in sync with techniques used in forecasting such that companies can plan their sales and conduct the inventory management effectively.

General Terms

Mining of customer purchase behavior and recommendation of products.

Keywords

Data mining, classification, association rules, clustering, purchase behavior.

1. INTRODUCTION

Customer's purchasing behavior varies according to various key factors like the seasonality, salary index, mode of shopping, etc. There are high contrasting differences observed in customer mindset while purchasing online or through offline mode. The behavioral mining paradigm basically begins with classification of the customer into categorical labels. These categorical labels are built using a classifier. The database consisting of tuple entries and their class names constitute the classifier. Based upon certain classification rules or constraints imposed, one can get a much refined and a filtered dataset to perform the mining operation. Once the classified data is achieved, we may work towards finding some useful knowledge in terms of association patterns or

prediction based on decisions. The association rules are used in relational databases to discover linkages between unrelated data using some specified conditions [4]. E.g. If the customer purchases milk, there is a high possibility that he will also purchase bread or biscuits, a concept called the market-basket analysis. By analyzing this kind of data, companies are able to perform various key business operations like direct marketing, sales promotion, etc. However, a much accurate and effective business operation can be achieved using a large dataset of homogeneous data elements i.e. segmentation of data such that all the elements of a data segment have similarity between each other. Two novel methods to achieve this are Clustering and Sequential Pattern mining. Hence, in the subsequent sections, the focus is laid upon understanding these methods with the help of certain real time system techniques. Firstly, in section II, the discussion is held with respect to the classification phase whereby we lay focus specifically on the decision tree and naive bayes approach to perform analysis and study different classes of customers formed based on the attributes of their purchase. This is followed by the association rules to find new linkages within the unrelated data. The study also focuses on the sequential patterns and the clustering approach to understand the procurement of a much refined unique dataset for analyses. Section III discusses the observations inferred from the study whereby we highlight the merits and demerits of these techniques. Finally, in section IV, we lay the conclusion and scope for the future work.

2. LITERATURE REVIEW

2.1 Classification

The classification technique is a novel approach used in data mining to create target classes and assign these classes, items from a heterogeneous collection of data. In the customer relationship management scenario, companies segment or classify their customers based upon numerous client profile attributes like property status, needs, preferences, etc. and design specific strategies and business models for recommending products and services to potential clients [1]. This section highlights the summary of analyzing the customer dataset using the decision tree and naive bayes algorithms and specific results are depicted:

2.1.1 Decision Tree

The classification method was applied on a dataset comprising of 969,228 customers for predicting the cross selling opportunity of a new service called WAP [2]. Firstly, the customers are segregated into two categories i.e. WAP and Non-WAP users [2]. Then, the decision tree algorithms is applied to understand the characteristics possessed by the WAP users and check for the customers possessing the same characteristics in the 'Non WAP' category. By performing analysis using the SAS 9.1, software, the following results were obtained as displayed in tabular form:

Input	Analysis Result
Consuming expense level	1. WAP users increase as expense level increases.
Age level	2. WAP users are between 18 and 50 years old.
Brand type	3. WAP users' proportion in Gotone and M-zone is higher than in other brands.
Roaming feature	4. Users roaming frequently are have higher WAP usage than others.
Customer type	5. New customers are not frequent WAP users.

Fig 1: Result using Decision Tree

2.1.2 Naive Bayesian Classifier:

The Naive Bayesian Classifier is an effective method preferred to categorize text, it has a fairly easy procedure of estimating parameters iteratively which makes it desirable for very large datasets [3]. In the recommendation system context, the Naive Bayes classifier approach is used to recommend highest ranking books to customers. Firstly, $P(h)$ is the prior probability which means that it is the probability of some hypothesis 'h' to be true. Now, $P(h/D)$ uses the predictor 'D' to support the hypothesis. The classifier assumes that effect of 'D' on 'h' is independent of all the other predictors. This is known as the class conditional independence. Consider the equation:

$$P(h/D) = P(D/h) * P(h) / P(D) \quad (1)$$

$P(h/D)$ is the posterior probability where h is the class and D is the given predictor.

$P(h)$ is the prior probability, h represents the class.

$P(D)$ is the prior probability where D represents the predictor. Let the total number of hypothesis be $h_1; h_2; h_3; \dots; h_n$, then:

$$P(h_1/D) = P(D/h_1) * P(h_1) / P(D) \quad (2)$$

Computing $P(h_2/D) \dots P(h_n/D)$, h_{map} i.e. hypotheses with maximum probability is chosen for the given data [3]. The given formula describes the calculation of h_{map} :

$$h_{map} = \arg - \max_{h \in H} P(h/D) \quad (3)$$

'H' is the set of hypothesis. For each hypotheses in 'H', we have:

$$h_{map} = \arg - \max_{h \in H} P(d/h) * P(h) / P(d) \quad (4)$$

Clearly, $P(D)$ remains constant for all hypotheses calculations. In order to calculate the most likely hypotheses, we can even eliminate $P(D)$ from the above equation (4). Hence, for recommending products like books to the customer that visits the website, the recommendation system uses this approach without no prior user profile data available [3].

2.2 Association Rules

Association rule mining is based on the notion that the association linkages between two related data entities can be useful in predicting possible linkages between any two unrelated data entities. In this section we highlight certain techniques followed which make use of associations to create new possible linkages.

2.2.1 Bipartite Network Representation

The study describes the association mining from the link prediction perspective. The link prediction scheme involves mining user item associations that are transitive and involve the customer's purchase behavior mapped onto a bipartite network as described below [4]. While performing a retail transaction, the customer entity and the product entity are selected as nodes and the links are created between them representing the relationship attribute. A number of customers and products are selected and two particular sets are formed:

Customer Set $C = \{c_1, c_2, \dots, c_n\}$

Product Set $P = \{p_1, p_2, \dots, p_n\}$

Based upon this data, a bipartite network is created which predicts future links based on the current linkages. These linkages depend on two conditions an association link between customer (c_i) and product (p_j) exists i.e. $a_{ij} = 1$, if purchase is made, else $a_{ij} = 0$ i.e. no purchase recorded. [4]

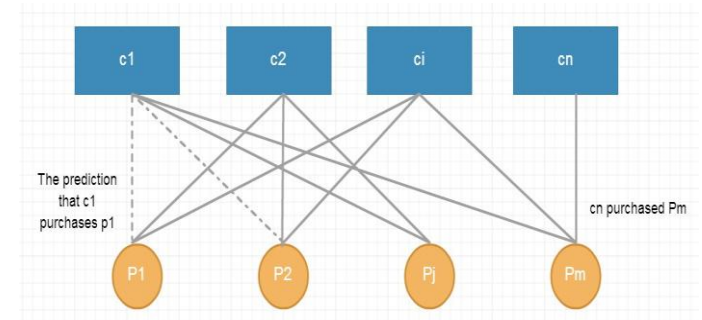


Fig 2: Link prediction in customer-product network

The study involves the use of Jaccard's coefficient in the bipartite network where the number of common nodes belonging to $f(x)$ and $f(y)$ in the customer-product network are compared with the activeness of the customer in purchasing the product [6]. The coefficient is defined as:

$$\text{Coefficient} = (r(x) \cap r(y)) / ((r(x) \cup r(y)) \quad (5)$$

The numerator denotes the common neighbors of x and y nodes. The objective behind using the jaccard's coefficient is to reduce the possibility of recommending highly popular products by using the denominator as a penalty factor,

however the most sought after products can be obtained by assigning the results in descending order of popularity. For recommendation purpose, the top-N recommendation procedure was followed. For the top 10 recommendations, standard metrics like the precision, recall and F-score were used [4]. Finally, the dataset used covered a total of 10281 customer transaction from the past two years spanning 1560 products [4].

2.2.2 Carrefour case

The Carrefour framework uses a storage database [5] and through implementation of association rules, studies the customer's perception towards the brand, the purchase experience to evaluate product and brand attributes [5] specifically. The procurement of data to apply association rules was done in this way:

1. A questionnaire was designed which included the company customers with their purchase details.
2. Product categories and brands were used to categorize the products purchased by the customer.
3. To analyze the product in more detail, three highly sought after products were chosen from each category [5].
4. Interviews with customers were held at different working hours of the store and during different days in a week in order to get a random set of data.

A total of 341 samples was selected to be valid from the total 363 samples selected [5]. The association rules used are defined as follows:

1. An itemset 'I' is made in which we have an array of items given by $(i_1, i_2, i_3, \dots, i_n)$, each item being a specific literal [5].

2. 'D' is a set of 'T' transactions made in the database, such that $T \subseteq D$ [5].

3. Association rules observe the rule $(X \rightarrow Y)$ [5], whereby this rule is applicable in database D based on two key factors:

- Support, denoted here by $\text{Sup}(X, D)$ [5] represents the rate of transactions in D containing the item set X [5], the importance of D depends upon the value of X, higher value denotes more importance.[5]
- Confidence, denoted by $\text{Conf}(X \rightarrow Y)$ represents that if a transaction made contains X, then the possibility of Y being present is relatively high [5]. Confidence is given by formula:

$$f(X \rightarrow Y) = \text{Sup}(X \cap Y) / \text{Sup}(X, D) \quad (6)$$

The objective behind using association rules is to generate all possible associations i.e. transactions having a pre-defined minimum support and confidence denoted by Minsup and Minconf respectively. In this scenario, association rule mining is done in a 2- stage process. The first stage involves creating a itemset whose support is greater than Minsup, subsequently the association rules are applied which depend on the following conditions [5].

$$\text{Sup}(X \cup Y, D) \geq \text{Minsup} \quad [5]$$

$$\text{Conf}(X \rightarrow Y) \geq \text{Minconf} \quad [5]$$

2.3 Clustering

In this technique, the entire focus is laid upon finding homogeneous groups of data values which may aid the analysis process. By performing clustering, the analyst can also chalk out certain data entries which do not match the homogeneity of any group and hence are termed as outliers. The PBS technique pertaining to clustering the customers is observed in this paper, it is discussed below:

2.3.1 PBS algorithm

This is a purchase-based segmentation algorithm which is used for conducting an efficient market segmentation. Firstly, the total number of customer clusters are selected and labelled as 'K'. After selecting the clusters, the initial cluster centers are obtained from the T^c records [8]. T^c is the cumulative transaction database which consists of purchase attributes like customer ID, transaction time, product item, quantity, expense [8]. Now, 'G' i.e. a set of K cluster centers and given by the following equation: $G = \{c^n | n = 1, \dots, K\}$, It is taken where c^n is the n^{th} cluster's center [8], $c^n \in T^c$. Hence, $(T^c - G) = \{c_i | i = 1, \dots, ||T^c - G||\}$ represents the set of remaining customers which are not the part of cluster centers [8]. Here, $c_i \neq G$. The similarity of the customers selected as cluster centers c_n and the remaining customers c_i is calculated as $\text{Sim}(c^n, c_i)$ [8]. Based upon maximum similarity between c_i and c^n , c_i is chosen and assigned to the n^{th} cluster G_n [8]. The expression describes the similarity measure between c_i and c^n [8]

$$\max_{1 \leq n \leq K} \{\text{Sim}(c_i, c^n)\} \text{ where } c_i \in (T^c - G) \quad (7)$$

Then we have the Pio equation:

$$\text{Pio}(c_i) = \sum_{c_j \in G, m \neq n} \text{Sim}(c_i, c_j) / \sum_{c_m \in G, m \neq n} \text{Sim}(c_i, c^m) \quad (8)$$

c_m represents the center of the cluster G_m [8]. In the G_n set, a customer with the highest priority is chosen as the new center [8]. The process can be denoted as:

$$c^n = \arg \max_{c_i \in G_n} \{\text{pio}(c_i)\} \quad (9)$$

After calculating the new cluster centers, the algorithm accomplishes one complete iteration and then repeats until no more change occurs in the cluster centers.

2.3.1.1 Quality Function:

After determining the clustering result, it's quality is determined and is given by a set equations:

$$p(k) = 1/K \sum_{n=1}^k \left(\min_{1 \leq m \leq K, m \neq n} \{(\eta_n + \eta_m) / \delta_{nm}\} \right) \quad (10)$$

$$\eta_n = 1/||G^n|| * (\sum_{c_i \in G_n} \text{Sim}(c_i, c^n)) \quad (11)$$

$$\eta_m = 1/||G^m|| * (\sum_{c_j \in G_m} \text{Sim}(c_j, c^m)) \quad (12)$$

$$\delta_{nm} = \text{Sim}(c^n, c^m) \quad (13)$$

From equation [11], η_n is the average of similarities between c^n and the other customers in G_n [8]. Subsequently, from equation 'c', η_m represents the average of similarities between c_m and rest of the customers in cluster G_m . From equation [13], δ_{nm} denotes the similarity relation between c^n and c_m . Finally, we perform the segmentation process, an optimal value for K is calculated whereby K ranges between lower boundary value 's' and the higher boundary value 't' [13].

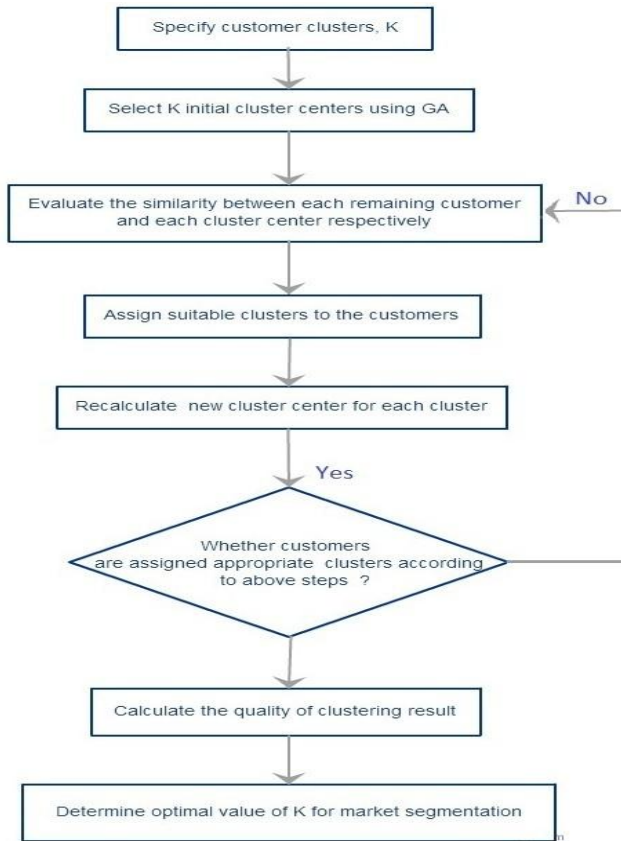


Fig 3: PBS algorithm framework

3. INFERENCE

This section involves evaluating the various methods discussed in the previous section with respect to their performance and making vital comparisons amongst them. Beginning with classification, decision tree and naive bayes classifier methods were studied. For analyzing the cross-selling opportunity for WAP, 21,150 user dataset was taken for which forecasting was performed [2]. The forecast result showed 9,517 WAP users [2], from which there were 5,234 actual WAP users [2], with an accuracy of 55 percent. However using association rule mining, 67,154 users were predicted, the result showed that WAP service would have 24,856 users as target customers [2], while in actual, WAP had 15,657 target customers [2]. The accuracy using this method came out to be 62.99 percent [2]. The book recommendation system used a test crawler to study all the reviews posted by buyers, all the negative and positive adjacent keywords were chosen in order to predict popularity of the books. The training data was provided by amazon.com [3]. Based upon any latest review updated, the system calculates the review sense. Also the category of reviews i.e. $f(\text{Review Sense} = X | \text{positive})$ and $f(\text{Review Sense} = X | \text{negative})$ are found out using Naive Bayes formula for the numerical data.[3] In the link prediction recommendation algorithm, the link prediction technique outperforms the collaborative filtering method in enhancing the recommendation quality considerably[4]. The method not only focuses on the topology of the structure but also takes domain knowledge into account. The algorithm recommends products to customers who performed more than 50 transactions [4]. After the pro-processing phase, a bipartite network was achieved which included 970 customer notes, 1560 product notes and 86692 links [4]. Notes are nothing but the product domain knowledge in the bipartite network [4].

The following table denotes the algorithm performance measures. The experiments denote that notes attribute enhances the quality of the recommendations list. Also, the product category strongly influences the purchase attribute [4].

Algorithm	Precision	Recall	F Measure
B_Jaccard's Coefficient	0.01116	0.0134	0.0124
DB_Jaccard's Coefficient	0.01103	0.0318	0.0163

Fig. 5. Algorithm Performance Measures

The PBS algorithm discussed above aimed at generating clusters of customers with similar purchase patterns. During validation, it was claimed that the pattern could also have been calculated by observing each item's frequency in one cluster [6]. The frequency at which an item in the item set is purchased is called 'support'. A comparative experiment was done to note the performance of the PBS algorithm against the K-Means algorithm whereby the number of clusters was set at 30[6]. For every cluster, average of all the supports was taken, the average measure of the top five supports and the support deviation were evaluated subsequently [6]. To check differences observed amongst these two algorithms with respect to the given three aspects, one-way Annova tests were adopted [6]. The results obtained showed that PBS algorithm generates clusters of customers with similar purchase behavior as the average of top five highest supports was much higher than that of the K-means. Also, PBS outperforms K-means in the standard support deviation calculation, as it is much higher for PBS [6]. PBS algorithm finally terminates in a few learning iterations [6].

4. CONCLUSION

In this paper, we have studied and observed the customer with respect to mining the purchase behavior and discussed several related attributes pertaining to this, e.g. Time of purchase, quantity purchased, etc. The goal of the study was to draw a comparative analysis of these mining techniques such that understanding the patterns hidden in the sales data could aid the product recommendation process in the future. The association rule mining proved out to be more efficient than the decision tree algorithm in terms of accuracy. Hence, the future work may involve creating a sync relation between these techniques discussed and the techniques used for sales forecasting e.g.: Delphi technique, Exponential smoothing, Regression analysis, etc. Several other mining techniques like sequential patterns, would also be considered for analysis and implementation. The use of big sales data and map reduce framework would also be a given a thought for finding the evolving trends in the data. Finally, the approach may also involve understanding the customer from the ethnicity point of view to understand the difference in the mindset of people across countries which could improve the cross-border sales of products.

5. REFERENCES

- [1] H. Gong and Q. Xia, "Study on application of customer segmentation based on data mining technology," in 2009 ETP International Conference on Future Computer and Communication, 2009, pp. 167–170.
- [2] X.-c. Yang, J. Wu, X.-h. Zhang, and T. jie Lu, "Using decision tree and association rules to predict cross selling opportunities," in Machine Learning and Cybernetics, 2008 International Conference on, vol. 3. IEEE, 2008, pp. 1807–1811.

- [3] A. S. Tewari, T. S. Ansari, and A. G. Barman, "Opinion based book recommendation using naive bayes classifier," in *Contemporary Computing and Informatics (IC3I)*, 2014 International Conference on. IEEE, 2014, pp. 139–144.
- [4] J. Li, L. Zhang, F. Meng, and F. Li, "Recommendation algorithm based on link prediction and domain knowledge in retail transactions," *Procedia Computer Science*, vol. 31, pp. 875–881, 2014.
- [5] S.-h. Liao, Y.-j. Chen, and Y.-t. Lin, "Mining customer knowledge to implement online shopping and home delivery for hypermarkets," *Expert Systems with Applications*, vol. 38, no. 4, pp. 3982–3991, 2011.
- [6] P. Giudici and G. Passerone, "Data mining of association structures to model consumer behaviour," *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 533–541, 2002.
- [7] B. Xiao, E. Aïmeur, and J. M. Fernandez, "Pfinder: An intelligent product recommendation agent for e-commerce," in *CEC*, 2003, p. 181.
- [8] C.-Y. Tsai and C.-C. Chiu, "A purchase-based market segmentation methodology," *Expert Systems with Applications*, vol. 27, no. 2, pp. 265–276, 2004.
- [9] X. Wang, J. Wang, T. Wang, H. Li, and D. Yang, "Parallel sequential pattern mining by transaction decomposition," in *Fuzzy Systems and Knowledge Discovery (FSKD)*, 2010 Seventh International Conference on, vol. 4. IEEE, 2010, pp. 1746–1750.
- [10] C. Wang and Y. Wang, "Discovering consumer's behavior changes based on purchase sequences," in *Fuzzy Systems and Knowledge Discovery (FSKD)*, 2012 9th International Conference on. IEEE, 2012, pp. 642–645.
- [11] Y. Zuo, A. S. Ali, and K. Yada, "Consumer purchasing behaviour extraction using statistical learning theory," *Procedia Computer Science*, vol. 35, pp. 1464–1473, 2014.
- [12] A. Sato, T. Tamura, R. Huang, J. Ma, and N. Y. Yen, "Smart business services via consumer purchasing behaviour modeling," in *Green Computing and Communications (GreenCom)*, 2013 IEEE and Internet of Things (iThings/CPSCoM), IEEE International Conference on and IEEE Cyber, Physical and Social Computing. IEEE, 2013, pp. 812–817.
- [13] T. Hong and E. Kim, "Segmenting customers in online stores based on factors that affect the customers intention to purchase," *Expert Systems with Applications*, vol. 39, no. 2, pp. 2127–2131, 2012.
- [14] S.-H. Liao, C.-M. Chen, and C.-H. Wu, "Mining customer knowledge for product line and brand extension in retailing," *Expert systems with Applications*, vol. 34, no. 3, pp. 1763–1776, 2008.
- [15] P. K. Bala, "Decision tree based demand forecasts for improving inventory performance," in *Industrial Engineering and Engineering Management (IEEM)*, 2010 IEEE International Conference on. IEEE, 2010, pp. 1926–1930.
- [16] K. Sun and F. Bai, "Mining weighted association rules without preassigned weights," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 20, no. 4, pp. 489–495, 2008.
- [17] E. T. Apeh, B. Gabrys, and A. Schierz, "Customer profile classification using transactional data," in *Nature and Biologically Inspired Computing (NaBIC)*, 2011 Third World Congress on. IEEE, 2011, pp. 37–43.
- [18] G. Linden, B. Smith, and J. York, "Amazon. com recommendations: Item-to-item collaborative filtering," *Internet Computing, IEEE*, vol. 7, no. 1, pp. 76–80, 2003.
- [19] C. C. Aggarwal and P. S. Yu, "Outlier detection for high dimensional data," in *ACM Sigmod Record*, vol. 30, no. 2. ACM, 2001, pp. 37–46.
- [20] Y. Chen, S. Alsbaugh, and R. Katz, "Interactive analytical processing in big data systems: A cross-industry study of mapreduce workloads," *Proceedings of the VLDB Endowment*, vol. 5, no. 12, pp. 1802–1813, 2012.
- [21] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," in *ACM SIGMOD Record*, vol. 22, no. 2. ACM, 1993, pp. 207–216.
- [22] K.-W. Cheung, J. T. Kwok, M. H. Law, and K.-C. Tsui, "Mining customer product ratings for personalized marketing," *Decision Support Systems*, vol. 35, no. 2, pp. 231–243, 2003.