NOVEL METHODS FOR PREDICTION OF CUSTOMER PURCHASE PROBABILITY

[1] Ashraf Ali Shaik, [2] P Dhana Lakshmi
[1] Assistant Professor, [2] Assistant Professor
[1 2] Department of Computer Science and Systems Engineering
[1 2] Sree Vidyanikethan Engineering College, A Rangampet,Tirupathi, Andhra Pradesh.

Abstract— Web is a huge repository for storing different types of information. As the number of webpage's and websites has increased rapidly, understanding and discovering web user's behavior are essential for the web monitoring and recommendation systems. user's behavior in the web can be predicted using the online databases such as log files and prediction models. The main problems of user's behavior analysis over online databases are: lot of transactions between user and the web and the volume of data which are not completely structured. Knowing the customer behavior for online product recommendation and prediction of customers purchase probability using click stream data and tracking of customer behavior is essential part of online recommendation system. This paper presents few efficient approaches that works using Bayesian belief networks (BBN) and Nearest-neighbor collaborative filtering that calculate the probabilities of inter-dependent events and provides a successful means of generating recommendations for web users.

IndexTerms— Purchase probability, Click stream data, Bayesian Network, Online Recommendation System

I. INTRODUCTION

Recommendation system is a key component for evaluating the customer behavior in online databases. The key component of a recommendation system is data[1,3]. This data may be collected from customer ratings of products, feedback/reviews from purchasers, etc. The collected data may be a source for recommendations to users. After data collection, recommendation systems use any of machine-learning algorithms like clustering or classification algorithms for finding similarities between products or web pages used by the user. Recommender logic programs uses this information as suggestions for target or newly arrived customer based on user profiles[2,5,6]. There are many factors which may influence a person in making these choices, and modeling of any one of these factor is possible in a recommendation system. Due to development of internet, it is possible to overcome time and space constraints a new concept called e-commerce which argued the importance of immediate response about customers needs[4,7]. Because of this online recommendation system is more focused to maintain a track of customers behavior [8,10]. There are many recommendation systems which are existing and focus mainly on previous data, but in online recommendation system focused on understanding customer intention to purchase through real time activity data at online storefronts as the part of online recommendation systems[9]. This paper outlines the implementation of Bayesian Belief Network to increase the product sales. This paper also proposes a recommender system for and prediction of customers purchase probability using click stream data based on collaborative filtering.

II. EXISTING MODELS FOR PURCHASE PROBABILITY PREDICTION:

State model-1

The purchase probability of customer is predicted as the average purchase probability of page typed where she/he has visited.

State model-2

The purchase probability of state transition is calculated from click stream data, and purchase probability of customer is predicted as average purchase probability of last state transition.

Example:

Let us consider two manufacturing companies A and B and their profit ranges from past 7 years.

Years	1	2	3	4	5	6	7
Profit of product 'A'	95%	90%	85%	80%	30%	15%	5%
Profit of product 'B'	5%	15%	20%	25%	95%	95%	98%

Product to be suggested for year 8=max(average profit(A),average profit(B))
Average profit(A)=57.14
Average profit(B)=50.42
Hence, Product suggested for year 8=product A
In this method average is the only parameter considered and there is no focus on future trend.

III. PROPOSED METHOD

Bayesian belief network is a tool for predicting customer purchasing probability for online database. Consider a customer purchasing any product using card or cash. Let P(X) denotes the customer purchasing a product using card and P(Y) denotes the customer purchasing a product using cash.

According to Bayes theorem $P\left(\frac{X}{Y}\right) = \frac{P\left(\frac{Y}{X}\right) * P(X)}{P(Y)}$

X:

True	False
P(X)=0.02	P(~X)=0.98

Y:

True	False
P(Y)=0.04	P(~Y)=0.96

X	True		False	
Y	True	False	True	False
True	P(Z/XY)=0.2	P(Z/X~Y)=0.4	P(Z/~XY)=0.5	P(Z/~X~Y)=0.5
False	P(~Z/XY)=0.8	P(~Z/X~Y)=0.6	P(~Z/~XY)=0.5	P(~Z/~X~Y)=0.5

Using known probabilities we may calculate the initialized probability of Z by summing different combinations of Z for which Z is true.

P(Z) = P(XYZ)+P(Z~XY)+P(ZX~Y)+P(Z~X~Y)

= P(Z/XY)*P(X)*P(Y)+(P(Z/~XY)*P(~X)*P(Y))+P(Z/X~Y)*P(X)*P(~Y)+P(Z/~X~Y)*P(~X)*P(~Y)

= 0.2*0.02*0.04+0.5*0.98*0.04+0.6*0.02+0.5*0.96*0.98

= 0.5

Similarly P(Y/Z) = 0.147

P(X/Z) = 0.07

So we could say that given Z is true, Y is more likely to causing X.

In the **second proposed** method we are introducing three important terms namely **total, remain, count** for predicting the customer behavior**.**

Total:

It includes the total number of items brought by a particular retail organization in a specific period of time.

Remain:

It includes the total loss occurred in terms of number of products brought to a retail organization in a certain period of time.

Count:

It includes the number of times the particular product has occurred in a transaction database and it indicates the support of that product.

IV. ALGORITHM

Step 1: Load data set, remaining, total values for every item.
Step 2: Calculate remaining/total score for each item.
Step 3: Enter minimum support, minimum confidence and minimum remain/total.
Step 4: Formulate the table with items which score is greater than minimum remaining/total score.
Step 5: Generate item sets which satisfies minimum support.
Step 6: Generate association rules which satisfies minimum confidence.
Step 7: suggest the item for which rule it gives high confidence.

Example:

TID	**ITEMS**
100	Beef, chicken, milk
200	Beef, cheese
300	Cheese, shoes
400	Beef, chicken, cheese
500	Beef, cheese, chicken, clothes, milk
600	Chicken, clothes, milk
700	Chicken, milk, clothes, shoes

Items	**Total**	**Remain**	**Count**
Beef	100	40	4
Chicken	80	20	5
Milk	40	10	4
Cheese	100	20	3
Shoes	60	40	2
Clothes	50	5	3

Calculations:

To calculate the value of REMAIN/TOTAL

Step 1:

Beef = 4/10 = 0.4

Chicken = 2/8 = 0.25

Milk = 10/40 = 0.25

Cheese = 20/100 = 0.2
Shoes = 40/60 =0.66
Clothes = 5/50 =0.1
Average of all=0.3 i.e the remain/total value = 0.3

Products that are less than remain/total in the given transaction database is chicken, milk, cheese, clothes.

Step 2:

Now generating rules

1st iteration:

A=chicken	5
B=milk	4
C=cheese	3
D=clothes	3

Let the minimum support =3

2nd iteration:

AB	4
AC	2
AD	3
BC	1
BD	3
CD	1

AB	3
BD	3
AD	3

3rd iteration:

ABD	3

The association rule generated is A and B then D.

V. CONCLUSION

In this paper we developed an efficient method for customer's purchase probability which is calculated by average purchase probabilities of web page types which she/he had visited. Internet bookstore click stream data considering state transition shows better performances than those of considering state probability and also weighted averaging method and bayesian networks shows better performance than without considering weights.

VI. REFERENCES

[1] Thakur, S. S., Kundu, A., Sing, J. K., A Collaborative Filtering Approach: Based on offline customers ratings for Prediction and Recommendation, published in 1st International Conference on Scientific Paradigm Shift in Information Technology & Management,IEMCOM – 2011, 5th – 6th January, 2011, Kolkata.

[2] Niu, L., Yan, X.W., Zhang, C. Q., Zhang, S.C., Product Hierarchy-Based Customer Profiles for Electronic Commerce Recommendation. 1st International Conference on Machine Learning and Cybernatics, Beijing, November, 2002.

[3] Thakur, S. S., Kundu, A., Sing, J. K., Bayesian Belief Networks – based Product Prediction for E Commerce Recommendation, accepted for publication in CNC 2011, Bangalore, India.

[4] Aggarwal, C.C., Wolf. J. L., and Yu. P.S., A new method for similarity indexing for market data. In Proceedings of the 1999 ACM SIGMOD
Conference, Philadelphia, PA June, 1999.

[5] B. Prasad, 2007, "A knowledge-based product recommendation system for e-commerce," International Journal of Intelligent Information and Database Systems, Volume 1, No. 1, ISSN:1751-5858, Inderscience Publishers.

[6] D. Paraschakis, B. J. Nilsson and J. Hollander, "Comparative Evaluation of Top-N Recommenders in E-Commerce: an Industrial Perspective", "IEEE 14th International Conference on Machine Learning and Applications", pp. 1024-1031, 2015.

[7] Li, X., & Chen, H. Recommendation as link prediction in bipartite graphs: A graph kernel-based machine learning approach. Decision Support Systems, 54(2), 880–890,2013.

[8] B. Shams and S. Haratizadeh, "Graph-based collaborative ranking", Expert Systems with Applications, vol. 67, pp. 59–70, 2017.

[9] J. Wang, S. Robertson, A. P. de Vries, M. J. T. Reinders, Probabilistic relevance ranking for collaborative filtering, Information Retrieval 11 (6), 477–497,2008

[10] B. Long, J. Bian, A. Dong and Y. Chang, "Enhancing Product Search by Best-Selling Prediction in E-Commerce", Proceedings of the 21st ACM international conference on Information and knowledge management - CIKM '12 ,pp.2479-2482, 2012.