Enhancing Recommendation on Customizable E-Commerce System with Association & Ranking

S. Saravanakumar[1], S. Diviya[2]

[1]Research Supervisor, Assistant Professor, Department of Information technology
Sri Jayendra Saraswathy Maha Vidyalaya College of Arts and Science, Coimbatore-5, India

[2]Research Scholar, Sri Jayendra Saraswathy Maha Vidyalaya College of Arts and Science, Coimbatore-5, India

Abstract: ***The propose system is a web recommender that models user habits and behaviors by constructing a knowledge base using temporal web access patterns as input. The rapid growth of e-commerce has caused product overload where the customer is no longer able to effectively choose the products which is exposed to. The proposed system introduces a personalized recommendation procedure by which it can get further recommendation effectiveness when applied to Internet shopping malls with respect to the financial status of the customer. The suggested procedure is based on Web usage mining of relationship between the customer and product as well as his income status and from the most sold out product using association rule mining and Fuzzy C Means. Fuzzy logic is applied to represent real-life temporal concepts and requested resources of periodic pattern-based web access activities.***

Keywords: web mining, Market Basket analysis, personalized product recommendation, Association rules mining, Support count

1. Introduction

Web mining essentially has many advantages which makes this technology attractive to corporations including the government agencies. This technology has enabled E-commerce to do personalized marketing, which eventually results in higher trade volumes. The government agencies are using this technology to classify threats and fight against terrorism. The predicting capability of the mining application can benefit the society by identifying criminal activities. The companies can establish better customer relationship by giving them exactly what they need. Companies can understand the needs of the customer better and they can react to customer needs faster. The companies can find, attract and retain customers; they can save on production costs by utilizing the acquired insight of customer requirements. They can increase profitability by target pricing based on the profiles created. They can even find the customer who might default to a competitor the company will try to retain the customer by providing promotional offers to the specific customer, thus reducing the risk of losing customers.

A typical approach to customer segmentation is a distance-measure-based approach where a certain proximity or similarity measure is used to cluster customers. It considers the statistics-based approach, a subclass of the distance-measure-based approaches to segmentation that computes the set of summary statistics from customer's demographic and transactional data, such as the average time it takes the customer to browse the Web page describing a product, maximal and minimal times taken to buy an online product, regency, frequency, and monetary statistics and so forth. These customer summary statistics constitute statistical reductions of the customer transactional data across the transactional variables. These variables are typically used for clustering in the statistics-based approach instead of the raw transactional data since the unit of analysis in forming customer segments is the customer, not his or her individual transactions. After such statistics are computed for each customer, the customer base is then partitioned into customer segments by using various clustering methods on the space of the computed statistics. To describe an implementation of the well-known Association rule mining algorithm for association rules mining. Association Rules Mining is an important data mining model studied extensively by the database and data mining community. Initially for Market Basket Analysis need to find how many items purchased by customers are related to each other. Association rule mining algorithm is a simple and effective for predicting the association of the consumer recommendations and their profiles about the hidden needs based on their income status. To propose a new recommendations models for e-commerce based on the financial status and the hidden needs as well as most sold out products, which have recommendation procedure, Association rule mining and FCM. Decision Behavior Model Graph (CBMG), that is used to Income-based predictive mining is adopted to minimize describe the behavior of groups of customers by making recommendations based on their budget. All the customers developed a user requirement analysis most likely to buy recommended products. Association rule mining model designed for supporting and tracking dynamic profile was applied to a real Internet shopping mall for user behavior in online recommendations.

2. Literature Survey

In previous works of Raorane A.A, Kulkarni R.V, and Jitkar B.D, (1) Decision making and understanding the behavior of the customer has become vital and challenging problem for organizations to sustain their position in the competitive markets. Technological innovations have paved breakthrough in faster processing of queries and sub-second response time. Data mining tools have become surest weapon for analyzing huge amount of data and breakthrough in making correct decisions. The objective of this is to analyze the huge amount of data thereby exploiting the consumer behavior and make the correct decision leading to competitive edge over rivals. Experimental analysis has been done employing association rules using Market Basket Analysis to prove its worth over the conventional methodologies.

The Work of Tom Brijs, Gilbert Swinnen (2) introduces a new methodology for behavior-based customer segmentation. More precisely, it will use the method of latent class mixture modeling to discover hidden customer segments on the basis of the contents of their shopping baskets. Furthermore, loyalty card data is used to find out if these customer segments differ in terms of socio-demographic or lifestyle characteristics and whether these characteristics can be used to target different customer segments with more relevant product offers. The method is carried out on market basket data from a major Belgian supermarket store. Results indicate that a number of distinct segments can be identified who differ significantly in terms of the average purchase rate within a pre-determined set of product categories.

3. Method

3.1 Data Collection Method

The data is collected as real time data from the user by a form once the entered into the site. This may include external factors, such as social profiles, their interest, hobbies it is stored in different categories i.e. product wise, interest wise and repeated purchase wise. Depending on the necessary statistical analysis, required data is filtered through these databases. The proposed work researches on interesting directions for additional important item ranking criteria should be explored for potential diversity improvements. This may include consumer-oriented or manufacturer-oriented ranking mechanisms, depending on the given application domain, as well as external factors, such as social profiles. Also, as mentioned earlier, association rule-based approaches could be used to achieve further improvements in recommendation diversity, although these improvements may come with a (possibly significant) increase in computational complexity. Moreover, because of the inherent tradeoff between the accuracy and diversity metrics, an interesting research direction would be to develop a new measure that captures both of these aspects in a single metric. In addition, user studies exploring users' perceptions and acceptance of the diversity metrics as well as the users' satisfaction with diversity-sensitive recommender systems would be an important step proposed. Furthermore the exploration of diversity recommends item bundles or sequences instead of individual items. In proposed work, the system wanted to make the next logical step by allowing any item to be treated as a class label—its value is to be predicted based on the presence or absence of other items. Put another way, knowing a subset of the shopping carts contents, it want to "guess" (predict/recommend) the rest.

3.2 Fuzzy C-Means Clustering Algorithm

This algorithm works by assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point. More the data is near to the cluster center more is its membership towards the particular cluster center. Clearly, summation of membership of each data point should be equal to one. After each, iteration membership and cluster centers are updated according to the formula

$$\mu_{ij} = 1 / \sum_{k=1}^{c} (d_{ij} / d_{ik})^{(2/m-1)}$$

$$V_j = \left(\sum_{i=1}^{n} (\mu_{ij})^m x_i\right) / \left(\sum_{i=1}^{n} (\mu_{ij})^m\right), \quad V_j = 1, 2 \ldots\ldots c$$

Where, 'n' is the number of data points. 'vj' represents the j^{th} cluster center. 'm' is the fuzziness index m € [1, ∞]. 'c' represents the number of cluster center. 'µij' represents the membership of i^{th} data to j^{th} cluster center. 'dij' represents the Euclidean distance between i^{th} data and j^{th} cluster center. Main objective of fuzzy c-means algorithm is to minimize

$$J(U,V) = \sum_{i=1}^{n} \sum_{j=1}^{c} (\mu_{ij})^m \|x_i - v_j\|^2$$

Where, '$\|x_i - v_j\|$' is the Euclidean distance between i^{th} data and j^{th} cluster center.

3.3 Association Rule Mining

In data mining, association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases. Piatetsky-Shapiro describes analyzing and presenting strong rules discovered in databases using different measures of interestingness. Based on the concept of strong rules, Agrawal et al. introduced association rules for discovering regularities between products in large scale transaction data recorded by Point-Of-Sale (POS) systems in supermarkets. For example, the rule {bread ,butter}=>{jam}found in the sales data of a supermarket would indicate that if a customer buys onions and potatoes together, he or she is likely to also buy beef. Such information can be used as the basis for decisions about marketing activities such as, e.g., promotional pricing or product placements. In addition to the above example from market basket analysis association rules are employed today in many application areas including Web usage mining, intrusion detection and bioinformatics.

3.4 Implementation

The proposed system relied on two sources of data. To start with, it followed the common practice of experimenting with the synthetic data obtained from the IBM-generator.1 this gave us the chance to control critical data parameters and to explore our programs under diverse circumstances. The generator employs various user-set parameters to create an initial set of frequent itemsets whose sizes are obtained by a random sampling of the Poisson distribution. Then, shopping carts are created in a manner that guarantees that they contain these itemsets or their fractions. The lengths of the shopping carts are established by a random number generator that follows the Poisson distribution the user-set parameters. In the data worked with is varied the average transaction length and the average size of the "artificially added" itemsets. The specific values of these parameters are indicated in the names of the generated files. For instance, T10:I4 means that the average transaction contained 10 items and that the artificially implanted itemsets contained 4 items on average. Apart from using the IBM generator, experimented with two benchmark domains from the UCI repository2 that are broadly known, with characteristics well understood y the research community. To be more specific, it used the shopper product voting domain and the SPECT-Heart domain. The journal shopping data set has the form of a table where each row represents one man or woman and each column represents a bill. The individual fields contain "1" if the person voted in favor of the bill, "0" if he or she voted against, and "?" otherwise. The system numbers the bills sequentially as they appeared in the table (from left to right) and then converted the data by creating for each shopper a "shopping cart" containing the numbers of those (and only those) bills that he or she voted for. For instance, the shopping cart containing ½1; 4; 8_indicated that the shoppers voted for bills represented by the first, fourth, and eighth column in the table.

3.5 Method of Implementation

The Proposed work identifies the users independently and they are registered initially. A web log is created to identify the browsed contents and the browsed person and all the information's are recorded with respect to the time of access. Key here is the maintenance of a user profile that records the user's interests, preferences and other information, and the use of this profile to aid in the personalization of information retrieval.

With regard to the type of useful information to store, an initial starting point was information about the technological platform for the communication. Additionally one can consider some user preferences information. These are useful to guarantee an optimal information delivery path, to improve the user motivation and save time searching for resources. When designing the schema to represent the user profile, there was a conscious effort made to make it extensible so that preferences could be added in the future without the need to rewrite the schema

This model was extended to also incorporate users' action history, which was missing from the specification. The matching of profiles to metadata is also discussed as it fulfils an important role in the personalization process. Although, the user profile model presented is focused on E-Learning, the general platform could be applied to other areas. The Administrator can add multiple categories of products to shop and subcategories and controls the entire system. One more essential page lets the user register as a customer in the database. A form prompts the customer for name, address, telephone number, credit-card number, and password. This information is entered in the database and the customer is issued a customer ID number. Every time the customer adds a product to his or her shopping cart, its contents are displayed. This page includes an option to place the order. The catalog page also provides the option to browse the contents of the customer's cart.

Table 1: Transaction Data Set

Transaction-Id	*Items from customers who brought more than 1 item*
18	I pad, paintings
19	Laptop, CD's, I Pad
20	Accessories, mobile
21	Mobile, CD's ,laptop

3.6 Cluster Analysis

Cluster analysis is an effective way to personalized product recommendation. By clustering, the similar browsing or purchasing behaviors of customers can be identified. Common characteristics of customers can be analyzed and then the e-commerce companies are able to understand their customers better. By means of this model, the system can recommend the interest products for customers and track the recommendation effect for future use. Through clustering, customers who are close to each other within a certain range are partitioned into a group. A number of different classes are produced after clustering. The customers with 1 group have high similarity with each other while the customers in different groups have low similarity. In many applications, the data objects in a cluster can be treated as a whole. By clustering, dense and sparse areas can be identified. Therefore, the global distribution patterns can be found including some interesting correlations among data attributes.

Table2: Clusters' according to Product Purchased

ID	*Laptops*	*I pad*	*Mobile*	*Cds*	*TV*
16	0	0	0	0	0
17	0	1	0	0	0
18	1	0	0	0	0
19	0	0	0	0	1
20	0	0	1	0	0
21	0	0	0	0	0

4. Result

To compute association rules of Market Basket Analysis is to apply two threshold criteria: minimum support and minimum confidence. Thus set these two thresholds into cells for example, set 10% minimum support and 50% minimum confidence. The following resultant will be obtained:

Table 3: Support Count and Confident Values

Rules	*No. of occurrence*	*No. of itemset*	*Support*	*confidence*
165->166	(24->18)9	2->1.5	175	75
165->167	(24->14)6	2->1.16	158	50
165->176	(24->8)5	2->0.6	133	41
165->180	(24->6)4	2->0.5	125	33
165->182	(18->14)6	2->0.75	137	50
166->167	(18->8)5	1.5->0.16	133	41
166->176	(18->6)4	1.5->0.6	108	33
166->180	(18->9)7	1.5->0.5	100	58
166->182	(18->4)4	1.5->0.75	112	33
167->187	(14->8)4	1.5->0.3	91	33
167->170	(14->9)4	1.16->0.6	91	33
167->182	(8->)4	1.16->0.75	35	33
176->182	(9->4)4	0.66->0.75	70	33
182->167	(24->18->14)6	0.75->033	54	33

4.2 Comparison of Support Count

A set of items is referred to as an item set. An item set that contain N- items is an N-itemset. The set (Laptops, I Pad) is a 2-itemset. The numbers of times an item set occurs to in a transaction is said to be the support count. The proposed work is compared with some existing algorithms for its support value.

Table 4: Comparison of support count with three Algorithms

Product code	*Support Count*			
	Apriori	*DynFP-Growth*	*FP-Growth*	*Proposed*
10	13.94	2.32	3.79	112.5
20	21.98	3.98	6.88	116.66
30	48.37	8.23	14.63	116.66
40	66.50	12.10	20.90	137.5
50	107.65	3.76	34.30	175

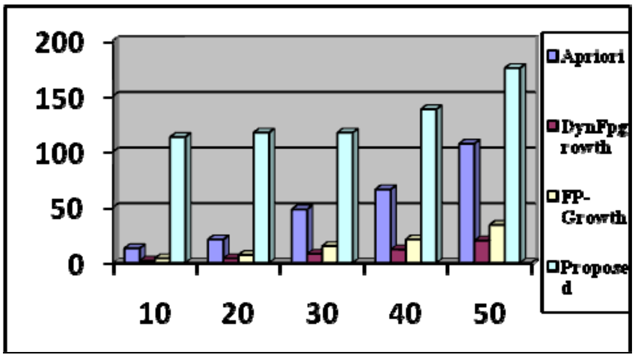

Figure 1: Shows the Comparison Rate of Support Count

4.3 Comparison of confident Value

Confidence is computed by taking ratio of support counts of the union of the dependent variable to the support count of dependent variable

Table 5: Comparison of confidence value with three algorithms

Support Factor	Confidence			
	Apriori	*DynFP-Growth*	*FP-Growth*	*Proposed*
10	13.94	2.32	3.76	75
20	21.98	3.98	6.88	50
30	48.37	8.23	14.63	50
40	66.50	12.10	20.90	50
50	107.50	19.50	34.30	41

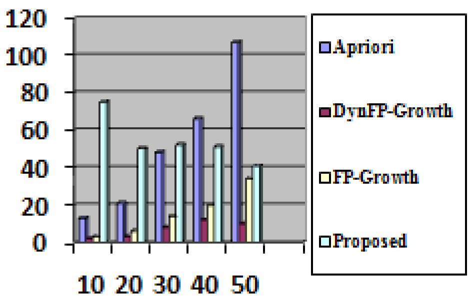

Figure 2: shows the comparison rate of confidence value

To demonstrate the significant improvement arising from our modified algorithm, comparably it reduces the processing time.

Table 6: Comparison of existing and proposed system

Measurement	*Existing*	*Proposed*
# of cases	59	59
# of corrected	43	47
Mean squared	0.01189	0.00594
Accuracy	72.881	79.661

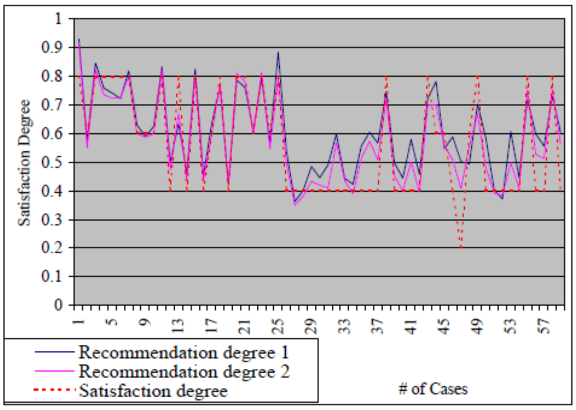

Figure 3: Shows the comparison of existing and proposed

5. Conclusion

The proposed system provides higher support count and confident value in which the recommend product will be according to the user desire and feasible to them. The error rate of predicting users mind is also reduced. The product predicted to users is much more reliable and likable by this system.

It is a simple system for shopping and understanding the customer personalization's and ranking and suggesting popular items designed to account for popularity bias. It focused on understanding the limit ranking of the items provided by the algorithms, and how it relates to that of the true popularity ranking and assessed the quality of suggestions as measured by the true popularity of suggested items. It believe that the problem posed opens interesting directions for future research including analysis of convergence rates of the ranking algorithms, consideration of alternative ranking and suggesting rules, and alternative user choice models.

6. Future Enhancement

The recommender system will be improved by predicting product according to Meta profile recommenders, social data mining systems, and temporal systems that recommend from location based perspective rather than income based. Temporal system is recommending products to the users according to the timely basis, like a festival or a new arrival or on vacation specials. Meta-recommenders are systems that allow users to personalize the merging of recommendations from a variety of recommendation sources employing any number of recommendation techniques can be combined. In doing so, these systems let users take advantage of the

ISSN (Online): 2319-7064

strengths of each different recommendation method. The product recommender system suggests new or previously un-purchased products to shoppers creating shopping lists on a personal digital assistant. The system considers a consumer's purchases across a store's product taxonomy. Recommendations of product subclasses are based upon a combination of class and subclass associations drawn from information filtering and co-purchase rules drawn from FCM data mining. Product rankings within a product subclass are based upon the products' sales rankings and income status within the user's consumer cluster, a less personalized variation of collaborative approaches. Also allows users to blend content requirements with personality profiles to allow users to determine which products they should select. It does so by merging more persistent and personalized recommendations, with ephemeral content needs such as the lack of offensive content or the need to be home by a certain time. More importantly, it allows the user to customize the process by weighting the importance of each individual recommendation for future work.

Reference

[1] "Association Rule – Extracting Knowledge Using Market Basket Analysis" Raorane A.A.1 Kulkarni R.V.2 and Jitkar B.D.3, 2012.

[2] T. Brijs, T. Swinnen, K. Vanhoof, and G. Wets, "Using Shopping Baskets to Cluster Supermarket Shoppers," AARTF, Amelia Island Plantation, Fla, 2001.

[3] Gediminas Adomavicius and YoungOk Kwon "Improving Aggregate Recommendation Diversity Using Ranking-Based Techniques", Vol. 24, MAY 2012

[4] Subhash K. Shinde, Uday V. Kulkarni "Hybrid Personalized Recommender System Using Fast K-medoids Clustering Algorithm", 2011

[5] G. Adomavicius and A. Tuzhilin, "Expert-Driven Validation of Rule-Based User Models in Personalization Applications," Data Mining and Knowledge Discovery, Vol. 5, 2001.

[6] G.M. Allenby and P.E. Rossi, "Marketing Models of Consumer Heterogeneity," J. Econometrics, Vol. 89, 1999.

[7] Subhash K. Shinde, Uday V. Kulkarni "Hybrid Personalized Recommender System Using Fast K-medoids Clustering Algorithm", 2011

[8] Using Shopping Baskets to Cluster Supermarket Shoppers",Tom Brijs Gilbert Swinnen Department of Applied Economic Sciences Limburg University Center B-3590 Diepenbeek, Belgium Koen Vanhoof Geert Wets.

[9] Jean- Francois Belisle, "An Introduction to Data Mining for Marketing and Business Intelligence in Analytics", 2011

[10] G.M. Allenby and P.E. Rossi, "Marketing Models of Consumer Heterogeneity," J. Econometrics, Vol. 89, 1999

Author Profile

S. Saravana Kumar pursing his PhD degree in computer science in Bharathiar University He is an assistant professor of information technology and placement officer in Sri Jayendra Saraswathi Vidyalaya College of arts and science. He is a Member of Board of Studies for Computer Science in S.N.R Sons College (Autonomous) Coimbatore from June 2005. He got **1st rank** in M.Sc Computer Science at college level and received **Gold Medal** for being the Topper from S.N.R Sons College. He has Presented a paper on "mobile learning" at the National Level Seminar on e-learning organized by SNMV College of Arts & Science, Coimbatore, Participated in the two days skill based training program on "J2EE and Best Practices in Application Development" organized by ICT Academy of TamilNadu in association with CTS & TCS at Bharathiar University., Participated in the three day workshop on "Masters Training" program organized by the Department of Career Guidance, Bharathiar University.

S. Diviya, currently pursuing her M. Phil degree in Computer Science in Sri Jayendra Saraswathi Vidyalaya College of arts and science. She worked as an assistant professor in computer technology department in K.G. College of arts and science for past three years. She has completed training in "Alice" software for creative programming with visual effects and also participated in one day workshop on effective mentoring. Her research areas include web mining and personalized product recommendation.