



Efficient Mining and Recommendation of Extensive Data Through Collaborative Filtering in E-Commerce: A Survey

N.Naveen¹, S. Ganesh Kumar²

²Asst Professor (S.G)

^{1,2}Department of Computer Science Engineering
SRM Institute of Science and Technology, Kattankulathur, Chennai

*Corresponding Author Email: naveenreigns@gmail.com

Abstract

E-Commerce is the most widely used technique nowadays. Buying and selling goods on the Internet has been most admired and frequently utilized. The humongous growth of the content available on the internet has made laborious for users to search and utilize information for classifying the products. Recommendation system regarded as the best way to help the customers in buying the related products. (GRS) group recommender system aims at enhancing the customer's benefits for buying the products. This paper summarizes the fuzzy tree matching, modeling user preference dynamics, web page recommendation, uncertainty analysis for keywords, recommender system application, temporal topic model for friend recommendation, autocratic decision-making system based on (GRS), modeling user recommender, evaluating recommender system and enhancing (GRS).

Keywords: E-commerce, Group recommender system (GRS), recommendation based system, User Preferences.

1. Introduction

During 1990's E-Commerce was introduced and progressed at a faster rate than they could ever imagine. Currently, many companies have the online presence and ability to conduct the business that has been so vital and essential. A-Z products can be bought by the customers through online. The usual form of E-commerce is business to customer, where the trading of the products deals between themselves. Illustration of a business to customer is Buying Jersey from Adidas website. Business to business is another variety of E-commerce trading. The particular organization will trade own websites, Domain names and hosting services to other organization. Illustration of this type is Go Daddy website. At last direct dealing between customers in e-commerce is by trading among themselves. Illustration of this type is OLX. The recommendation has changed the evolution of the E-Commerce. The Different types of the recommendations are used and other factors that are customer driven system are deployed in order to achieve the target. These methods have made done by getting the user preferences and applying for the betterment of the model. Even in some models, keywords have been used as a key to enhance online shopping. Group recommender system (GRS) is the advanced technique of the recommendation system. Opinions of the users are also been observed in form of the ratings and likings of the recommendations. The interest of the products differs from time to time and customer to customer but the recommendation has to be in such a way that it benefits both the

customers and the sellers. Some models have even analyzed and evaluated by feedback given by the users as their preferences.

A. Fuzzy Tree Method

The researchers say that the advancement of web-based information and E-commerce technologies has made their way for E-learning a widespread system. It has been applied in many of the fields nowadays. This has changed and evolved into a new scope of learning and gaining knowledge through online. One of the most renowned and frequently used in recent trends helps in recommending called as the recommender system. It attracts the customer needs by trying to match their expectations. This helps the user by predicting based on the preference. Fuzzy tree defines categories that are to be specified which roughly belongs. The measures are done to enhance the semantic similarity. This is a productive method that deals with new learner and new learning techniques which has to be suggested. Resemblance between learning and learner is a major measure in this method. But it is a tough process to find a way of learning by which it can be further divided into two types [1].

B. Modeling user Preference

This suggested model reads the mind of the customers by user preference dynamics (UPD). The innovation and introduction of a new product will attract the minds of the customers. Illustration of this method is, a user chooses a laptop of the normal version but tends to get attracted to the new features that are included in another laptop. It has increased at a rapid growth. This feature makes it a unique from other models. The wide range of products makes it a

customer driven model. (UPD) changes from time to time for a particular product. In certain circumstances UPD will influence in filtering the process that is very complicated [2].

C. Web Usage and Domain Knowledge

The Effectiveness of the recommendation lies in the method by analyzing the model and knowledge used. Training datasets are taken to measure the performance. If the training datasets are more accurate, then efficiency will be higher. In the web page recommendation, the probabilistic model and association rules are applied. Web log files will be collected from the server for efficiency and will be recommended for some time. List of URLs of the web page and weblog files have been accessed by the users. Discovered knowledge has succeeded in making efficient WebPages. Probabilistic models made a systematic approach in web usage data. It exploits side information with regards to the private attributes for the recommendation of the scheme [3].

D. Keywords System

Since web pages are innumerable, it is difficult for the users to surf. Researchers suggest by detecting the web page and summarizing the content. Certain information's are needed always like anti-government acts and armed conflicts. Some uncertain part of information's has evolved with the rise of web event. For example, Result of 10th and 12th students. These web pages are for a shorter period of time. Study of nature which is semi automated structured (Domain on to WP) and the method is efficient in recommendation of the web pages by means of satisfaction. Complex network metric needed for two successive time stamps hence Time complexity plays a major role in this model [4].

E. Recommender System Application

The one who operates the location, time and quadruples is based upon this system. Since it is a four-dimensional data it is harder to find the occurrence from high dimensional data. User spatial activity preference is not reachable when the continuity of the location occurs in a continuous manner. Many users do not implement login due to the shortage of time and privacy. This login suffers from a data scattered problem that acts as a hindrance for this model. Exploitation activity occurs to check temporal activity preference. To solve this complexity (STAP) model does the spatial specificity and temporal method. It is tough to directly manage four dimensional data which lags in data sparse problem [5].

F. Recommendations By Microblogging Systems

This is the most prevalent media basement for Internet users. Micro blogging became a user-friendly choice for searching and expressing customer's point of view. This will enhance to find friends of the same domain interest and will help in improving the social recommendation. This method improves the interest of a particular customer on the product by means of recommendation. This leads to social attraction and will make customers drive towards the particular product. It predicts their capabilities to find friends and engage with social environment by micro blogging. Similar interest will improve good relationship between the friends and serves for business trading process. This is complex method where other data sets will be validated. The analysis of the interest by the usage pattern is a failure due to insufficient keywords [6].

G. Recommendation By Autocratic Decision

This phenomenon depends on the limit or the range of the linguistic terms and likelihood. The autocratic decision making is done effectively by means of Group recommendation system (GRS). It imports all limits within the range of the linguistic terms to perform the matrix operation. This will validate the smaller threshold frequency between 0 and 1. The limit of the linguistic range will be updated. The obtained matrix is of the degree form and condition needs to be satisfied. The performance of the product will be better in performance when it is in large rather than in smaller groups. It calculates the score of the each and every possible matrix that are obtained. Ordered weighted average (OWA) of group decision is by analysing the linguistic model. The prescribed model is not cost effective and more confusing to find linguistic intervals [7].

H. User Activity Preference in Lbsns

Location Based Social Network (LBSNs) plays a vital role for millions of users to access. It helps in understanding mobile users about (STAP) as a personalized context-aware location recommendation. It became popular by the method called the point of interest (POI). This system allows the user to express their views and suggestions regarding the products. This will improve the social ratings and reviews for the particular products and acts a key for a customer-driven method. Thus by understanding the user's intention and freedom to express makes it unique from all the models. It suggests the recommendation by user feedback and their views. Factorization is done to filter for the user preference product for similarity cases. This model suffers from data scarcity problem and makes it complex process to evaluate. There is no clear solution for this method as data are scattered in different dimensions [8].

I. Evaluating Recommender Systems

This is one of the best ways to provide user-friendly service even when users are at different locations. Algorithm deployed in this method is incremental algorithm which has been well constructed. The important feature in evaluating recommendation system is to establish more diverse, comprehensive, scalability; reliability etc. Facets will depend on accuracy, coverage, diversity and risk factor of the model. The positives in recommender system need to be taken and rest have to be discarded for development. It determines the efficient way of recommendation and can go beyond for accuracy, coverage, risk factor; adaptability etc. The accuracy is not enough for the need of recommended quality [9].

J. Group Recommendation By Social Interactions

The Group recommendation is a testing problem based on the dynamics of the members and diversity. Digital is becoming popular day by day as it makes user connected by means of digital form. It can be described in various scenarios. For instance, watching TV programs with family and watching movies with friends. It has created a good impact on both types. Users work preference and social activities are those of the types. These scenarios will make the system build in an efficient manner to help a proper recommendation by a means of the social interactions. This also allows users to share their views and exchanging the information for benefits. It utilizes the social and content interest of the members in the group. Effectiveness is high in the method because user can think and analyze the recommendation that was put forward. It is testing problem that lacks in consistency for longer period of time. The social characteristics within the group are ignored that results in sub optimal GRS performance [10].

K. Interest Circle Based Recommendation

One of the most significant methods of recommendation is done by interest circle based model. In this process they utilize the point of interest (POI) to get the views and likings of the user for a particular product. The past interest of the users will be used for the purpose of recommendation and currently used in many applications scenarios. Hence the social relationship is enhanced which may benefit the users opinions. By making use of social interactions cold start problem can be improved. The set back in this model is if a customer has rated many items which means his interest is broad but not focused [11].

L. Collaborative Filtering Methods

Another way of recommendation is by using collaborative filtering (CF). CF is segregated into two types as user based and item based. Thus in this method social recommender system is established to acquire the network socially. In the process the collective information of ratings by the users are filtered depending on the disorder and unrelated information. The disadvantage of this process is the elimination of limitation of similarity dimension but not differences of dimensions [12].

M. Leveraging Items Connections

Using social relations, user generated reviews and items connections are combined into matrix factorization. User similarity, item similarity, ratings and reviews of the customer are integrated to form useful information in recommender system. Hidden topic matrix factorization (HFT) to combine based on the reviews and the ratings of the users. A real world datasets consists of item connection and user connection which has to be improved for the efficiency purpose [13].

N. Social Trust Level Model

It calculates the social trust level between two users is by explicit trust on one user to another based on user's behaviour or similarity. In this model they employ a technique called level of trust (LO TRUST) which is used for personalized recommendation systems. This depends on user attraction and interest for a particular period of time. An online social network (OSNs) acts as a medium for the people who can share the information's. Trust can be notated by value and may change according to the time period. Complete willingness and interactions are based only on the trust level [14].

O. Unsupervised Ensemble Based Model

This model will identify threats from the normal needles in context setting. These entities will allow following the unsupervised model used. When compared in diverse many people will gather information that may be sensitive aspect. Insiders will be with complete overview about the internal working system and tasks that has to be performed [15].

P. Analysis

The web lays the foundation for E-commerce to reach out to customers. Recommender system targets to vision of the customers and their attributes. There are two techniques mainly used in this

proposed system. Fast diffusion and information sharing capability are techniques deployed here. This has to be done for an established network and for the needs of the users. Opinion dynamics dependency on (GRS) based method is a measure of matrix weight to obtain the desired product.

Q. Current Research

This research is for the analysis based on frequently purchased items to achieve the optimal algorithm. The algorithm that used been used has more execution time than FUP algorithm. In this method, the dynamic recommendation for the frequently purchased items is used. The earlier models are used to derive the algorithm which is more efficient and tends to read the customer's choice. It plays an important role in enhancing the marketing strategy for marketers for their sales. The transaction history database is subdivided into many parts to find usable groups which are made to split up. Plotting the customer needs and based on that groups are implemented. Transaction history and sequence is done for the past purchases customers. The prediction of the purchase history that is purchased by the customers and extracting the frequently purchased items used from the transactions. Agglomerative hierarchical clustering, collaborative filtering and association rule mining are some of the algorithms that have been used.

R. Original Database Discovery

With the help of the dynamic database new transactions can be inserted. It will activate new association rules and will simultaneously invalidate existing database. This has been a major issue for the developers. It should maintain association rules for database that are dynamically arranged. Therefore a new algorithm was introduced to rectify such a problematic occurrence. Assumption is new way of adaptability for transactions that change from original transactions for statistics purposes. These can be collected by previous method in which mining would have taken place to extract the data. Support count and confidence was brought in order to gain more knowledge about the user's preference.

S. Modules

For the purpose of the future enhancements these modules are used for this project. The modules are done for betterment of the proposed system to overcome the system that was proposed. The following are used for the certain part and are listed. User Interface, Clustering transaction history, Allocation Phase, Intra Cluster dissimilarity, Inter Cluster similarity, Refinement phase, Transactional SL Ratio, Incremental Association Rule, Product Recommendation are the modules used. User interface acts as key element for human-machine interaction. It acts as the communication between the human and the machine. This target is most effective form of the interface. Manipulation can be accessing the inputs from user. The user then will implement login if he was as existing user or else he needs to register. This will lead to next stage of the process. Transaction history that is clustered is based on the number of inputs. Input can be described by number of transaction history databases. The clustered set of transactions will be obtained as the result.

The initial stage is to find the frequent items when compared to the history transactions databases that have to be grouped. The transaction history database contains the previous transactions made by the customers. The details include customer id, the set of items bought along with the transaction id. Allocation phase is the transactions in a sequentially manner. Transaction (t) will be

implemented to an existing cluster or a new cluster. The new cluster is formed by number of accommodation (t) in order to be cost efficient for the clustering. Every transaction identifier is written back to the database for allocated cluster. Depending on the cost few or more clusters may be added to obtain new product. Intra-cluster dissimilarity tells us how different the transactions are within a cluster. It depends on intra cluster dissimilarity (U), Small items (S_m) and maximum ceiling (E). The maximum ceiling is the maximum number of transactions that might contain an item to call it a small item. Thus intra cluster dissimilarity is the union of distinct small items present in all the clusters. Inter cluster similarity is another type of the clustering module. Depending on the cost efficiency the process may be further divided. In refinement process small large ratio (SLR) can be explained by ratio of number of the small items to that number of large items. The ratio is compared to threshold value of SLR. If transaction in SLR is more than threshold value then transactions are shifted to excess pool. The incremental association rule mining relies on the transaction history database. The output obtained will be the frequent items sets and promised frequent item sets.

T. Algorithms

Agglomerative hierarchical clustering is method of clustering from bottom to top of the branch. It is also known as bottom-up clustering method. These will be further divided into further clusters till it reaches the root node. It is successive iteration by which it merges the nearest or closest pair of the clusters by fulfilling criteria's. Then it is evaluated with pair wise of clusters that are formed and calculating the distance between them. A distance matrix is constructed based on the values that are obtained. Ordering of objects are produced which may be informative for the purpose of data display. Even small groups are formed which may tend to be helpful for finding a new product. The next process is further done by association rule mining. It is a process of finding a probability of the items that may occur. It is pattern that defines when A occurs, B also occurs with some probability. Target is to find the user specified minimum support and minimum confidence. Finally collaborative filtering (CF) is the technique used by the recommender system. It is a process by which predictions are the done automatically regarding the user's interest by collecting information's or likings from user. It can be also processed by predicting the rating or the preference for a particular item. The datasets are calculated by similarity measures and these values are used to predict the user's preference in the form of ratings.

U. Future Enhancement

In future, the evaluation based on the other techniques that are employed. It also plans to find other suitable datasets to make more tests and compare the efficiency of our method among different datasets. It plans to use patterns extracted using other Data Mining techniques such as clustering and classification in the process of designing a website for some grocery store or company. The automation of the process of building the improved prototype belongs also to the future work. Product recommendation defines the finally recommended items that are filtered. It also includes the purchase rate and fluctuation mind of the customer's interest. Then items that are liked or used in their shopping cart or wish list.

2. Conclusion

The literature reviewed in this explains how group recommendation based system (GRS) has applied and other features that have been

dedicated to the past. Researchers have highlighted a common theme of giving importance to various factors that describe the trading techniques in E-Commerce. These techniques have helped in gaining knowledge as well as by implementing. Using the positive notes of various data mining techniques along with various study completed by researchers earlier, the proposed model has been more efficient.

3. Acknowledgment

The author would like to thank Dr. S. Ganesh Kumar (SRM IST) for his constant support and guidance throughout the survey. We would also like to thank Dr. Amutha [head of CSE Department SRM] and other faculty members of SRM IST.

References

- [1] D. Wu, J. Lu, and G. Zhang, "A fuzzy tree matching-based personalized E-learning recommender system," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 6, pp. 2412–2426, Dec. 2015.
- [2] D. Rafailidis and A. Nanopoulos, "Modeling users preference dynamics and side information in recommender systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 6, pp. 782–792, Jun. 2016.
- [3] T. T. S. Nguyen, H. Y. Lu, and J. Lu, "Web-page recommendation based on Web usage and domain knowledge," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 10, pp. 2574–2587, Oct. 2014.
- [4] J. Xuan, X. Luo, G. Zhang, J. Lu, and Z. Xu, "Uncertainty analysis for the keyword system of Web events," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 6, pp. 829–842, Jun. 2016.
- [5] J. Lu, D. Wu, M. Mao, W. Wang, and G. Zhang, "Recommender system application developments: A survey," *Decis. Support Syst.*, vol. 74, pp. 12–32, Jun. 2015.
- [6] N. Zheng, S. Song, and H. Bao, "A temporal-topic model for friend recommendations in Chinese microblogging systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 45, no. 9, pp. 1245–1253, Sep. 2015.
- [7] S.-M. Chen and B.-H. Tsai, "Autocratic decision making using group recommendations based on intervals of linguistic terms and likelihood based comparison relations," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 45, no. 2, pp. 250–259, Feb. 2015.
- [8] D. Yang, D. Zhang, V. W. Zheng, and Z. Yu, "Modeling user activity preference by leveraging user spatial-temporal characteristics in LBSNs," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 45, no. 1, pp. 129–142, Jan. 2015.
- [9] A. Gunawardana and G. Sani, "Evaluating Recommender Systems," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, and B. Shapira, Eds. New York, NY, USA: Springer, 2015, pp. 265–308.
- [10] M. Gartrellet al., "Enhancing Group Recommendation by Incorporating Social Relationship Interactions," in *Proc. 16th ACM Int. Conf. Supporting Group Work (GROUP)*, Sanibel, FL, USA, 2010, pp. 97–106.
- [11] X. Yang, H. Steck, and Y. Liu, "Circle-based recommendation in online social networks," in *Proc. 18th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2012, pp. 1267–1275.
- [12] X. Yang, Y. Guo, Y. Liu, and H. Steck, "A survey of collaborative filtering based social recommender systems," *Computer Communications*, vol. 41, pp. 1–10, 2014.
- [13] J. Tang, X. Hu, and H. Liu, "Social recommendation: a review," *Social Network Analysis and Mining*, vol. 3, no. 4, pp. 1113–1133, 2013.
- [14] Y.-M. Li, C.-T. Wu, and C.-Y. Lai, "A social recommender mechanism for e-commerce: combining similarity, trust, and relationship," *Decision Support Systems*, vol. 55, no. 3, pp. 740–752, 2013.
- [15] H. Ma, I. King, and M. R. Lyu, "Learning to recommend with social trust ensemble," in *Proc. 32nd Int. ACM SIGIR Conf. Research and Development in Information Retrieval*, 2009, pp. 203–210.
- [16] T. Padmapriya and V. Saminadan, "Improving Throughput for Downlink Multi user MIMO-LTE Advanced Networks using SINR approximation and Hierarchical CSI feedback", *International Journal*

of Mobile Design Network and Innovation- Inderscience Publisher, ISSN : 1744-2850 vol. 6, no.1, pp. 14-23, May 2015.

- [17] S.V.Manikathan and K.srividhya "An Android based secure access control using ARM and cloud computing", Published in: Electronics and Communication Systems (ICECS), 2015 2nd International Conference on 26-27 Feb. 2015, Publisher:IEEE,DOI: 10.1109/ECS.2015.7124833.