

E-COMMERCE PRODUCT RECOMMENDATION SYSTEM**¹ Ms.Sk.Raziya Sulthana, ² I. Thanuja, ³ K. Naveen Kumar, ⁴ M. Vamsi**

¹ Assistant Professor, Department of Computer Science & Engineering, Chalapathi Institute of Engineering and Technology

^{2,3,4} Department of Computer Science & Engineering, Chalapathi Institute of Engineering and Technology

Abstract: — Web is accelerating and altering how day to day errands like web based shopping, covering service bills, observing new films, imparting, and so on, are achieved. For instance, in more seasoned shopping techniques, items were efficiently manufactured for a solitary market and crowd yet that approach is at this point not suitable. Markets in view of long item and improvement cycles can never again get by. To remain cutthroat, markets need to give various items and administrations to various clients with various requirements. The shift to internet shopping has made it officeholder on makers and retailers to tweak for clients' requirements while giving a larger number of choices than were conceivable previously. This, notwithstanding, represents an issue for clients who should now dissect each contribution to figure out what they really need and will profit from. To help clients in this situation, we examine about normal recommender frameworks methods that have been utilized and their related compromises.

I. INTRODUCTION

Recommender frameworks were first presented as Cooperative Sifting by its creators in which they talk about how individuals team up and channel email reports that are applicable to them and of purpose to their crowd. The sifting system included investigations of normal properties among at least two records. Properties having a place with the records that were dissected included message, answer or its explanations. This was viewed as more successful than straightforward investigation of the archive's items which numerous other mail frameworks gave. Human mediation of the separating system prompted additional fascinating records being chosen [1] [2].

Recommender frameworks permit fast and computerized customization and personalization of internet business destinations. They permit the locales to create more deals by fitting to the requirements of the guests and transforming them into shoppers, up-selling additional items by packaging firmly related things together, and expanding client devotion [3] [4]. Client dependability is accomplished by showing clients that they require some investment to figure out their necessities and to look into them [5]. This is apparent when the site structure, the items, and presentation of products changes to customers' needs and preferences. Customers revisit these websites rather a competitor's because they are accustomed to it and do not have to go through a learning process. Even if the competitor were to offer similar experience, customers will return to a site they already know.

II. TYPES OF RECOMMENDATIONS

Recommender systems can be personalized, non-personalized, attribute-based, item-to-item correlation, and people-to-people correlation. Recommendations are either short-lived or

long-lived depending on the implementation. The system is considered automatic if it requires minimal or no input from the active user and manual if it requires some work [3]. Personalized recommendations are automatic and based on the user's preferences such as favorite color, movie genre and music group. They are often compared against hand-picked products by content-providers and experts for user's preferences and tastes to provide recommendations [10]. Non-personalized recommenders generate recommendations based only on product ratings from other users of the system [10]. These recommendations are straight forward since they require very little effort to produce and considered automatic manual since the user must explicitly search for a certain type of product to base the recommendations on [3] [10].

These recommendations can be short-lived or not depending on how long the system remembers user's preferences for. Item-to-item correlation recommenders recommend items based on other items the user has displayed interest in. These recommendations are prevalent in e-commerce sites where new products are recommended based on what the user has in their shopping cart [3]. These recommendations are manual since user must have a non-empty cart, and short-lived because the user does not have full shopping carts. Association rules are most often used in this system [11]. People-to-people correlation system finds similarity between the active and other users in the system, recommends products other customers have purchased or rated in the past [4]. Collaborative filtering is the most commonly used approach in this system [12]. Since it requires users to have purchased or rated products in the past this method is very manual. The recommendations can last depending on system's design.

III. RECOMMENDATION TECHNIQUES

Different algorithms and techniques are used by recommender systems to generate recommendations. The most popular ones are association rules, collaborative filtering, content-based filtering and

hybrid filtering.

A. Association Rules

Association rules are used to recommend products based on their presence along with other products [4] [13]. When two products are purchased together, the presence of one item in a transaction can be used to determine the second product also being in the same transaction. This is very useful when making recommendations to new users who wish to make purchases. To define association more formally, a collection of products m products $\{P_1, P_2, P_3, \dots, P_m\}$ belongs to set P . We say a transaction T from set of transactions D is a subset of P , $T \subseteq P$ such that the transaction contains products from P . Each transaction can be uniquely identified as TID. A transaction T contains set X , a subset of products from P and it is a subset of T . Association rules implies that there exists Y , subset of P and there is no mutual product between X .

This means that whenever products from X exist in a transaction T , there is high likelihood that products from Y will also exist in the same transaction [11] [14]. Two variables, confidence c , and support s [11] are used to measure the quality of the associations made [4].

Support measures how frequent the association happens in the entire set of transactions as shown in (1) and confidence measures the frequency of both products occurring whenever one product exists in the transaction as shown in equation

$$s = \frac{\text{number of transactions containing X or Y}}{\text{total number of transactions}}$$

$$c = \frac{\text{number of transactions containing X or Y}}{\text{number of transactions containing X}}$$

Collaborative Filtering

Collaborative filtering approach uses customer details, ratings, and reviews aggregated from all the customers to build recommendations [6] [12] [8]. The strength of this approach is that it analyzes existing active customers with similar preferences and characteristics of the current customer to build the recommendations. The filtering method is achieved through a heuristic-based, a model-based method, or a hybrid model that combines characteristics from both heuristic and model-based approaches [2] [4]. The heuristic based or memory-based collaborative filtering model takes in rating data, whether product was purchased or not, and duration of viewing products to calculate the recommendations [2] [12]. Active customers whose information is used is done by selecting all the customers who are neighbors of the current customer using similarity measures including personal information, cosine metric, and jaccard coefficient for binary data [2]. Then, utilizing k-nearest neighbor classification method, prediction value is computed for each product that current customer has not viewed but the other active customers have. With the newly calculated set, recommendation is created based on products with the highest scores. There are many different algorithms and technique that can be used in heuristic based collaborative filtering includes k-nearest neighbor algorithm, web mining algorithms, decision trees, and support vector machines [2]. The model based collaborative filtering technique uses training data such as the active user's ratings and reviews to build a model using

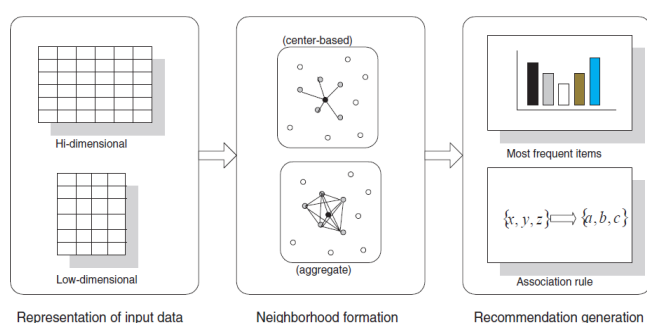


Figure 1 - Part of Recommendation Systems [4]

the i -th customer bought j -th product and zero otherwise. The matrix is called original representation [4]. Collaborative filtering has challenges with sparsity, scalability and synonymy. Synonymy occurs because similar products are labeled differently in real life, and recommender systems cannot always associate between them, and treat them as different. A reduced dimensional representation is constructed to alleviate the weaknesses. A matrix of

size $n \times k$ is constructed where all values in the matrix are nonzero, which implies that each customer has had an association with the k product. Due to decreased size, it also helps alleviate the problem with synonymy.

The neighborhood formation forms the heart of the recommendation system. In this step, the similarities between customers are computed and used to create proximity based neighborhood between the target customer and likeminded customers [13]. For each customer u and N customers where N

$= \{N_1, N_2 \dots N_l\}$, the customer u does not belong to set of N and the similarity $\text{sim}(u, N_k)$ $\text{sim}(u, N_k)$ is greater than $\text{sim}(u, N_{k+1})$ $\text{sim}(u, N_k)$ with $\text{sim}(u, N_1) \text{sim}(u, N_k)$ being the maximum. Proximity measures can be calculated using (3) or (4).

different data mining and machine learning algorithms [2] [12]. The model is then validated using the testing data and list of products and rating is predicted for them if customers have not given any rating to it yet or been exposed to it. While the heuristics based model uses the entire database and the customers to create recommendations for the active customer, the model based approach only relies on the active customer's information as the input. Techniques and algorithms from fields such as Bayesian model, clustering, association rules, artificial neural networks, linear regression, maximum entropy, latent semantic analysis, and Markov process can be used [2].

Collaborative filtering is the most successful technology used in recommender systems and it is the most widely used on the internet [4]. The recommender system is split into three components: representation, neighborhood formation, and Equation 3 calculates the correlation between two different variables in terms of how the variables are related. The correlation between user a and b is defined as the summation over i are over the items for which both user a and b have voted [12] [15]. The notations r_{ai} and r_{bi} represent the rating given to i -th item by user a , and user b respectively. r_a and r_b represent the averages. The result is between -1 and 1 with -1 being a perfect negative correlation. In equation 4 both a and b are vectors in the m dimensional product space and the distance between them is calculated as the cosine of the angle between the two vectors. For n customers, a similarity matrix S of size $n \times n$ is computed using either one of the proximity measures.

There are two methods to forming a neighborhood: center-based and aggregate neighborhood [4]. Centre based techniques form a neighborhood for a customer c of size k by selecting l nearest customers where both k and l are arbitrary. An aggregate neighborhood creates a neighborhood of size l for customer c by selecting the closest customer. The rest of the $l - 1$ neighbors are selected similarly. At a certain point jj , when $\bar{C} \rightarrow$ there are j neighbors in N and $j < l$, the centroid of $N, \bar{C} \rightarrow$ is calculated using (5). Then a new customer w who is not in N is selected as the $j+1^{\text{th}}$ if w is the closest to the centroid $\bar{C} \rightarrow$. The centroid is then recomputed for $j + 1$ neighbor and continues until the number of neighbors in N is l . gives recommendations based on items the user has viewed in the past. The contents can be described using labels and the labels are given a weight of how well they describe the article. Using these labels and user preferences, nearest neighbor or clustering algorithms can be used

to recommend other articles to the active user. However, new users with limited information and a limited number of labels pose a challenge to this method. Common algorithms that are applicable include k-nearest neighbor, clustering, Bayesian, and artificial neural networks [2]. Information filtering systems are usually used with structured data that can be easily analyzed to gain insights. Vast amounts of data are usually analyzed by filtering systems to give recommendations because it is The user profiles are obtained explicitly through questionnaires and forms or implicitly using behavioral information.

actual recommendations which is to calculate top m recommendations from the computed neighborhood of customers. Two prominent techniques that are used are most-frequent item recommendation, and association rule-based recommendations [4]. In Most-Frequent Item Recommendation, neighborhood N is scanned frequency count of purchases is calculated for each neighbor. All the product recommendations to the user. The attributes are compared with keywords describing the recommendations as mentioned. Keywords used to make recommendations are weighted using term frequency/inverse document frequency (TF-IDF) method to measure importance. Term frequency TF is calculated from N items that could potentially be recommend to user as [6] [16].

are then sorted according to the frequency and m most frequently bought products that is not purchased by the current customer are recommended [4]. In Association Rule-Based

Recommendations: L neighbors taken into account while using association rules to generation recommendations. Association rules work by recommending a product that a neighbor bought with the presence of another product [14]. However, having a limited number of neighbors to work limits the effectiveness of the recommendations made [4]. Collaborative filtering has a major disadvantage since it requires data to exist in order to be useful. It has two major customers cannot easily purchase products such that they buy even 1% person of the store's products. A recommender system that uses nearest neighbor algorithms is ill suited to make recommendations for an active user in those sites. This is commonly known as reduced coverage. It also leads to poor recommendations due to lack of enough data [4]. Nearest neighbor algorithms grow with the number of customers and products available, thus leading to scalability issues.

B. Content-based filtering

Content-based filtering is based on being able to analyze products and find similarity with active user to recommend products. Unlike collaborative filtering or association rules, this method does not require an active database of purchase history. It is based on information retrieval, analysis and filtering [2] [6] [16] [17]. This approach is used mainly in places where content can be read or analyzed such as news articles, movies and anything with metadata attached. Then we can simply get the weight for keyword k_i in document $k_i d_j$ as [6] [16]:

$$w_{i,j} = TF_{i,j} \times IDF_i \quad (8)$$

Content-based filtering systems also recommend new items based on what the user had liked previously [6]. A content based profile can be constructed for a user from their previously

liked items, ratings, search keywords, and other behavioral data. This information is aggregated to create a profile for the user. These types of systems are highly dependent on the items being easy to analyze. In order for recommender systems to be able to generate recommendations, content must be structured and easy to parse. If this is not, then the item must be described manually [6]. Another problem is being able to differentiate between a bad item and a good item based on retrieved information. A bad item using same keywords as good item will also get recommended.

Two other major drawbacks are lack of information about a user, and overspecialization. When a new user is introduced into the system, their preferences and profiles are not aggregated. The user would not have given enough ratings, and reviews to products. This leads to insufficient information to generate recommendations [6]. When the system is only able to recommend certain items based on user's profile, it leads to overspecialization. This is due to the user having rated a specific item, the recommender system is only able to provide recommendations for similar products. This also leads to the user never being recommended outside of their previous ratings [6]. In such cases, genetic algorithms which evolve information filtering agents to provide recommendations have been proposed. This is done by using an iterative method where previous output is used to learn and adapt dynamically [18] [19].

C. Hybrid filtering

To avoid problems that exist in both content-based and collaborative filtering systems, hybrid solutions have been proposed [6]. Solutions include: implement both filtering separately and combine the results, incorporating characteristics of content-based filtering to collaborative adding characteristics of collaborative filtering to content-based filtering systems and new algorithms that incorporate both systems' techniques. Combining different recommender systems approach involves building two different recommender systems based on collaborative-based and content-based approaches. The recommendations can be separately generated and then combined linearly [20]. The algorithm assigns a weight to the generated recommendations per user based on its relevance to the user. The recommendations are then added in order to be presented to user. The second method is to use the level of confidence each system produced for the results that are more consistent with the user's past ratings and provide them to the user [21]. Many recommender systems are implemented using collaborative-based approach with content-based user profiles generated through content-based approach [6]. The profiles are then used to find similarity between users rather than items which help the system overcome some of the sparsity-related limitations. Recommendations can be generated through collaborative filtering first. They are then compared against current user profile to determine if it's interesting to the user or not and to present it [19]. Curse of dimensionality occurs when a lot of features exist per item that makes it difficult to cluster or compare them [13]. The most common approach is to use dimensionality reduction algorithm on a group of content based profiles [6]. This allows performance improvements since it reduces the amount of preferences/features that must be compared to generate the recommendations.

CONCLUSION

Recommender frameworks permit web based business destinations to be profoundly adaptable for the client and purchaser. They permit organizations to better grasp their clients, give customized stores, and thus increment consumer loyalty and steadfastness. They are executed by using different existing information mining instruments and adjusting them to current necessities. Famous methodologies incorporate utilizing affiliation rules, cooperative separating and content-based sifting and crossover separating. Suggestions utilizing affiliation rules are created in view of past exchanges the client has proactively shown interest in. Cooperative separating permits the dynamic client to get suggestion in light of items that clients with comparable premium have bought and evaluated emphatically, and by utilizing the dynamic client's past appraisals and exchange history to construct a model that gives another arrangement of comparable items. Content-based separating analyzes the client's very own profile and inclinations with the data set to find items that are of interest and line up with the dynamic client and present them. Suggestions can go from being customized to local area driven and consider a great many prospects. The suggestions are likewise being invigorated because of the idea of changing hunt history, evaluations, and appearance of new items. This likewise presents many difficulties which incorporate virus start, taking care of mysterious clients, making a social recommender framework that can oblige more than one dynamic client, taking care of different various information sources and versatility with expanded information.

REFERENCES

- [1] D. Goldberg, D. Nichols, B. Oki and D. Terry, "Using Collaborative Filtering to Weave an Information Tapestry," in ACM, New York, 1992.
- [2] C.-P. Wei, M. Shaw and R. Easley, "A Survey of Recommendation Systems in Electronic Commerce," National Sun Yat-Sen University, Kaohsiung,, 2001.
- [3] B. Schafer, J. Konstan and J. Riedl, "Recommender systems in e-commerce," in ACM, New York City, 1999.
- [4] B. Sarwar, G. Karypis, J. Konstan and J. Riedl, "Analysis of recommendation algorithms for e-commerce," in ACM, New York, 2000.
- [5] J. Pine, Mass Customization, Boston: Harvard Business School Press, 1993.
- [6] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," Knowledge and Data Engineering, IEEE Transactions, vol. 17, no. 6, pp. 734-749, 2005.
- [7] M. Claypool, P. Le, M. Waseda and D. Brown, "Implicit Interest Indicators," in ACM, New York, 2001.
- [8] A. N. Regi and R. Sandra, "A Survey on Recommendation Techniques in E-Commerce," vol. 2, no. 12, 2013.
- [9] R. J. Brachman, T. Khabaza, W. Kloesgen, G. Piatetsky-Shapiro and E. Simoudis, "Mining business databases," in ACM, New York, 1996.
- [10] B. Schafer, J. Konstan and J. Riedl, "E-Commerce Recommendation Applications," in Applications of Data Mining to Electronic Commerce, Minneapolis, Springer US, 2001, pp.

115-153.

- [11] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," in 20th International Conference on Very Large Databases, Santiago, 1994.
- [12] J. S. Breese, D. Heckerman and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," in Morgan Kaufmann Publishers, San Francisco, 1998.
- [13] X. Amatriain, A. Jaimes, N. Oliver and J. M. Pujol, in Recommender Systems Handbook, New York City, Spring US, 2011, pp. 39-71.
- [14] W. Lin, S. Alvarez and C. Ruiz, "Efficient Adaptive-Support Association Rule Mining for Recommender Systems," Data Mining and Knowledge Discovery, vol. 6, no. 1, pp. 83-105, 2002.
- [15] B. Sarwar, G. Karypis, G. Konstan and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in ACM, New York City, 2001.
- [16] G. Salton, Automatic text processing: the transformation, analysis, and retrieval of information by computer, Boston: Assison-Wesley Longman Publishing Co., 1989.
- [17] N. Belkin and B. Croft, "Information filtering and information retrieval: two sides of the same coin?," vol. 35, no. 12, 1992.
- [18] B. Sheth and P. Maes, "Evolving agents for personalized information filtering," Orlando, 1993.
- [19] M. Balabanovic and Y. Shoham, "Fab: content-based, collaborative recommendation," Communications of the ACM, vol. 40, no. 3, pp. 66- 72, March 1997.
- [20] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes and M. Sartin, "Combining content-based and collaborative filters in an online newspaper," in ACM, Berkeley, 1999.
- [21] T. Tran and R. Cohen, "Hybrid Recommender Systems for Electronic Commerce," in AAAI Workshop, 2000.
- [22] E. Suh, S. Lim, H. Hwang and S. Kim, "A prediction model for the purchase probability of anonymous customers to support real time web marketing: a case study," Expert Systems with Applications, vol. 27, no. 2, pp. 245-255, 2004.