A Review of collaborative filtering Recommendation System

Sura I. Mohammed Ali[1] and Sadiq Sahip Majeed[2]

Department of Mathematics and Computer Applications,

Collage of Science, Al-Muthanna University, Iraq

Emails: suraibraheem@mu.edu.iq

Corresponding Emails: [1]suraibraheem@mu.edu.iq and [2]sadiqss1983@mu.edu.iq

Received 26-04-2021, Accepted 10-05-2021, published 12-05-2021.

DOI: 10.52113/2/08.01.2021/120-131

Abstract: Recommended systems, also known as systems of recommendation, are a part of information filtration systems which are utilized to predict the user's estimation or choice for an object. In recent years, recommended systems have been extensively used in e-commerce programs. Music, news, books, research papers, and goods are likely to be the most popular E-commerce pages. This article provides an analysis of the scope of recommendation systems and discusses recommended systems that include Collaborative filtering (CF), one of the farthest common recommended methods, which are typically divided into three major categories: Approaches to recommendation that are content-based, collective, or hybrid.

Keywords: Recommender System, Collaborative Filtering, Content-based, Hybrid Recommendation Approach.

I. INTRODUCTION

The growing amount of heterogeneous knowledge available on the Internet has made it difficult to suggest specific products that match the needs of end users. After the invention of the first recommender systems (RS) in middle of -1990s[1,2, 3], it has drawn theawareness of researchers and has become an important research field.

Recommender systems are software that may exist as a standalone or as part of a larger scheme a part of a website Recommender programs use product knowledge and the views of group members to help individuals in the community determine the items that are ultimate probable to be motivating to them or applicable to their requirement. The primary objective of recommender systems is to assist consumers in finding helpful objects by automatically reducing the space of alternatives in order to become more suitable for users. The origins of RS, according to Adomavicius and Tuzhilin[4], can be backwards to works in, knowledge processing, cognitive science ,approximation theory, forecasting theories, management science, and customer elite modeling in marketing. Recommender systems are commonly classified into many kinds. These forms are categorized based on the method used to calculate the recommendation[5,6]. There are various types of recommender systems described in the literature, including collaborative-based, mixed, content-based knowledge-based, social, and semantics-based recommender systems. This article conducts a literature analysis on the Collaborative filtering (CF) suggestion methodology. To recommend appropriate products to consumers, recommender schemes typically use two approaches: filtering of content-based and filtering of collaborative [7]. In the filtering of content-based, the user's knowledge is gathered to create a profile, and then any products purchased are recommended). As a result, content-based filtering recommends products to people with identical profiles contents to the recommended items score is the most widely used tool for gathering customer feedback. The benefit of filtering of content is that it will suggest previously unrated products to consumers with specific preferences while also providing reasons for the recommendations [8]. Collaborative filtering, on the other hand, is commonly used in recommender schemes and is the important recommendation method to date[9].

II. Collaborative filtering (CF)

Collaborative filtering (CF) is a method of filtering objects based on the views of other people. In essence, the CF algorithms use a series of ratings to predict the ranking that a target consumer u will assign to an object i. This rating collection includes all ratings r for an object i got by certain users. Typically, for each person, a collection of "nearest neighbor" users whose previous scores have a strong degree of similarity is discovered scores for unknown objects are predicted using a composite of known scores from nearby neighbors. The definitions of objects are not needed in the collective filtering approach. The calculation of the forecast of the N-top suggested products is based on the scores of the target user's closest neighbors. As a result, identifying a target user's closest neighbors is an important phase in CF. Collaborative filtering methods can be classified in many ways Some studies [4, 5]classified them as memory-based (or heuristic-based) and model-based methods, while others [10] classified them as probability-based and non-probability-based methods. The cold-start problem, which occurs when historical data is too scarce (called as sparsity problem), new users have not scored enough objects (known as the new user problem), or both, is a major issue restricting the effectiveness of filtering of collaborative in some purpose domains.

Different of content-based filtering , a collaborative filtering is predicated on the premise that individuals who have similar predilections on certain products will have similar preferences on other items [11], the function of this method is to identify people with common relationship, and it provides suggestions based on the interests of these "similar neighbors" [9]. The basic premise of this method is that individuals with identical tastes would rank objects similarly.

III. Limitation of Collaborative filtering

Cold-start issue: Other users in the system must be stuffed in order to find a contest. Collaborative filtering is fully dependent on similar neighbors in the system; however, if these similar neighbors are not present in the system during the arch process, this is referred to as the cold-start problem. The hybrid solution will help to prevent this problem.

Data sparsity: When there are a vast number of items to recommend, the user/ratings matrix becomes sparse, making it difficult to locate users who have rated the relevant items. Using shared and distinct methods, recommender frameworks typically create user neighborhoods based on their profiles. If a customer has only reviewed a few things, it is difficult to assess his or her taste and relate to him or to the incorrect neighborhood Sparsity is a problem caused by a lack of detail.

Scalability: recommendations for various situations in which consumers and products occur As a result, vast amounts of computing power are required on a regular basis to compute recommendations.

Filtering of Memory-based collaborative and filtering of model-based collaborative are the two types of filtering of collaborative techniques.

- Memory Based approach: to create recommendations, memory-based CF approaches usually use a ranking matrix to manage a user-item repository. In general, memory-based shared filtering techniques use neighborhood item datasets to gather user attention, and is intended to be used in the future with all ratings by indicating to users or items whose ratings are identical to the other user or items[12]. Memory-driven approaches measure user and object commonalities based on whether they are entity or user-based strategies. They are dependent on their neighbors. They are referred to as the "k Nearest Neighborhood method (KNN)." KNN filtering recommends objects to the customer based on resemblance procedures . It is the most widely used method, and it implements the three phase to generate recommendations.

1. Locate similar users (neighbors) to user a.
2. Use an optimization technique,
3. Use the top N suggestions from phase 2.

Memory-based techniques are splited into 2 types: user-based and item-based approaches.

1. User Based Approach: Recommendations are made to users based on the assessment of objects by other users of the same category with which he or she shares similar interests. If the item was already rated by an individual on the street, the customer would like it.

2. Item Based Approach: examines the range of items rated by the home person, computes their resemblance to the prospective object, and then selects the most comparable items At the same time, their reflecting similarities are calculated. Following the

discovery of the most comparable items, the forecast is determined by taking a weighted mean of the target customer's scores on these similar items [4].

- Model Based approach: the key disadvantage to memory-based collaborative filtering strategy is that it manages whole datasets connected to user object datasets, which causes this method to run slower than most collaborative systems and causes scalability issues. When generating real-time entries in the recommendations software log, there is a problem. Researchers propose model-based recommendation systems to address these issues.

- Model-based recommendation systems: This model is designed by extracting certain information from a large database connected to a certain parameter/attribute and using this model every time instead of using the whole database; as a result, the model improves both the scale and speed of the recommendation system[13].This method uses less memory and takes less time to process. This method allows the structure to be visualized more effectively and efficiently. This technique allows the machine to visualize more effectively while still reducing error. There are some methods for locating concealed (latent) elements. MF (Matrix factorization) and SVD (Singular value decomposition) are the most widely used methods.

IV. Measures of Similarity

In user-based CF similarity, the differ of user-based similarity and item-based similarity is determined. The several functions for detecting correlations are there, including the well-known cosine-based function, Pearson correlation, , Manhattan distance and Jaccard coefficient. All of these functions can be used as long as they have the relevant input style within the relevant domain and restoration a rate that shows for higher values, there is a high degree of similarity [13, 14]. Data interpretation defines similarity in terms of a function of distance. The function of distance might be computed using either the Manhattan distance or the Euclidean distance.

$$\mathrm{Dist}_{x,y} = \sqrt{\sum_{k=i}^{m}(x_{ik-}x_{jk})^2} \text{ [10]}$$

$$\mathrm{Dist}_{x,y} = \left|x_{ik-}x_{jk}\right|$$

Similarity Based on Cosine It is a measure of resemblance between two vectors in an interior product space that computes the cos. of the angle among them. The cos. of 1 is 0, and it is smaller at every different angle.

$$\cos\theta = \frac{x.y}{\|x\|\|y\|} = \frac{x_1 * x_2 + y_1 * y_2}{\sqrt{x_1^2+y_1^2}\sqrt{x_2^2+y_2^2}} \text{ ... [10]}$$

Pearson Coefficient Correlation Pearson correlation can also be used to measure the resemblance between any two vectors. The result varies from -1 to 1. Here, 1 and -1 indicate that they are highly correlated. If the answer is yes, then they are fixedly connected. If -1, they are inversely connected. In a passive linkage, the value of one variable diminution as the value of the other vector increases. The degree of correlation is decreasing towards zero.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \text{...[10]}$$

Formula for foretelling calculated as the measured medium of variation from the neighbor's mean; using the concept of weighted number, there can estimate the ranking for the entire user-item pair. First, we need to find all the objects that are close to our target object, and then we keep the things that the successful consumer has rated. and use the similarity b to weight the user's ranking on each of these objects between that and the desired object in final, scale the forecast by the number of similarities to arrive at a fair expected ranking score.

$$k_{x,y} = sim(u_x, u_y) =$$

$$\frac{\sum_{h=1}^{\dot{n}}(r_{u_x,h_i} - \bar{r}_{u_x})(r_{u_x,h_i} - \bar{r}_{u_y})}{\sqrt{\sum_{h=1}^{\dot{n}}(r_{u_x,h_i} - \bar{r}_{u_x})^2}\sqrt{\sum_{h=1}^{\dot{n}}(r_{u_x,h_i} - \bar{r}_{u_y})^2}}$$

…[10]

V. LITERATURE REVIEW

The most widely used collaborative filtering algorithms for recommendation are neighborhood-based. A traditional community-based collaborative filtering system's process divided into three stages: input data, neighborhood creation, and recommendations generation.

- The initial input data is a list of n customers' past buying purchases on m goods. To fix the scalability, sparsity, and synonymy problems, the original m * n representation can be converted into the decrease dimensional impersonation.
- Neighborhood creation: For an active user, the correlations among all other users and the live user are calculated to compose a proximity-based neighborhood for the active one with a volume of likeminded users. There are several methods for calculating the similarity of two items Pearson's correlation coefficient, Spearman's rank correlation coefficient, restricted Pearson's correlation coefficient, cos. resemblance and mean-squared discrepancy are some examples.
- Recommendation generation: The last step is to create suggestions based on the interests of the active user's closest neighbors, for example, propose two separate strategies for completing this mission. They are the most common item advice and the connected rule-based recommendation.

This paper is a systematic analysis that includes an in-depth examination of current research on the classical CF-based advice approach. We explore some possible CF concerns and highlight prospective study avenues for addressing data sparsity and cold start. In[14] objects excerpt uses for recognizing similar among items, offer an efficient way of output value set rules with higher outputs based on a genetic algorithm to boost the standard of recommendation Evaluations were carried out on the MovieLens data collection. The genetic algorithm-Association Rule Mining is generated by the recommender scheme (GARM). The laws of association are discovered in the form of (I_1, I_2) $\rightarrow I_t$. If the target customer enjoys items I_1 and I_2, he or she would like item It as well. There are two standards for assessment. The first is the overall rule quality of each iteration. The other is execution time. To assess the influence of the parameters, 100 target items were chosen at random.

[15] implemented a hybrid system that employs various recommendation strategies to get better the consistency and preciseness of the recommendation process. The BookCrossing dataset was evaluated, and it was used in all experiments. This dataset decomposes a community of book lovers who exchange books and share their experiences all over the world. This paper introduces a novel semantics-based recommendation system that combines the benefits of knowledge-based, content-based, and collaborative filtering recommendation techniques. As a result, an ontological user profile showing user interest is generated and correlated with the objects domain ontology. The enhanced spreading activation algorithm is used to update user interest for various objects based on the Euclidean distance between the vectors of interest scores in the users' profiles.

Also,[16] to address some of the shortcomings in CF frameworks, we added a shared filtering approach to improve the suggestion accuracy extracted from user-created tags. Collaborative labeling is used as a method to understand and filter consumer expectations for objects. Furthermore, they investigate the benefits of mutual tagging for data sparsity and a cold-start user. These applications represent significant difficulties in collaborative filtering. Figure depicts a system with two phases: a model construction phase and a probabilistic recommendation phase. During the model building process, we first use a shared filtering scheme to produce the latent tags that would be of interest to a target customer, which is referred to as a Candidate Tag Set (CTS). According to the CTS, to evaluate the products to suggest stochastically, a naive Bayes technique is used.

In[17] proposed a method for modeling consumer expectations using item domain features and combining them with CF for customized suggestion The UPM-B-IDF (modeling user preferences matrix based on item domain features), UPV (modeling user preferences vector), and CF-B-UCM algorithms comprise this structure (CF based on the user preference model). The UPM-B-IDF algorithm aims to model the user preferences matrix using object domain features; the UPV algorithm aims to derive the user preferences vector from the user preferences matrix; and the CF-B-UCM algorithm aims to provide customized suggestions by combining user preference models with collective filtering. This approach not only incorporates domain features into customized recommendations, but it also helps in identifying latent customer associations that the traditional CF process misses. Their approach yields improved results, demonstrating that the consumer choice model is more efficient for suggestion.

VI. CONCLUSION:

Collaborative filtering is one of the key tools that will power the adaptive web, and it will be used to get better the efficiency and thoroughness of the recommendation process.. In this paper, we tried to provide a snapshot of existing knowledge about interactive filtering systems and processes. When massive amounts of information become more widely accessible, collaborative filtering may become essential. Throughout the method, we can obtain a better understanding of the mechanisms of collective filtering.

REFERENCES

[1] K. Inagaki et al., ‘A case of hepatitis E virus infection in a patient with primary biliary cholangitis’, Acta Hepatol. Jpn., vol. 58, no. 3, pp. 183–190, 2017, doi: 10.2957/kanzo.58.183.

[2] U. Shardanand and P. Maes, ‘Social information filtering’, pp. 210–217, 1995, doi: 10.1145/223904.223931.

[3] W. Hill, L. Stead, M. Rosenstein, and G. Furnas, ‘Recommending and evaluating choices in a virtual community of use’, Conf. Hum. Factors Comput. Syst. - Proc., vol. 1, pp. 194–201, 1995, doi: 10.1145/223904.223929.

[4] G. Adomavicius and A. Tuzhilin, ‘Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions’, IEEE Trans. Knowl. Data Eng., vol. 17, no. 6, pp. 734–749, 2005, doi: 10.1109/TKDE.2005.99.

[5] G. Ali and M. El, ‘A SEMANTICS-BASED FRAMEWORK FOR RECOMMENDER SYSTEMS By’, no. December, 2014.

[6] J. S. Breese, D. Heckerman, and C. Kadie, ‘Empirical Analysis of Predictive Algorithms for Collaborative Filtering’, pp. 43–52, 2013, [Online]. Available: http://arxiv.org/abs/1301.7363.

[7] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, ‘Recommender systems survey’, Knowledge-Based Syst., vol. 46, pp. 109–132, 2013, doi: 10.1016/j.knosys.2013.03.012.

[8] B. Hu and Y. Zhou, ‘Content Semantic Similarity boosted Collaborative Filtering’, Proc. - 2008 Int. Conf. Comput. Intell. Secur. CIS 2008, vol. 2, pp. 7–11, 2008, doi: 10.1109/cis.2008.177.

[9] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, ‘Item-based collaborative filtering recommendation algorithms’, Proc. 10th Int. Conf. World Wide Web, WWW 2001, pp. 285–295, 2001, doi: 10.1145/371920.372071.

[10] B. J. Schafer, D. Frankowski, J. Herlocker, and S. Sen, 'Collaborative Filtering Recommender Systems - CollaborativeFilteringRecommender Systems.pdf', Lncs, vol. 4321, no. January 2007, pp. 291–324, 2007, [Online]. Available: http://www.eui.upm.es/~jbobi/jbobi/PapersRS/CollaborativeFilteringRecommenderSystems.pdf.

[11] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl, 'Recom Syst', Commun. ACM, vol. 40, no. 3, pp. 77–87, 1997, [Online]. Available: http://portal.acm.org/citation.cfm?id=245108.245126.

[12] P. B.Thorat, R. M. Goudar, and S. Barve, 'Survey on Collaborative Filtering, Content-based Filtering and Hybrid Recommendation System', Int. J. Comput. Appl., vol. 110, no. 4, pp. 31–36, 2015, doi: 10.5120/19308-0760.

[13] P. Nagarnaik and A. Thomas, 'Survey on recommendation system methods', 2nd Int. Conf. Electron. Commun. Syst. ICECS 2015, no. Icecs, pp. 1496–1501, 2015, doi: 10.1109/ECS.2015.7124835.

[14] B. S. Neysiani, N. Soltani, R. Mofidi, and M. H. Nadimi-Shahraki, 'Improve Performance of Association Rule-Based Collaborative Filtering Recommendation Systems using Genetic Algorithm', Int. J. Inf. Technol. Comput. Sci., vol. 11, no. 2, pp. 48–55, 2019, doi: 10.5815/ijitcs.2019.02.06.

[15] G. Ali and A. ElKorany, 'Semantic-based collaborative filtering for enhancing recommendation', KEOD 2014 - Proc. Int. Conf. Knowl. Eng. Ontol. Dev., no. December, pp. 176–185, 2014.

[16] H. N. Kim, A. T. Ji, I. Ha, and G. S. Jo, 'Collaborative filtering based on collaborative tagging for enhancing the quality of recommendation', Electron. Commer. Res. Appl., vol. 9, no. 1, pp. 73–83, 2010, doi: 10.1016/j.elerap.2009.08.004.

[17] J. Zhang, Q. Peng, S. Sun, and C. Liu, 'Collaborative filtering recommendation algorithm based on user preference derived from item domain features', Phys. A Stat. Mech. its Appl., vol. 396, pp. 66–76, 2014, doi: 10.1016/j.physa.2013.11.013.