ISSN: 2277-9442 Vol. 5, No.2 December, 2020, Netherland

International Journal of Applied Science & Engineering Research

Taxonomy Paper on Recommendation Systems: A Review

Dr. Stuti Asthana

Independent Researcher and Analyst, UT Administration of Dadra and Nagar Haveli and Daman and Diu

Date of Submission: 12th August 2020 Revised: 26th October 2020 Accepted: 18th December 2020

***Abstract*-- Today with the coming of digitalization there is a tremendous growth in the use of an efficient recommendation System. The major impact of recommendation system is seen in the field of digital marketing where several approaches to offer new and better things to customers have came into role. As there are vast number of internet users and easy availability of internet there is a rapid increase in the recommendation system handles this difficulty in the best possible manner. There are a number of situations where users come into dilemma what to choose will be best or not with respect to their requirements , in that scenario the options provided by the recommender systems on the basis of previous reviews plays a prominent role in decision making and to filter out their preferences from the myriad possibilities. The present paper shows a review work done on different techniques adopted in recommender system. Today most of the successful companies like Amazon and Netflix are using recommender system to provide ample better opportunities to their customers.**

***Keywords*: recommendation system, digitalization, content based filtering, collaborative filtering, Netflix**

INTRODUCTION

In olden times there was a decline to purchase items from Internet, due to poor bandwidth, inefficient transactions etc. but with time and by the coming of recommendation system there was a vast increase in information according to the interest of customers, products and transactions . Further in order to supply personal services to customers personal recommender systems with techniques of recommendation have been widely adopted. The main advantage of personal recommender system was that it enhances the search efficiency by reducing the searching time for interesting items. On the basis of preferences, activities, behavior etc. recommendations are provided by the recommendation system and assist them in taking correct decision. Also the prediction to ratings and preference are given for an item [1]. Whenever there are several options for the users and user have to select a particular thing and take decision then recommendation system are adopted. This can be achieved by monitoring the users' past purchasing history and transactions of individuals of similar interest.

Amazon is an early leader in enhancing its e-commerce sector by implementing the principle of the recommendation system [2].

Section 2 of the paper deals with the description of the types of suggested systems. It addresses both the acquisition and presentation of the profile of consumer interest and the profiles of goods. Section 3, discuss the work being done by different researchers and their outcome and technique used. Section 4, discusses the mathematical equations and expressions involved in recommendation systems. Finally we discuss the future research and conclusion of our paper.

RECOMMENDER SYSTEMS

This section discusses the techniques adopted in recommender systems. The figure below illustrates the different steps involved in the framework of recommendations. We obtain the data from the resources initially. The consolidated data from various databases, data cubes, archives, etc are accessed here and the work is performed here to eliminate inconsistencies. Further for selection of data we use the different feature selection and extraction techniques. As the data comes from heterogeneous sources there are possibilities for noise, missing values and inconsistencies, so it is mandatory to preprocess the data in order to get correct data. At the preprocessing stage these task are performed:

• *Data Cleaning*- Data cleaning means filling in the missing values. Smoothing techniques are used to convey meaningful information to noisy data. The noise function is generated in order to remove useful information from data and to disregard it.

Data Transformation- We use the technique of discretization of data or generation of hierarchy for the transformation. To transform expression data into finite value, we use discretization of data. Further in order to obtain the dersired result. We use suitable techniques for data mining. For generating suggestion to user our technique uses data mining technique. A graded list of things is recommended to a consumer by the recommender.

For a given item rating a preference score is adopted [3] to forecast it. Recommendation strategies can be typically classified into collaborative filtering[4], content-based filtering[2], knowledge-based system[5], hybrid systems, etc.

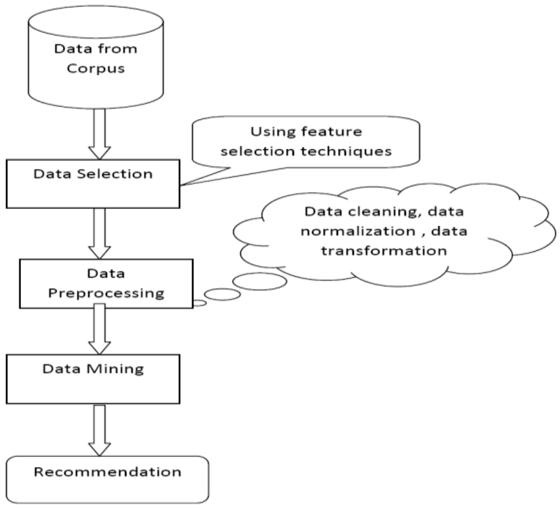

Fig: 1 Steps in Recommendation system

A. Content Based Recommendation –

Here the decision on the most popular item for the user is made according to his interest by analyzing item descriptions [6]. A list of item profiles is created on basis of data provided by the user. This uses a word frequency TF metric and the IDF. To define the value of the object, the TF*IDF brand, commonly referred to as the term weight, is used. The term frequency defines how many times an object in a document occurs. It's given mathematically by:

TF (t) = No. of times term 't' comes in a document / Total No. of terms in the document

$$TF_{i,j} = \frac{f_{i,j}}{\mathbf{Max_z f_{z,j}}} \qquad \textbf{............ (1)}$$

An object's value is determined by the IDF.

IDF (t) = log_e (Total No. of documents / No. of document with term t in it)

$$IDF_i = \log \frac{N}{ni} \qquad \textbf{............ (2)}$$

Term Weight is gives as :

$$W_{i,j} = TF_{i,j} \text{ X } IDF_i \qquad \textbf{............ (3)}$$

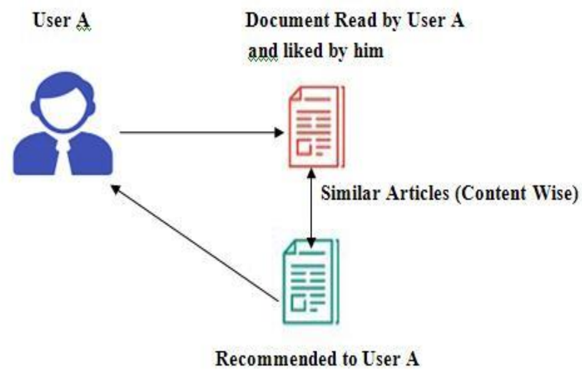

Fig 2: Content based recommender system

For knowing the preference of an item to a user we use the vector space model after the calculation of TF*IDF. A vector is generated for item and its attributes in VSM. To define the similarity between consumer and product, we use the cosine angle between the vectors. For content based recommendation correlation based approach is used [2].

B. Collaborative filtering based recommendation

It works on the preference for the same item for similar type of users. Here the main technique of KNN algorithm is deployed. If a user set has the strongest association in the past, the nearest neighbor is found. Scores of new objects are estimated on the basis of the nearest neighbor's scores[7]. In order to identify the desired items for the user we use the Pearson correlation or Log-likelihood ratio. We define Pearson correlation as [8].

$$sim(a,b) = \frac{\sum_{p \in P}(r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P}(r_{a,p} - \bar{r}_a)^2}\sqrt{\sum_{p \in P}(r_{b,p} - \bar{r}_b)^2}}$$

Where , a and b are users

$r_{a,p}$ = rating for item p by user a

P = items set read by both a and b

similarity values lies between -1 and +1

While content-based recommendations focus on item attributes as well as user attributes, item-based recommendations are based on a rating provided to a specific item by a user.

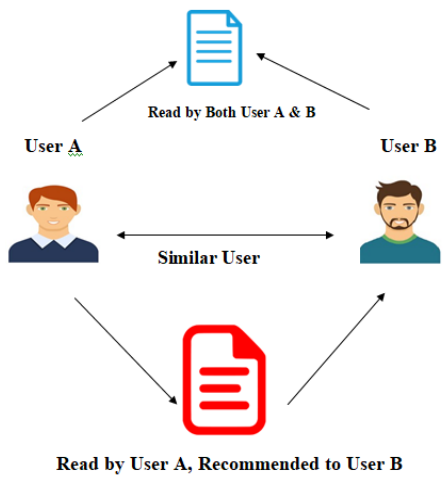

Fig 3: Collaborative based recommender system

C. Knowledge Based Recommendation System

It operates on the basis of knowledge of the needs of the consumer, his interests, etc., which forms a recommendation basis. Its basis is the perception of the need of the consumer for a specific object [7].

D. Hybrid Recommendation Sytem

We combine the features of different recommender systems to get a model known as the hybrid recommendation system[7], to enhance the accuracy of recommendations. This approach blends the scores of many methods of recommendation together. Via function combination. The Cascade Advice uses a variety of suggestions. Function augmentation uses output as an input feature to another from one technique.

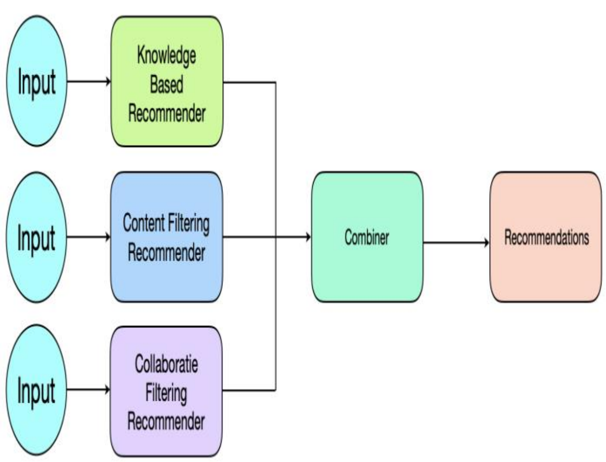

Fig 4: Hybrid based recommender system

A major challenge to be dealt with collaborative filtering is cold [9, 10–12]. It comprises of cold user & cold item. The former arises when a new user register himself on a website whereas the later arises when there is a new user.

Whenever a new item arises and there is insufficient information about it, then it is termed as cold item.

Enhancing and improving the output of the recommender system concerns the most towards the study of the recommender systems. The users who rated items in the scale of 2 to 20 items are the cold users [12]. Many techniques which focus on techniques for improving the recall and precision of the recommender system are discussed.

Related Work

The present section deals with the works related to recommendation system by various authors:

In his dissertation, **Mei-Hua Hsu** used a combination of two techniques, namely a collaborative filtering technique and a content-based approach for students' customized english learning recommendation system[13]. The clustering method is used for the grouping of students into different subjects. Finally, the rule of association is used to produce suggestions for different lessons.

Nguyen, Lucas, Artus and Lars used educational data for predicting student performance in Recommender system in their research work [14]. For generating recommendation and to validate their approach they used matrix factorization technique and logistic regression.

Boticario, Olga and Jesus used an approach that can be used in e-learning scenarios & their work as customized recommenders [15].

All the above related work didn't focused for generating recommendations if size of data was big. So in order to meet with the challenges of increasing data at present techniques like hadoop, big data, mahout have to be used to generate recommendations.

Sarwar et al. (2001) [16] told that there are 2 divisions of collaborative filtering namely memory-based and model-based algorithms for collaborative filtering.

Memory-based algorithms are based on the concept of all past experiences that the user had in purchasing a given item and to generate prediction. Most related user similar to target user is found on the basis of data. Since they have statistically common interest hence these neighbors are similar. Several statistical techniques are used for it and finally topmost items are recommended.

Model-based collaborative filtering are based on building model based on rating given by the user. Rule based and Bayesian network approaches are used for this technique such as clustering. Collaborative filtering such as classification problem is formulated by the clustering model while rule-based model as an association-rule model and the Bayesian network model treats it as a probabilistic model. Sometimes these techniques are also termed as item-based collaborative filtering algorithms.

Lops et al. (2011) [17] stated that comparing the attributes of a content object with that of a user profile is the basis for recommendation of a content based recommender system.

User's interest in an object is the outcome of this process. For a content based model to give accurate results it is necessary that the user profile should be accurate. As stated by Lika et al., the limitation of collaborative and content-based filtering is the issue of handling new users or items (2014)[18].

Burke (2007) [19] In his research, he proposed that the cold-start problem could be solved by hybrid systems. In his research, he merged content-based and collaborative model characteristics to achieve an advanced hybrid recommendation scheme. An ensemble is the technique of integrating several algorithms into a single one. Hybrid filtering is the common ensemble method in the area of recommender systems. Burke defined hybrid recommender system as the one which combines multiple recommendations into a single model (2007).

Bushra Alhijawi et al. [20] discussed a novel genetic-based recommender system (BLIGA) based on semantic data and historical rating data was addressed. (BLIGA) findings were compared against the results of recommendations from alternative methods of collective filtering.

Sneha V. et al.[21] suggested that the approach that recommends items has been solved by identifying users who have similarities in ranking items by analyzing the user-item ratings matrix. Only a small number of items are rated by the user in the list which results in high sparsity which makes it difficult to identify similar users and recommend items reliably. In order to find ratings of unrated items we use Collaborative filtering on the basis of the user's rating history and similarity to other users.

Jianrui Chen et al. [22] claimed that making use of obsolete sources of information is the the major challenge in the recommendation process. For solving this problem the current paper proposes an efficient time-weighted collaborative filtering algorithm. The new matrix that was discovered expresses many users' desires. Consumers and higher objects, then Correlations are classified into the same category according to differential equations. Stable values of the user state mean that they have similar goals and are distributed to the same population afterwards. Finally dynamic similarity measurements are used to obtain the results of real-time prediction.

Mathematics involved in traditional recommendation system are as follows:

In the beginning we find the similarity between users and target users (cold users). The technique which is best suited for it is the cosine similarity which is given as follows:

$$\mathrm{Sim}(x,y)^{COS} = \frac{\sum_{p \in I} (r_{x,p} \times r_{y,p})}{\sqrt{\sum_{p \in I} r^2_{x,p}}} \quad \text{….. (4)}$$

Where p= item Here

'I'= set of common rating items by user 'x' and 'y'.
'$r_{x,p}$' and '$r_{y,p}$' = ratings

NHSM composite is another enhancement in similarity measure. [16]. An enhanced feature of this is that it considers local context information as well as global preference of the behavior of his user. Similarity measures of NHSM are formed by the combination of three modified techniques Jaccard, user rating preference (URP) and proximity-significance-singularity. The equation for it is given as :

$$\mathbf{Sim(x,y)^{NHSM} = Sim\,(x,y)^{JPSS}} \times \mathbf{Sim\,(x,y)^{URP}} \quad \text{…… (5)}$$

JPSS is generated by a modified Jaccard PSS (proximity-significance-singularity) and is determined as:

$$\mathbf{Sim\,(x,y)JPSS = Sim\,(x,y)PSS \times Sim\,(x,y)Jaccard} \quad \text{…(6)}$$

$$\mathbf{Sim\,(x,y)^{Jaccard}} = \frac{|I_x \cap I_y|}{|I_x \cup I_y|} \quad \text{…..(7)}$$

Where 'Ix' and 'Iy' are all the things rated respectively by 'x' and 'y' users. PSS is the updated mode of proximity-impact-popularity similarity tests (PIP). PSS consists of a combination of three proximity-significance-singularity parameters, which are measured as

$$\mathbf{Sim\,(x,y)^{PSS}} = \sum_{P \in I} \mathbf{PSS}\,(r_{x,p}, r_{y,p}) \quad \text{….(8)}$$

$$\mathbf{PSS}\,(r_{x,p}, r_{y,p}) = \mathbf{Proximity}\,(r_{x,p}, r_{y,p}) \times \mathbf{Significance}\,(r_{x,p}, r_{y,p}) \times \mathbf{Singularity}\,(r_{x,p}, r_{y.p}) \quad \text{….(9)}$$

Here Proximity shows difference between two ratings, whereas Significance depicts distance from the median rating.

$$\text{Proximity}\,(r_{x,p}, r_{y,p}) = 1 - \frac{1}{1+\exp(-|r_{x,p} - r_{y,p}|)} \quad \text{...(10)}$$

$$\text{Significance}\,(r_{x,p}, r_{y,p}) = \frac{1}{1+\exp(-|r_{x,p} - r_{med}| - |r_{y,p} - r_{med}|)} \quad \text{…..(11)}$$

Where r_{med} = median range of ratings

EVALUATION CRITERIA

There are a variety of criteria that can be tested in the Recommendation method, the most important of which is Recall & Precision. These parameters determine the consistency of the proposed process.

The mean ratio of test data items that can be observed in the training data list is called Recall. The criteria for recall is determined as:

$$\text{Recall} = \frac{TP}{TP + FN} \quad \textbf{.....(12)}$$

Here, TP = True Positive
TN = True Negative
FP = Fasle Positive
FN = False Negative

The ratio of the recommended items to the items that are currently the precision criterion for evaluating data users [22]. It is computed as:

$$\text{Precision} = \frac{TP}{TP + FP} \quad \textbf{.....(13)}$$

CONCLUSIONS

A vital role is played by collaborative filtering using similarity measures in recommendation systems, the present study shows that the accuracy and recall of the recommendation system can be enhanced for cold users with categorized objects. This paper demonstrates a study of the different methods used in the recommendation system, such as collaborative content-based & hybrid filtering, and the recommendation system's various implementations. It may be possible to determine a quick and efficient method for categorizing objects in future work.

REFERENCES

[1] www.educationaldatamining.org (accessed April 2017).

[2] Adomavicius G., Sankaranarayanan R., Sen S., and Tuzhilin A. "Incorporating Contextual Information in Recommender Systems Using a Multidimensional Approach." ACM Transactions on Information Systems Vol. 23, No. 1, January 2005: 103–145.

[3] Mahout.http://mahout.apache.org/users/basics/algorithms.ht (accessed April 20, 2017).

[4] Herlocker J., Konstan J., Tervin L., and Riedl J. "Evaluating Collaborative Filtering Recommender Systems." ACM Transactions on Information Systems, 2004: 5-53.

[5] Burke, Robin. "Hybrid Recommender Systems: Survey and Experiments." User Modeling and User-Adapted Interaction Volume 12, Issue 4, November 2002: pp 331–370.

[6] Pazzani M.J., Billsus D. "Content-Based Recommendation Systems." The Adaptive Web. Lecture Notes in Computer Science, vol 4321. Berlin, Heidelberg: Springer, 2007. pp 325-341.

[7] Herlocker J., Konstan J., Tervin L., and Riedl J. "Evaluating Collaborative Filtering Recommender Systems." ACM Transactions on Information Systems, 2004: 5-53.

[8] Pazzani M.J., Billsus D. "Content-Based Recommendation Systems." The Adaptive Web. Lecture Notes in Computer Science, vol 4321. Berlin, Heidelberg: Springer, 2007. pp 325-341.

[9] Kim, H.-N.; El-Saddik, A.; Jo, G.-S. Collaborative error-reflected models for cold-start recommender systems. Decis. Support Syst. 2011, 51, 519–531. [CrossRef]

[10] Patra, K.B.; Launonen, R.; Ollikainen, V.; Nandi, S. A new similarity measure using Bhattacharyya coefficient for collaborative filtering in sparse data. Knowl. Based Syst. 2015, 82, 163–177. [CrossRef]

[11] Leung, W.-K.C.; Chan, C.-F.S.; Chung, F.-L. An empirical study of a cross-level association rule mining approach to cold-start recommendations. Knowl. Based Syst. 2008, 21, 515–529. [CrossRef]

[12] Zhang, M.; Tang, J.; Zhang, X.; Xue, X. Addressing cold start in recommender systems: A semi-supervised co-training algorithm. In Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '14), Gold Coast, Australia, 6–11 July 2014; ACM Press: New York, NY, USA, 2004.

[13] Hsu, Mei-Hua. "A personalized English learning recommender system for ESL students." Expert Systems with Applications, January 2008: 683–688.

[14] Nguyen Thai-Nghe, Lucas Drumond, Artus Krohn-Grimberghe , Lars Schmidt-Thieme. "Recommender System for Predicting Student Performance." 1st Workshop on Recommender Systems for Technology Enhanced Learning, Procedia Computer Science 2010: 2811-2819.

[15] Boticario, Olga C. Santos and Jesus G. "Requirements for Semantic Educational Recommender Systems in Formal E-Learning Scenarios." Algorithms, 20 July 2011: 132-154.

[16] Sarwar et al. (2001), "Item-Based Collaborative Filtering Recommendation Algorithms".

[17] Lops et al. (2014), "Content-based Recommender Systems: State of the Art and Trends".

[18] Lika et al., "Facing the cold start problem in recommender systems" (2013).

[19] Burke R. (2007) "Hybrid Web Recommender Systems", In: Brusilovsky P., Kobsa A., Nejdl W. (eds) The Adaptive Web.

[20] Bushra Alhijawi, Yousef Kilani, "A collaborative filtering recommender system usinggeneticAlgorithm",https://doi.org/10.1016/j.ipm.2020.102310, 2020.

[21] Sneha V.1, Shrinidhi K.R.2, Sunitha R. S.3, Dr.Mydhili K. Nair.4, "Collaborative Filtering based Recommender System using Regression and Grey Wolf Optimization Algorithm for Sparse Data", IEEE Conference Record # 45898; IEEE Xplore ISBN: 978-1-7281-1261-9.

[22] Jianrui Chen, Bo Wang, Liji , Zhiping Ouyang , "Personal recommender system based on user interest community in social network model",https://doi.org/10.1016/j.physa.2019.04.197, 2019.

[23] Draz, R., Atlam, E.-S., Tawfik, M.A., Keshek, A. "Non_linear approach for automatic extraction of field association words" (2020) International Journal of Emerging Technology and Advanced Engineering, 10 (6), pp. 6-9.

[24] Díaz, I.R., López, E.S., Góngora, D.M." Development of investigative competence in electrical engineering students" (2020) International Journal of Emerging Technology and Advanced Engineering, 10 (7), pp. 47-52.

[25] Fonseca, D.J., Heim, I. "Development of an automated MATLAB-based platform for the analysis of massive EEG datasets" (2020) International Journal of Emerging Technology and Advanced Engineering, 10 (11), pp. 7-11.

[26] Goel, A., Bhujade, R.K. "A functional review, analysis and comparison of position permutation based image encryption techniques" (2020) International Journal of Emerging Technology and Advanced Engineering, 10 (7), pp. 97-99.

[27] Goel, A., Bhujade, R., "A prototype model for image encryption using zigzig blocks in inter-pixel displacement of RGB value" International Journal of Advanced Trends in Computer Science and Engineering, 2020, 9(5), pp. 6905–6912

[28] Nwankwo, P.C. "An improved method for determining static bottom hole pressure in dry natural gas wells: A computer programming approach" (2020) International Journal of Emerging Technology and Advanced Engineering, 10 (5), pp. 171-175.