Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 7.056

IJCSMC, Vol. 10, Issue. 4, April 2021, pg.113 – 117

A Review on Filtering Techniques used in Restaurant Recommendation System

T. Choenyi; T. Tseyang; S. Choikyong; P. Tsering; T. Gurme
DOI: 10.47760/ijcsmc.2021.v10i04.016

Abstract— These days, recommendation systems are extremely common. Recommendation systems are used in suggesting personalized content and services. These are the following recommendation techniques described in this paper: collaborative-based filtering, Content-based filtering, demographic-based filtering, hybrid filtering, along with some applicational research. The paper points over some of the drawbacks and benefits of these techniques. And How with the use of Hybrid filtering these techniques are combined together to remove these individual drawbacks and improve effectiveness. Further with reviews of some research based on these techniques, this paper studies on restaurant recommendation system, mainly suggesting restaurant to the customer based on similar customer satisfaction rating. This model uses content-based filtering for suggesting.
Keywords-component; Recommendation systems; collaborative-based filtering; Content-based filtering; demographic-based filtering; hybrid filtering

I.INTRODUCTION

The exponential growth of data has resulted in the modern information age. Every day, the number of internet users grows. By 2025, it's expected that 463 exabytes of data will be created globally everyday. Although companies collect these generated datas, more than half of these datas are unused due to the lack of useful tools and of the skills to analyse them along with cost of storage.

Recommendation System is one of the applications where this generated data is used.

A recommendation system is an application that predicts outcomes in a variety of domains across the internet. The internet transmits a vast volume of data and offers a wealth of information about user searches. The information extract from the pattern of previously searched data can be moulded into a user-friendly prediction of relevant data. The system's implementation can be accomplished using a variety of methods. There are broadly four types of recommendation systems:

Collaborative-filtering, Content-based filtering, Demographic-Filtering and Hybrid-filtering [1].

Collaborative-filtering: In this approach, each user's recommendation is based on a comparison of their preferences with those of other users who have rated the product similarly to the active user[1].

Content-based filtering: In this approach, the recommender system recommends products based on their similarities. It recommends based on product association rules among the other products.

Demographic-based filtering: It makes recommendations based on the demographic attribute of the user.

Hybrid Filtering: In this approach, above any two techniques are combined together to form hybrid filtering. It eliminates the weakness of individual filters and combines the strengths of more than two recommender systems.

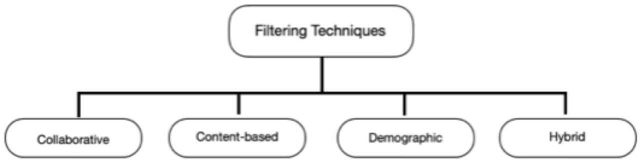

Figure 1. Types of Filtering Techniques

II.LITERATURE REVIEWS

Collaborative-filtering (CF) is the most robust recommendation system and it is often referred to as people-to-people correlation [2]. Its recommended based on action, activities, and preferences of other users who have similar tastes and have similarly rated the product [3]. The collaborative filtering algorithm has two distinct techniques.

Memory-based collaborative filtering uses some or an entire database on user-preferences. Forming clusters of new users with similar tastes and identifying the neighbors. There are further two methods in which neighbors can be identified, User-based Nearest Neighbor, which looks at users' similarity to predict product preference. And Item-based Nearest Neighbor analyses related products that users have reviewed to predict based on users' preference [4]. Amazon also uses Item-based Collaborative filtering for their recommendation system [5].

Model-based collaborative filtering builds a model by extracting information from a dataset, which is then used for recommendation. Markov decision processes (MDPs) are used to solve sequential determination problems and need less memory to run.

Memory-based CF techniques are simple to implement and easy to add new features to the model. Model-based CF increases the algorithm's prediction accuracy. However, Cold-Start: CF systems require a large amount of data to predict recommendations [6]. Scalability: because of Cold-Start CF needs massive computation power to compute recommendations. Sparsity: Only a few popular items in a large e-commerce dataset are rated [7].

Deep neural network are used in recommendation system. This model mainly consist two part: candidate generation and ranking. They used a deep collaborative filtering model with age as one of the input training features, which eliminates algorithmic bias and allows them to generate real-time candidate generation models. They used logistic regression with watch times as weight for ranking, which further generates independent scores for the candidates and orders them based on their scores. The use of deep neural networks in recommendation systems allows for the addition of more features without concern for complexity [8]

Content-based filtering (CDF) recommends products based on their similarities. It works on the principle of "Show me more of what I have liked" [8]. CDF recommends items based on similarity count and uses association between these products to produce related features. CDF can be performed through the following steps: First, the item profile is created by extracting features from items. Second, user profiles are created from the features of items purchased by users. Then, similarity scores between user-profiles and item profiles are calculated to recommend items with the highest similarity scores [9]. This method is commonly used to recommend documents such as news, websites, movies, and books through the keywords from profiles.

The Content-based recommender system provides personalised recommendations by user profile and it also provides transparency, explaining how it works. The limitations of these advocates are similar to items that users have already purchased and it is difficult to generate the attributes for items in certain areas.

There is a Content-Based Citation Recommendation Model. This model is divided into two stages: (i) a fast, recall-oriented candidate selection, (ii) a feature-rich, precision-oriented reranking. The model uses a supervised neural model to embed all available documents in candidate selection (NNSelect) and with a three-layered feed-forward neural network with two ELUs and a sigmoid layer for Reranking Candidates (NNRank). As a result, this model was able to develop a citation recommendation system without relying on the metadata available for the baseline methods [10].

Collaborative Filtering model can be made based on association based K-means clustering. Since it draws on shifts in tendencies, this model employs implicit collective filtering as they derive more of their data from user experience and background. Association rules with clustering assist in defining the rules on how events are related to other events in a group. Collaborative filtering enables identifying purchasing information of similar users for the types of businesses discovered, to generate a recommendation list, based on businesses that the consumer has not attempted [11].

Demographic filtering recommendation is based on the demographic profile of the user. It uses the information provided by the user and is considered to be similar according to demographic parameters such as nationality, age, gender etc [12].

In some cases Demographic recommenders outperform collaborative and content-based filters because the system do not need historical data for recommendation and do not suffer from cold start. Demographic recommender based on [13] uses three Machine Learning approaches to the test: Naive Bayes, Bayesian Network, and SVM. According to the experimental results, three machine learning methods based on demographic information outperformed the baseline method, especially the SVM method. These preliminary results indicate that demographic information alone is inadequate for accurate rating prediction; however, more detailed studies are needed to validate their findings, such as managing unbalanced data and considering other forms of attractions.

Hybrid filtering is a combination of multiple recommending techniques to overcome some of the limitations of the above filtering technique like cold start, overspecialisation, and sparsity problems. Hybrid filtering is also used to improve the accuracy and efficiency of the recommendation system [12].

Basilico and Hofmann [14] were the first to attempt to unify collaborative and content-based filtering. They used the similarity function between user-item pairs to produce a generalisation across either user or item dimensions. San Pedro and Karatzoglou [15] proposed a supervised Latent Dirichlet Allocation model to model deals with collaborative question answering communities with question recommendation. Liu et al [16] proposed to add a virtual profile based on observed user-item interactions in LinkedIn.

Hybrid filtering used with Content-based Collaborative Filtering approach for news topic recommendation. They used rich contexts and users personalised By combining the advantages of the Content-based Filtering approach and the features of the Collaborative Filtering approach [17].

Recommendation Techniques	Definition	Example	Advantages	Problems
Collaborative	This type of recommendation system predicts what would interest a user based on the preferences of many other users.	It assumes that if person A likes burger, and person B likes both burger and fries, Then it is likely that person A will like fries too.	No need for domain knowledge. Capture's the change in user interests over time.	Cold Start. Scalability. Sparsity.
Content-based	This type of recommendation system relies on the products themselves and recommends other products with similar features.	If a user likes the web page of "real Madrid", "PSG", "Bayern Munich", the CBF will recommend pages related to the football.	Independent recommendation without users info. Recommends new to user. Transparency to their active user.	Difficult to generate attributes for the items. Overspecialization. It is hard to acquire feedback from the user.
Demographic	This type of recommendation systems focuses on the demographic of the user.	If a user is in country X then it will likely recommend based on country X.	They are fast and straightforward to obtain result. Overcomes Cold Start and Scalability	Stability. Mainly based on user interest. Information are impractical considering security and privacy.
Hybrid	This type of recommendation uses the combination of any two above system to recommend.		It combines the advantages of these individual systems.	It reduces the problems of individual systems.

Table 1. Quick Comparison between the filtering techniques

In this paper, we are targeting a problem with limited user information. Address, Name, Order, rate, votes, location, cuisines. Of this limitation, we won't be able to apply all the techniques discussed before to build the recommendation system. Therefore, we are promising a novel approach by building a Content-based recommendation system with user ratings and reviews. Thus through the use of a content-based recommendation system, we would predict the restaurant based on the user like or user choice alike restaurant based on ratings, cuisines, and locality.

III.PROPOSED METHOD

The proposed method used in this recommendation system is shown below.

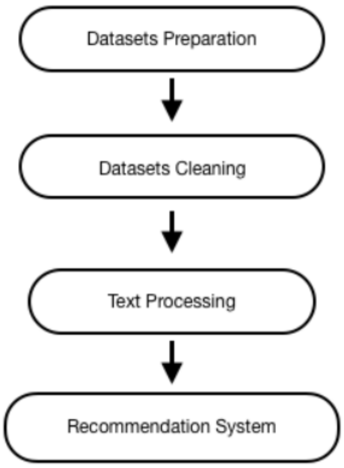

Figure 2. Proposed method

3.1 Dataset Preparation

The dataset used in this paper is obtained from kaggle.com. The dataset consists of restaurants in Bangalore, India collected from Zomato.

3.2 Dataset Cleaning

```python
#Removing the Duplicates
df.duplicated().sum()
df.drop_duplicates(inplace=True)
```

Table 2. After Dataset Cleaning

address	name	online_order	book_table	rate	votes	location	rest_type	cuisines	cost	reviews_list	menu_item	type
942, 21st Main Road, 2nd Stage, Banashankari, ...	Jalsa	True	True	4.1	775	Banashankari	Casual Dining	North Indian, Mughlai, Chinese	800.0	[('Rated 4.0', 'RATED\n A beautiful place to ...	[]	Buffet
2nd Floor, 80 Feet Road, Near Big Bazaar, 6th ...	Spice Elephant	True	False	4.1	787	Banashankari	Casual Dining	Chinese, North Indian, Thai	800.0	[('Rated 4.0', 'RATED\n Had been here for din...	[]	Buffet
1112, Next to KIMS Medical College, 17th Cross...	San Churro Cafe	True	False	3.8	918	Banashankari	Cafe, Casual Dining	Cafe, Mexican, Italian	800.0	[('Rated 3.0', "RATED\n Ambience is not that ...	[]	Buffet

In this stage, we are going to remove the redundant columns, duplicate values, and NaN values. We would also be renaming and feature engineering few columns for better clarity and understanding.

3.3 Text Processing

In this stage, Texts in the dataset which are lowercased, Punctuations, stop-words, URLs are removed. Term Frequency Inverse Document Frequency (TF-IDF) is used for the retrieve information from words with high-frequency are called high TF. For IDF, if the word appears in fewer documents(larger IDF) more weight that word. So, We begin with the most frequent words (TF), then IDF helps in recreate the previous list and get better results. Taking into account the document frequency, we can try to penalise generic words by reducing their relevance. In short, It is used to gather information and look for similarity in the reviews.

3.4 Recommendation System

```python
# Creating tf-idf matrix
tfidf = TfidfVectorizer(analyzer='word', ngram_range=
(1, 2), min_df=0, stop_words='english')
tfidf_matrix = tfidf.fit_transform(df_percent['review
s_list'])
```

For the Recommendation System, we are going to use a content-based recommendation technique. Since is algorithm try to recommend item similar to consumers previous rated or purchased item. It compares scores of candidates generated by association rule among the items and ranked accordingly to the score. Providing the best matching item recommended.

IV. RESULT

We are able to create a Content-Based Recommendation System based on information available in dataset. The Recommender system analyse and studies the reviews of the restaurants in dataset. The System recommends a restaurant for the user based on the restaurant that user selects(write) with similar in reviews and sort the restaurant on the basis of rating.

As shown above, our recommendation system recommend a list of restaurant with similar reviews.

V. CONCLUSION

```
recommend('Ande Ka Funda')
```

```
TOP 9 RESTAURANTS LIKE Ande Ka Funda WITH SIMILAR REVIEWS:
```

	cuisines	Mean Rating	cost
Vasanth Vihar - Since 1965	South Indian, Street Food	3.32	150.0
Sri Krishna Sagar	South Indian	3.26	300.0
Desi Dhaba	North Indian, Chinese, Rolls	3.19	300.0
Ande Ka Funda	North Indian	2.99	200.0
Ande Ka Funda	North Indian, Chinese, Rolls	2.99	200.0
Punjabi Tasty Khana	North Indian, Chinese, Biryani	2.68	450.0
Sri Lakshmi Dhaba	North Indian, Chinese	2.50	250.0
Foodiction	North Indian, Fast Food, Chinese, Burger	2.35	500.0

Recommendation Systems are becoming increasingly popular, with a growing number of websites utilising them. They provide a useful feature that suggests products to users based on their needs and preferences. We discuss all of the different techniques available in the recommendation system in this study. Every technique has advantages and disadvantages. Currently, hybrid recommendation approaches appear to be the most effective and accurate. Because it is a combination of different recommendation techniques, it mitigates some of the limitations associated with a single technique. Although this individual recommender technique is proved to be effective when used with the right data.

REFERENCES

[1]. B.Thorat, P., M. Goudar, R., & Barve, S. (2015). Survey on collaborative filtering, content-based filtering and hybrid recommendation system. *International Journal of Computer Applications*, *110*(4), 31-36. https://doi.org/10.5120/19308-0760

[2]. Ekstrand, M. D., Riedl, J. T., & Konstan, J. A. (2011). *Collaborative filtering recommender systems*. Now Publishers.

[3]. Lee, Y. (n.d.). Recommendation system using collaborative filtering. https://doi.org/10.31979/etd.5c62-ve53

[4]. Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, *7*(1), 76-80. https://doi.org/10.1109/mic.2003.1167344

[5]. Kim, H., El-Saddik, A., & Jo, G. (2011). Collaborative error-reflected models for cold-start recommender systems. *Decision Support Systems*, *51*(3), 519-531. https://doi.org/10.1016/j.dss.2011.02.015

[6]. E., & ., D. (2018). Collaborative filtering: Data sparsity challenges. *International Journal of Multidisciplinary and Current Research*, *6*(6). https://doi.org/10.14741/ijmcr/v.6.6.13

[7]. Covington, P., Adams, J., & Sargin, E. (2016). Deep neural networks for YouTube recommendations. *Proceedings of the 10th ACM Conference on Recommender Systems*. https://doi.org/10.1145/2959100.2959190

[8]. Sharma, R., & Singh, R. (2016). Evolution of recommender systems from ancient times to modern era: A survey. *Indian Journal of Science and Technology*, *9*(20). https://doi.org/10.17485/ijst/2016/v9i20/88005

[9]. Keerthana, T., Bhavani, T., Priya, N., Prathyusha, V., & Sri, K. (2020). FLIPKART PRODUCT RECOMMENDATION SYSTEM. *Journal of Engineering Sciences*, *11*(4), 515-522

[10]. Bhagavatula, C., Feldman, S., Power, R., & Ammar, W. (2018). Content-based citation recommendation. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). https://doi.org/10.18653/v1/n18-1022

[11]. Jooa, J., Bangb, S., & Parka, G. (2016). Implementation of a recommendation system using association rules and collaborative filtering. *Procedia Computer Science*, *91*, 944-952. https://doi.org/10.1016/j.procs.2016.07.115

[12]. Ortega, F., Sánchez, J., Bobadilla, J., & Gutiérrez, A. (2013). Improving collaborative filtering-based recommender systems results using Pareto dominance. *Information Sciences*, *239*, 50-61. https://doi.org/10.1016/j.ins.2013.03.011

[13]. Wang, Y., Chan, S. C., & Ngai, G. (2012). Applicability of demographic recommender system to tourist attractions: A case study on trip advisor. *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*. https://doi.org/10.1109/wi-iat.2012.133

[14]. Basilico, J., & Hofmann, T. (2004). Unifying collaborative and content-based filtering. *Twenty-first international conference on Machine learning - ICML '04*. https://doi.org/10.1145/1015330.1015394

[15]. San Pedro, J., & Karatzoglou, A. (2014). Question recommendation for collaborative question answering systems with RankSLDA. *Proceedings of the 8th ACM Conference on Recommender systems - RecSys '14*. https://doi.org/10.1145/2645710.2645736

[16]. Parker, C. T., & Garrity, G. M. (2003). Exemplar abstract for Wagnerdoeblera megaterium (Liu et al. 2014) Hördt et al. 2020 and Gemmobacter megaterium Liu et al. 2014. *The NamesforLife Abstracts*. https://doi.org/10.1601/ex.24974

[17]. Lu, Z., Dou, Z., Lian, J., Xie, X., & Yang, Q. (2020). Content-Based Collaborative Filtering for News Topic Recommendation. *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 217-223

[18]. *WTF is TF-IDF?* (n.d.). KDnuggets. https://www.kdnuggets.com/2018/08/wtf-tf-idf.html

