DOI:10.9756/INTJECSE/V14I5.888 ISSN: 1308-5581 Vol 14, Issue 05 2022

News Recommendation System Review

Niran A. Abdulhussein, Ahmed J. Obaid

nirana.alasadi@student.uokufa.edu.iq; ahmedj.aljanaby@uokufa.edu.iq

Faculty of Computer Science and Mathematics, University of Kufa, Iraq

Abstract

Nowadays, it is a very important way for researchers and people to find their desired meaning instead of searching for specific topics. Recommendation systems allow users to quickly search for and locate the topics they need in complex information environments. RS can show us the related results are close to what we want. In this paper, we listed the rate of the most used types and techniques in RS and their strategies.

Introduction

Recommendation systems (RS) are subclasses of information filtering system that pursues to predict "rating" or "preferences" a user might give to an item (such as music, books, or movies) or a social element (such as people or a group). Haven't considered yet, using a model built on item properties (content-based approaches) or the user's social environment (collaborative filtering approaches) [1]. Although many different approaches to recommendation systems have been developed in the past years, interest in this field remains high due to the growing demand for practical applications, providing personalized recommendations, and dealing with redundant information. Data has become the main factor in everything, but nowadays the data size is increasing exponentially, including news data also increases over time [2]. According to IBM, "Internet users generate (2.5) trillion bytes of information every day." Much of the information on Earth today (90%) has been generated over the last two years. This information can be obtained from anywhere. Examples: social media posts, images, videos, both e-commerce and non-e-commerce migration records, satellite data, etc. This data is called big data. The Tech America Foundation describes big data as follows: "Big data is a term that defines large amounts of fast and complex variable data that require advanced techniques and technologies that enable the storage, collection, distribution, management, and analysis of information." [2]. RS is widely explored and applied in both E-commerce and non-E-commerce to maximize profit and meet the precision marketing aim. Some researchers looked into how RS helps E-commerce firms make the most money and analyzed the advice from a variety of market-leading websites. Using this technique, Amazon saw a 20 percent to 30 percent increase in sales [2].

RS is a sort of information filtering system that filters and creates information for users based on their choices, interests, or, in the case of e-commerce, past transactions. The information generated can also be based on the user's activity while visiting any website. Based on a user's profile, the algorithm can forecast which things they will like [3].

RS allows users to quickly search for and locate the news they need, as well as propose related or nearby content based on previous clicking patterns, saving time and effort for the user and ensuring they get what they need in a short amount of time.

Types of Recommendation Systems

Based on supplied data, RS attempts to forecast if an item will be valuable to a user [4]. These systems have been slowly increasing in popularity in recent years, with companies such as eBay and Amazon [5] using them in retail and E-commerce. These companies collect a lot of data from their consumers and personalize the RS to their needs [6]. RS are commonly used in E-commerce and retail, as well as in other areas like healthcare, transportation, and agriculture [7]. We've already covered a few key applications of recommendation systems, but we'll put them all in one spot here:

1. **Product Recommendations*:*** Online retailers are perhaps the major users of recommendation systems. We've noticed how Amazon and other online retailers work hard to give each returning user with product recommendations. These recommendations are not created at random, but rather are based on similar consumers' purchase decisions or other methodologies [8].
2. **Movie Recommendations**: Netflix recommends movies to its customers based on their preferences. These suggestions are based on the ratings supplied by users. The necessity of accurately predicting ratings is so great that Netflix offered a million-dollar prize for the first algorithm that could outperform its own recommendation system by a factor of ten (10 percent) [8].
3. **News Articles**: Based on previous articles read, news services have attempted to identify articles of interest to readers. The resemblance could be based on the similarity of keywords in the documents or on articles read by people with similar reading preferences [8].

Furthermore, RS is employed in Business Adoption and Applications [5] such as E-commerce, transportation, agriculture, healthcare, and media.

Techniques of Recommendation Systems

It is critical for a system to use efficient and accurate recommendation strategies in order to deliver effective and relevant recommendations to its users. This explains why it is critical to understand the characteristics and possibilities of various recommendation approaches. To design any system, two things are required: the content of the users and items, as well as their implicit interactions [10]. The techniques are:

1. Content-Based Filtering (CBF)

A CBF algorithm constructs a recommender by comparing the user and item profiles based on the content of a shared attributes space. This technique divides people into groups based on how similar their tastes are [3]. CBF is a domain-dependent algorithm that focuses on analyzing item properties to create predictions. CBF is the most successful when it comes to recommending documents like web pages, publications, and news. CBF makes recommendations based on user profiles and attributes taken from the content of things previously evaluated by the same user [11]. CBF use many sorts of algorithms to discover similarities across documents in order to generate meaningful recommendations [9].

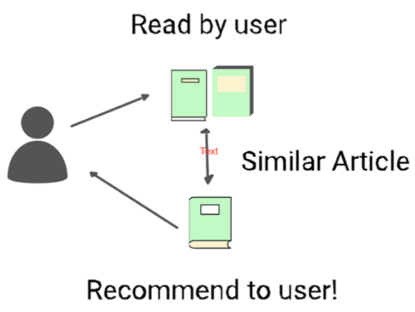

Figure 1: Example of CBF.

As a result, these strategies enable the system to handle a large number of users. CBF is user-independent because it simply requires assessing objects and user profiles to make suggestions. CBF, unlike collaborative filtering, does not have cold-start difficulties. New items or products are suggested before a large number of users rate them. CBF has various disadvantages. For starters, if not enough information is provided in the text to clearly differentiate products, the recommendation will be inaccurate. These strategies necessitate in-depth domain understanding. Second, because they must match the qualities of profiles and commodities, content-based systems are limited in their originality [5].

2. Collaborative Filtering (CF)

A content-free strategy is used when the features of things are not known ahead of time. CF is a domain-independent prediction technique for content such as movies and music that cannot be simply and fully represented by metadata [9]. CF works by collecting a database of user preferences for items. It then makes recommendations by matching people with appropriate interests and preferences based on similarities between their profiles [12]. CF makes use of user behaviors such as ratings, history, and interactions with objects. This strategy presupposes those customers will buy comparable products to the ones they have already purchased [3].

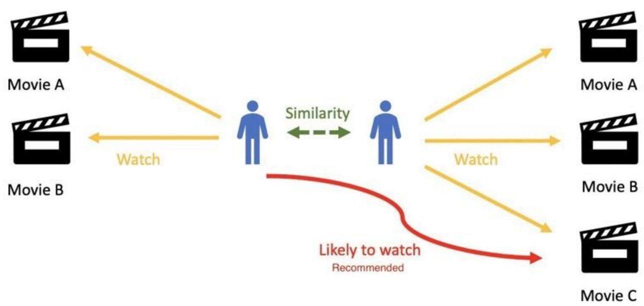

Figure 2: Example of CF.

CF promotes things depending on the user's previous activity. This approach is more accurate than CBF, the second most commonly used method. This is because CBF cannot produce an appropriate suggestion if there is insufficient information provided to differentiate products, implying that CBF requires extensive domain expertise [3].

A CF, the most prevalent recommendation model, suffers from a cold-start problem, in which models encounter difficulties due to data sparsity. The problem is gathering enough data is referred to as data sparsity. Meanwhile, cold-start refers to the models' difficulty in producing correct recommendations. CF is the most commonly used method in many businesses.

The system's most prevalent choice is CF, which does not require domain expertise because the embeddings are automatically taught. The mapping of things to a sequence of numbers is referred to as embedding items in a recommender system. This method of describing items with learnt vectors is used to train algorithms for determining the link between items and extracting their features. Next, the benefit of CF is that it develops models that assist users in discovering new interests. Finally, because the system just requires the rating matrix to construct a factorization model, CF is an excellent starting point for other systems [5]. CF approaches are:

A. Memory-based techniques

Items that the user has previously rated play an important role in his hunt for a neighbor who shares his appreciation [13]. Once a user's neighbor is identified, several techniques can be used to aggregate the preferences of neighbors to provide suggestions. Because of their effectiveness, these strategies have found broad success in real-world applications. Memory-based CF can be performed using:

User-based the collaborative filtering technique determines user similarity by comparing their ratings on the same item, and it then computes the predicted rating for an item by the active user as a weighted average of the ratings of the item by users similar to the active user, where weights are the similarities of these users with the target item.

Item-based the similarity between things, not the similarity between users, is used to derive predictions in filtering approaches. It constructs a model of item similarities by obtaining all items rated by an active user from the user-item matrix, determining how similar the retrieved items are to the target item, and then selecting the most similar items and determining their related similarities [9].

B. Model-based techniques

This technique uses prior ratings to train a model to improve the CF Technique's performance. Machine learning or data mining techniques can be used to develop models. Because they use a pre-computed model, these strategies may swiftly recommend a set of items and have shown recommendation outcomes that are similar to neighborhood-based recommender techniques [9]. To forecast item evaluations for a specific user, user-based approaches go through two major steps. The first stage seeks out users who are similar to the target user. The second stage collects rates from users who are similar to the active user and utilizes them to generate recommendations. There have been numerous collaborative filtering algorithm measures that compute user similarities [5].

Dimensionality Reduction techniques such as the Singular Value Decomposition (SVD), Matrix Completion Technique, Latent Semantic approaches, and Regression and Clustering are examples of these techniques. Model-based strategies examine the user-item matrix for relationships between items, which they then use to compare the list of top-N recommendations. Model-based strategies are used to overcome system sparsity issues [9].

The use of learning algorithms has also transformed the nature of recommendations, from advising users on what to consume to advising them on when to consume a product. As a

DOI:10.9756/INTJECSE/V14I5.888 ISSN: 1308-5581 Vol 14, Issue 05 2022

result, it is critical to investigate various learning techniques employed in model-based recommender systems:

Association Rule: Association rules mining algorithms [14] extract rules that anticipate the occurrence of an item based on the existence of other items in a transaction. For example, given a set of transactions, each of which is a set of things, an association rule applies the form A to B, where A and B are two sets of items [15]. Association rules can provide a relatively compact representation of preference data, which may enhance storage efficiency as well as performance.

Clustering: Clustering techniques have been used in a variety of applications, including pattern recognition, image processing, statistical data analysis, and knowledge discovery [16]. The clustering algorithm attempts to partition a set of data into sub-clusters in order to uncover meaningful groups within them. After forming clusters, the opinions of other users in the cluster can be averaged and used to produce suggestions for specific users. In some clustering systems, a user can participate in multiple clusters only partially, and suggestions are based on the average across all clusters of involvement, weighted by a degree of participation [17].

Decision Tree: The decision tree is built using tree graph methods, which involves examining a set of training samples for which the class labels are known [9]. They are then used to classify previously unseen cases. They can produce very accurate predictions when trained on very high-quality data [18]. Decision trees are more interpretable than other classifiers such as Support Vector Machine (SVM) and Neural Networks because combine simple queries about data in a comprehensible manner.

Artificial Neural Network (ANN): is a structure composed of many linked neurons (nodes) arranged in layers in a systematic manner. Weights are assigned to neural connections based on the degree of impact one neuron has over another. In some specific problem circumstances, neural networks have some advantages [9]. ANN can estimate nonlinear functions and capture complicated relationships in data sets; they are also efficient and can function even if a portion of the network fails. The main disadvantage is that it is difficult to determine the optimal network topology for a given problem, and once determined, the topology serves as a lower bound for the classification error.

3. Hybrid Filtering

A hybrid filtering technique integrates several recommendation algorithms to improve system optimization and avoid some of the limitations and issues associated with pure recommendation systems. Their primary goal is to eliminate the disadvantages of the individual ones. The assumption behind hybrid approaches is that a combination of algorithms will produce more accurate and effective suggestions than a single algorithm since the shortcomings of one algorithm can be compensated for by another [9].

Using various recommendation strategies in a combined model can mitigate the shortcomings of individual techniques. Integrating techniques can be accomplished in one of three ways: separately implementing algorithms and combining the results, using some CBF in a collaborative approach, using some CF in a content-based approach, or developing a unified recommendation system that combines both approaches.

Weighted, Switching, Mixed, Feature Combination, Cascade, Feature Augmentation, and Meta-Level are some hybrid filtering combination strategies [5].

DOI:10.9756/INTJECSE/V14I5.888 ISSN: 1308-5581 Vol 14, Issue 05 2022

References

[1] Lalita Sharma and Anju Gera, “A Survey of Recommendation System: Research Challenges”, International Journal of Engineering Trends and Technology (IJETT) - Volume4Issue5- May 2013. ISSN: 2231-5381, http://www.ijettjournal.org, Page 1990-1991.
[2] Debashis Das, Laxman Sahoo, and Sujoy Datta, “A Survey on Recommendation System”, International Journal of Computer Applications (0975 – 8887) Volume 160 – No 7, February 2017.
[3] Daphne Bunga Dwiputriane, Zuraida Abal Abas, And Nanna Suryana Herman, “Systematic Literature Review on Enhancing Recommendation System By Eliminating Data Sparsity”, Journal of Theoretical and Applied Information Technology, 15th April 2022. ISSN: 1992-8645, www.jatit.org, E-ISSN: 1817-3195, pages 2254- 2270.
[4] Ricci, F.; Rokach, L.; Shapira, B. Introduction to recommender systems handbook. In Recommender Systems Handbook; Springer: Boston, MA, USA, 2011; pp. 1–35.
[5] Zeshan Fayyaz, Mahsa Ebrahimian, Dina Nawara, Ahmed Ibrahim, and Rasha Kashef. “Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities”, Applied science, MDPI. Appl. Sci. 2020, 10, 7748; doi:10.3390/app10217748, www.mdpi.com/journal/applsci.
[6] Schafer, J.B.; Konstan, J.A.; Riedl, J. E-commerce recommendation applications. Data Min. Knowl. Discov. 2001, 5, 115–153.
[7] Mustaqeem, A.; Anwar, S.M.; Majid, M. A modular cluster-based collaborative recommender system for cardiac patients. Artif. Intell. Med. 2019, 102.
[8] Jure Leskovec, Anand Rajaraman, and Jeffrey David Ullman, “Ch9: Recommendation Systems”, Mining of Massive Datasets, Published online by Cambridge University Press: 05 December 2014, pp. 292 – 324, DOI: https://doi.org/10.1017/CBO9781139924801.010.
[9] Shaina Raza and Chen Ding, “News recommender system: a review of recent progress, challenges, and opportunities”, Published online: 21 July 2021, Springer, Artificial Intelligence Review, https://doi.org/10.1007/s10462-021-10043-x, 2021.
[10] Mobasher B, Jin X, Zhou Y. Semantically enhanced collaborative filtering on the web. In: Web mining: from web to semantic web. Berlin Heidelberg: Springer; 2004. p. 57–76.
[11] F.O. Isinkaye, Y.O. Folajimi, and B.A. Ojokoh, “Recommendation systems: Principles, methods and evaluation”, 2015.
[12] Pan C, Li W. Research paper recommendation with topic analysis. In Computer Design and Applications IEEE 2010;4, pp. V4-264.
[13] Zhao ZD, Shang MS. User-based collaborative filtering recommendation algorithms on Hadoop. In: Proceedings of 3rd international conference on knowledge discovering and data mining, (WKDD 2010), IEEE Computer Society, Washington DC, USA; 2010. p. 478–81. doi: 10.1109/WKDD.2010.54.
[14] Yoon HC, Jae KK, Soung HK. A personalized recommender system based on web usage mining and decision tree induction. Expert Syst Appl 2002;23:329–42.
[15] Ku_zelewska U. Advantages of information granulation in clustering al-gorithms. In: Agents and artificial intelligence. NY: Springer; 2013. p. 131–45.
[16] Linden G, Smith B, York J. Amazon.com recommendation: item-to-item collaborative filtering. IEEE Internet Comput 2003;7(1): 76–80.
[17] Caruana R, Niculescu-Mizil A. An empirical comparison of supervised learning algorithms. In: Cohen W. Moore AW, editors. Machine Learning, Proceedings of the twenty-third international conference, ACM, New York; 2003. p. 161–8.