

A review of techniques used in e-commerce recommendation system

Ke Yan

University of Liverpool, Liverpool, UK

yank.fitzgerald@gmail.com

Abstract. In the age of the internet, the explosive growth of information makes the e-commerce platform users need more and more requirements for the shopping experience. Recommendation systems are developed to retrieve information efficiently and provide a personalised shopping experience for users. There are three types of algorithms that are most frequently employed in recommendation systems: content-based filtering, collaborative filtering, and hybrid techniques. In this paper, content-based filtering and collaborative filtering are firstly explained in terms of working principles, limitations, and advantages. Then, some hybrid techniques that combine content-based filtering and collaborative filtering are proposed to overcome the problems of these two conventional methods. It is found that while collaborative filtering may struggle with scalability, sparsity, and cold start issues, content-based filtering may have quality issues and a lack of variation. The hybrid techniques are introduced to address the drawbacks of these two algorithms. It is, however, more complex and requires more memory resources to use hybrid techniques. Implementing deep learning in recommendation systems would be a promising area in the future. Through the analysis above, this paper would provide a systematic review of the development of the recommendation system and inform the future study. Students or researchers who are interested in big data can quickly grasp the concepts of techniques used in e-commerce recommendation systems as well as the benefits and drawbacks of different techniques.

Keywords: Content-Based Filtering, Collaborative Filtering, E-Commerce Recommendation System, Hybrid Techniques.

1. Introduction

The boom of the internet has spawned a large number of new industries. E-commerce, one of the most representative new industries, is playing a more and more important role in people's daily lives. Meanwhile, the burgeoning of the internet leads to the rapid growth of information, which means it is more challenging to retrieve useful information. This phenomenon makes the online shopping experience less enjoyable for most users. Therefore, to improve customer satisfaction and brand loyalty, e-commerce companies like Amazon and Netflix are devoted to constructing a recommendation system that can provide users with personalised recommendations.

The recommendation system is based on data-mining technology. It can analyse users' preferences and anticipate what the user would purchase. Content-based Filtering and Collaborative Filtering are two typical techniques that are used to develop recommendation systems. There are also Hybrid Techniques which combine features of Content-based Filtering and Collaborative Filtering have been introduced to enhance the quality of recommendations [1]. This paper will first briefly introduce the

structure of the recommendation systems, then give a critical review of Content-based Filtering and Collaborative Filtering by describing their working principles and analysing their limitations and advantages. To overcome the limitations of the two conventional approaches, some hybrid techniques that merge content-based filtering with collaborative filtering will be presented. Finally, we will have a glimpse into the future development.

2. Recommendation System

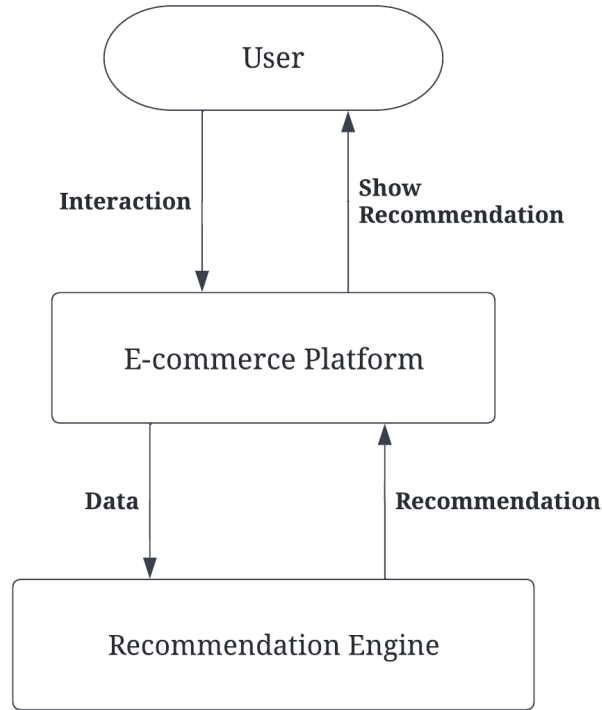


Figure 1. Recommendation System.

Figure 1 depicts a simple recommendation system. Users' interaction with the e-commerce platform will generate a series of data such as browsing history, purchase records and ratings. This data will be collected and pre-processed by the e-commerce platform and then fed to the recommendation engine. The engine will then generate a collection of recommended items, and the e-commerce platform will display the items (usually in a top-N list form) to the user. The recommendation engine is the crucial part of the whole system. There are three main techniques for developing a recommendation engine: Content-based Filtering, Collaborative Filtering and Hybrid Techniques.

2.1. Content-based Filtering

2.1.1. Working Principle. Content-based Filtering assumes that if a user is interested in one item, then they will also be interested in other items with similar characteristics [2]. Content-based Filtering suggests items to those the user has already bought or liked. Most content-based recommendation systems build user profiles and item profiles [2, 3, 4]. A user profile consists of preferences and requirements, while an item profile includes a list of attributes about the item [3].

There is not always an efficient way to extract item attributes. For a book, the keywords, the category, the author and the language can be easily retrieved by algorithms. For a refrigerator, however, its attributes can be vague and sometimes have to be selected manually. Therefore, the majority of content-based filtering research tends to concentrate on items with textual content [5]. For example, books, web pages, and movies. Term frequency-inverse document frequency (tf-idf) is one of the most successful information retrieval methods, which assigns a weight for each term that appears in the text [4]. For example, for item A and item B, two vectors of weights, A and B can be

calculated through tf-idf. Then the degree of similarity between two items is measured by Cosine Similarity $\text{sim}(A, B)$.

$$\text{sim}(A, B) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (1)$$

After calculating all the similarities between each item, the recommendation system will recommend the top-n items to the user.

2.1.2. Limitations. As mentioned above, it is challenging to collect attributes for some items.

The recommended items may have quality problems. Conventional content-based filtering only measures the similarity between items while not considering other users' ratings or comments about the item. As a result, the recommended items' quality varies greatly.

The category of recommended items can be monotonous, as content-based filtering can only generate recommendations based on the user's current preferences. In other words, the model has limited ability to extend users' existing interests and cannot predict their future interests [4]. Therefore, this content-based strategy might not be appropriate for comprehensive e-commerce platforms with multiple categories.

2.1.3. Advantages. Information regarding other users is not needed because the recommendations only serve the current user. The growing user base does not significantly affect the performance of the recommendation system. Content-based filtering is straightforward and easy to use, which makes it easier for e-commerce companies to employ it for their platform.

2.2. Collaborative Filtering

2.2.1. Working Principle. Collaborative filtering is first introduced in the first commercial recommendation system Tapestry [6], which can find emails that a current user may be interested in based on how other users rate the emails. Collaborative filtering follows such an assumption, that is, similar users will share similar interests. Usually, similar users can be found using the User-Item Matrix shown in Figure 2.

	Item						
	1	2	3	...	j	...	k
User	1	1			3		
	2		4				1
	3	5		3		3	2
	:						
	i		3	5		4	
	:						
	a	5		4	?		2

Figure 2. User-Item Matrix.

Each element of the user-item matrix corresponds to a rating r . For instance, r_{ij} indicates user i rated item j as 4. Collaborative filtering is such an algorithm to predict the rating of item j by active user a , that is, the value of $r_{a,j}$. Since it is clear that user 3 and user a share similar interests, we call user 3 a

neighbour of user a. Usually, the similarity between user A and user B is measured by Pearson Correlation:

$$r_{A,B} = \frac{\sum_{i=1}^k (A_i - \bar{A})(B_i - \bar{B})}{\sqrt{\sum_{i=1}^k (A_i - \bar{A})^2} \sqrt{\sum_{i=1}^k (B_i - \bar{B})^2}} \quad (2)$$

Where:

- A and B are row vectors in the user-item matrix.
- k is the dimension of vector.

In practice, there are a couple of methods to measure how similar the users are in collaborative filtering, such as Cosine Similarity and Pearson Correlation. Nevertheless, Breese et al. [7] pointed out that Pearson Correlation performs better than Cosine Similarity (Vector Similarity). Having calculated the similarity, collaborative filtering will find several neighbours similar to user a, and then get the estimated value of $r_{a,j}$ by weighted average.

The conventional collaborative filtering mentioned above, also known as user-based collaborative filtering, can have a problem with scalability [8]. The quick expansion of users and items makes the complexity of computing users' similarity increase dramatically. Therefore, researchers proposed an item-based collaborative filtering [9, 10], which compares new items to those that active users have rated highly, rather than finding similar users. This approach has shown significant improvement in the computation speed and recommendation quality [9, 10].

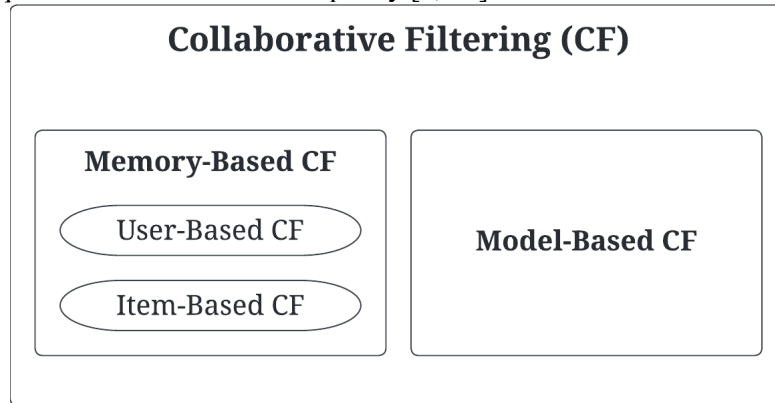


Figure 3. Collaborative Filtering Hierarchy

Generally, there are two categories of collaborative filtering: one is memory-based collaborative filtering, which includes user-based collaborative filtering and item-based collaborative filtering. Another is model-based collaborative filtering. Memory-based collaborative filtering uses whole or partial data for each computation, while model-based collaborative filtering generates recommendations based on pre-trained models such as Cluster Models and Bayesian Network Models. The experiments conducted by Breese et al. [7] show that the Bayesian Network Model-Based Collaborative Filtering is faster and consumes less memory than the Memory-Based Collaborative Filtering.

2.2.2. Limitations. In addition to the scalability problem mentioned above, collaborative filtering has these common problems:

1. Sparsity

The user-item matrix is quite sparse, which means most of the elements are empty [5, 8]. This is because users do not always rate what they have bought. The data sparsity makes the overlap of rated items between two users minor, which lowers the probability of finding users who have similar ratings.

2. Cold Start

Cold start, also known as first-rater problem, is a common problem among new items and users who have not used this e-commerce platform before [5]. On the one hand, a new item will not be recommended unless it has been rated. On the other hand, no similar user will be found until the new user has rated something.

2.2.3. Advantages. Compared with content-based filtering, collaborative filtering can be used for items with implicit attributes. In addition, collaborative filtering can recommend unexpected items for users, which helps to expand their interests. Because of this, collaborative filtering recommendation systems are more popular with e-commerce giants.

2.3. Hybrid Techniques

Hybrid techniques attempt to combine content-based filtering and collaborative filtering. The simplest hybrid technique is to execute content-based filtering and collaborative filtering simultaneously, then merge the results together. Cotter and Smyth [11] have developed such a hybrid recommendation approach for a digital television system. This hybrid method runs content-based filtering and collaborative filtering in parallel, which utilises the advantages of each algorithm. Therefore, the user can receive TV guides based on previously liked channels and other users with similar interests. Also, Claypool et al. [12] proposed a weighted average approach to merge content-based filtering with collaborative filtering. The weight is adjusted dynamically. As the item's number of users and ratings rise, the collaborative filter will carry higher weight. This approach alleviates the cold start problem in the early stage and increases the quality and accuracy of recommendations.

Melville et al. [13] introduced another hybrid method named Content-Boosted Collaborative Filtering. The lack of user ratings is the cause of the sparsity problem, but content-based filtering can predict users' interests. Thus, they use content-based filtering to fill the sparse user-item matrix with predicted rating values. Then, recommendations are generated through collaborative filtering. Specifically, they implement a naive Bayesian text classifier in content-based filtering, which can predict the rating of unrated items. This method solves the problem of matrix sparsity in conventional collaborative filtering, and the cold start problem is also alleviated to some degree. The content-boosted collaborative filtering has been proven to perform better than the simple merging of content-based filtering and collaborative filtering [13].

Content-based filtering has no cold start problem as it only cares about the attributes of items, but its recommendation quality cannot be guaranteed, and it suffers from monotonous recommendation results. On the contrary, collaborative filtering can provide various high-quality items, but it has cold start and sparsity issues. To some extent, content-based filtering and collaborative filtering are complementary. It is intuitive for hybrid techniques to arrogate them to generate better recommendation results.

3. Future Work

While hybrid techniques solve most of the problems with content-based filtering and collaborative filtering, it inevitably increases the computational complexity and consumes more memory resources. Therefore, it is still necessary to dig deeper to find more efficient and accurate methods. Deep learning, a powerful tool for solving complex problems, has been widely used in data mining, image processing and natural language processing. Recently, it has been used in recommendation systems to improve the capability to process big data [14]. Devooght and Bersini [15] utilise recurrent neural networks (RNN) to transform collaborative filtering into a sequence prediction problem. The results show that RNN produces more diverse recommendations and is well suited to dense data sets. Therefore, it can be predicted that the introduction of deep learning can provide more personalised and higher-quality recommendations in the future.

4. Conclusion

The recommendation system can improve customer satisfaction and bring profit growth for e-commerce industries. There are three techniques commonly used for recommendation engines: Content-based Filtering, Collaborative Filtering, and Hybrid Techniques. This paper has described the

working principles for each technique. It is noteworthy that content-based filtering can have quality problems and lack variety, while collaborative filtering might face scalability, sparsity, and cold start problems. To overcome the drawbacks of these two approaches, hybrid techniques that merge content-based filtering with collaborative filtering have been introduced. Both the simple hybrid approach and the content-boosted approach perform better than pure content-based or collaborative filtering. However, the hybrid techniques require more memory resources and increase complexity. Deep learning shows great potential for improving computational capabilities. It is predictable that the introduction of deep learning in recommendation algorithms will be an important direction for future studies. There are also limitations to this paper. For instance, many variants of content-based filtering and collaborative filtering have been presented in recent years, all of which are remarkable and have improved the original algorithm to some degree. To keep it straightforward and explore the most extensively used techniques in e-commerce recommendation systems, this paper cannot provide a detailed description of all the variants. In the future, researchers who are interested in this topic could choose one algorithm and dig deeper to analyse the limitations and advantages of different variants.

References

- [1] T. Badriyah, E. T. Wijayanto, I. Syarif, and P. Kristalina, "A hybrid recommendation system for E-commerce based on product description and user profile," in 2017 Seventh International Conference on Innovative Computing Technology (INTECH), 2017: IEEE, pp. 95-100.
- [2] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen, "Collaborative Filtering Recommender Systems," in *The Adaptive Web: Methods and Strategies of Web Personalization*, P. Brusilovsky, A. Kobsa, and W. Nejdl Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, p. 291-324.
- [3] H. Li, F. Cai, and Z. Liao, "Content-based filtering recommendation algorithm using HMM," in 2012 Fourth International Conference on Computational and Information Sciences, 2012: IEEE, pp. 275-277.
- [4] R. Van Meteren and M. Van Someren, "Using content-based filtering for recommendation," in *Proceedings of the machine learning in the new information age: MLnet/ECML2000 workshop*, 2000, vol. 30, pp. 47-56.
- [5] P. Melville and V. Sindhwani, "Recommender systems," *Encyclopedia of machine learning*, vol. 1, pp. 829-838, 2010.
- [6] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry," *Communications of the ACM*, vol. 35, no. 12, pp. 61-70, 1992.
- [7] J. B. D. CarlKadie, "Empirical analysis of predictive algorithms for collaborative filtering," *Microsoft Research Microsoft Corporation One Microsoft Way Redmond, WA*, vol. 98052, 1998.
- [8] D. Liu, "A Study on Collaborative Filtering Recommendation Algorithms," in 2018 IEEE 4th International Conference on Computer and Communications (ICCC), 2018: IEEE, pp. 2256-2261.
- [9] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the 10th international conference on World Wide Web*, 2001, pp. 285-295.
- [10] G. Linden, B. Smith, and J. York, "Amazon. com recommendations: Item-to-item collaborative filtering," *IEEE Internet computing*, vol. 7, no. 1, pp. 76-80, 2003.
- [11] P. Cotter and B. Smyth, "Ptv: Intelligent personalised tv guides," in *AAAI/IAAI*, 2000, pp. 957-964.
- [12] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes, and M. M. Sartin, "Combining Content-Based and Collaborative Filters in an Online Newspaper," in *SIGIR 1999*, 1999.
- [13] P. Melville, R. J. Mooney, and R. Nagarajan, "Content-boosted collaborative filtering for improved recommendations," *Aaai/iaai*, vol. 23, pp. 187-192, 2002.

- [14] Z. Batmaz, A. Yurekli, A. Bilge, and C. Kaleli, “A review on deep learning for recommender systems: challenges and remedies,” *Artificial Intelligence Review*, vol. 52, no. 1, pp. 1-37, 2019.
- [15] R. Devooght and H. Bersini, “Long and short-term recommendations with recurrent neural networks,” in *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, 2017, pp. 13-21.