# Comparison of the Effectiveness of Various Algorithms on a Recommendation System

Bektemyssova Gulnara and Akhmer Yerassyl

International Information Technology University, Almaty, Kazakhstan

**Abstract.** Recommender systems attempt to identify user information by proposing related products or resources that customers may be interested. Recommender methods have attracted attention in the fields of information technology, ecommerce, and so on, by essentially fertilizing from a standard collection of decisions that led consumers to find information of interest. This research focuses on the three common recommendation systems: Collaborative Filtering, Content-Based Filtering, and Hybrid recommendation systems. For the purposes of this analysis, the well-known MovieLens dataset has been used. The assessment considered both the quantitative and qualitative dimensions of the recommendation systems. This paper describes the field of various recommendation approaches and related fundamental techniques. Any algorithm in this field has both benefits and drawbacks. The goal of the research is to bring various algorithms to the test in order to find the right one based on the layout of the dataset and the researchers' goals.

**Keywords:** Recommender Systems, Collaborative Filtering, Content-Based Filtering, E-Commerce, Hybrid Recommendation System.

## 1 Introduction

Recommender systems are an integral part of e-commerce today. The active transition from traditional offline sales to online makes the introduction of machine learning technologies and algorithms for recommendations more and more popular in retail. [1]. Recommendations simplify shopping for store customers, and allow sellers to increase customer loyalty by saving time and an individual approach to product offerings, as well as increasing the product matrix and average customer check. Unlike e-commerce, grocery chains do not represent how customers react to promoted products in real time. However, thanks to loyalty programs and check databases, it is possible to build a recommendation system from scratch. [2].

In this paper, we will look at various concepts of recommender systems. We will introduce how they perform, define their theoretical background, and start debating their strengths and limitations for each of them. A comparative analysis

of these algorithms is carried out from the point of view of the criteria of the accuracy of the results obtained and the performance.

In the first part, we will address the two main methodologies of recommender systems: collaborative and content-based approaches. The following two parts would then go through different collaborative filtering methods, such as user-user, item-item, and matrix factorization. The part that follows presents contents-based approaches and their operation. Finally, we will go over how to assess a recommender system.

In retail, three types of recommendations are commonly used: content, collaborative, and hybrid. Recommendation systems are frequently divided into three large categories:

- Content-Based systems, which are using keywords to propose products to a client that are close to those historically favored [7];
- Collaborative Filtering methods, that propose products based on information recently seen or purchased.
- Hybrid Recommendation methods, which provide a variation of Content-Based and Collaborative Filtering techniques to overcome some of the shortcomings that occur in the above-mentioned systems.

## 2 Approaches

### 2.1 Collaborative filtering

Collaborative recommendation is quite certainly the commonly used, and advanced of the approaches. Collaborative recommender frameworks combine item ratings or suggestions, identify common threads among customers based on their scores, and produce new suggestions based on inter-user correlations. This approach may be Memory-Based Collaborative Filtering, which measures customer's access using similarity or other metrics, or Model-Based Collaborative Filtering, that derives a template from past prescriptive analytics and uses it to make forecasts. [3, 4].

### 2.2 Content-based filtering

Even though Collaborative Filtering is well known and effective, it has drawbacks. One of them is the sparsity dilemma, which happens when users give no scores; throughout this situation, our model is unable to produce fair suggestions. To address the sparsity problem, study suggests Content-based Recommender Systems, which are focused on the analysis of adjunct data such as text, photographs, and videos, as well as customers' accounts. [5]. Assume anyone loves science fiction, romance, and action films but not fantasy films. Through period, the algorithm could collect this knowledge and decide that the

client has a high approval rating for genres such as science fiction, romance, and action, and a negative rating for fantasy. The algorithm could even discover which actors the client likes and dislikes. Also with tiny remarks, the customer's choice may be inferred in this manner. The critical point between Content Based Filtering and Collaborative Filtering is that Collaborative Filtering proposes new products depending on the taste of the customers who have common preferences for many other products, while Content Based Filtering is focused on the analysis of source data and is not associated with the expectations of many other clients.

## 2.3 Hybrid recommender system

The term "hybrid recommendation strategy" applies to a recommendation system that employs two or more sources of recommendation methods in order to achieve better results while minimizing the disadvantages of each particular one. Collaborative filtering is often paired with another method.

## 3  Related works

When working with items containing textual data, content-based systems yield outcomes that are more accurate. However, these systems are incapable of distinguishing between a well-written text definition and a poorly written one, particularly when similar or different phrases are used [8]. Furthermore, these systems are sometimes constrained by the over similarity issue; when a system suggests products that have a higher correlation to a customer's profile, the client is likely to be recommended with products that are identical to those which have already been seen [10]. Besides that, when a new customer enters in the system with little or no rankings, he or she is very likely to be given low accuracy suggestions (this is recognized as the cold-start or new-user problem) [8]. As mentioned in [10]. Content-based systems need a great amount of scores before recommending products to a consumer with high precision. Collaborative Filtering methods, in comparison to content-based systems, result in bias due to the sparsity problem [8]. Since the amount of items on e-commerce websites is immense, the most frequent users normally rank only a portion of the given data. It implies that some of the most common products have very few scores and therefore have a low probability of being suggested by the system [8, 9]. Collaborative Filtering systems, like Content-based systems, should have a large number of relevant data on a user account before producing correct predictions. Furthermore, new products must be assessed by a wide range of users; otherwise, the RS would be unable to offer suggestions for items [11]. In specific, RS face technical challenges; given the massive quantities of data available on websites and apps, a significant amount of computing effort has been put to generate suggestions [9].

# 4 Preliminary experiments

For preliminary study, we used the 'MovieLens 1M Dataset.' The dataset includes 1,000,209 anonymous reviews of roughly 3,900 movies submitted by 6,040 MovieLens subscribers who entered the site in 2000. We explicitly selected two documents: ratings and movies. There were four fields in the ratings file. They are as follows: UserID (scale from 1 to 6040), MovieID (varies from zero to 3952), Ratings (a 5-star ranking), and Timestamp (in seconds after the epoch). Each consumer does have at least 20 ratings. There were three basic forms in the movies log. They are as follows: MovieID, Title, and Genres. Titles are much the same as given by IMDB (including year of release). Genres are tube and chosen from the categories listed: Children's, Comedy, Crime, Documentary, Drama, Fantasy, Film-Noir, Horror, Musical, Mystery, Romance, Sci-Fi, Thriller, War and Western.

We conducted preliminary research study on the datasets. Figure 1 depicts the histogram of average ratings posted by customers. As we can see, this plot resembles a normal distribution with a strong left tail. The majority of users have average scores between 3.5 and 4.
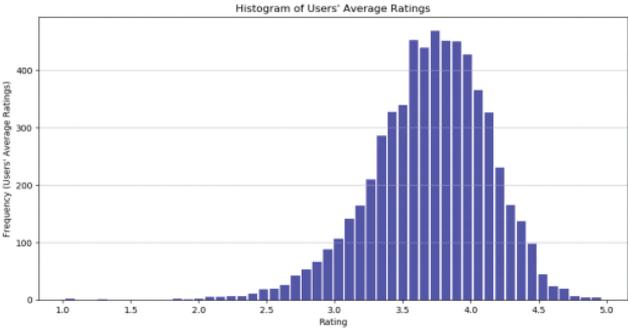


**Fig. 1.** Histogram of users' average scoring.

Fig. 2 depicts a histogram of user-rated products. According to these two graphs, most consumers score just a few objects.
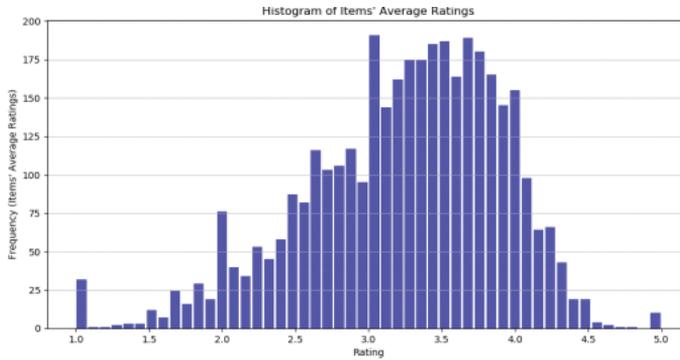
**Fig. 2.** Histogram of items rated by users.

## 5    Results and discussion

Quantitative analysis starts by examining the RMSE and MAE errors of a Collaborative Filtering-based and a Hybrid system. Since the Content-Based Filtering approach has quite a statistical attribute. In this section, we select the top-recommended movies from both methods for ten clients and compute the RMSE errors for each method for analysis. The RMSE graph for ten clients in Fig. 3 shows that perhaps the hybrid model has a relatively lower RMSE. Fig. 4's typical RMSE plot also illustrates the hybrid system's supremacy.
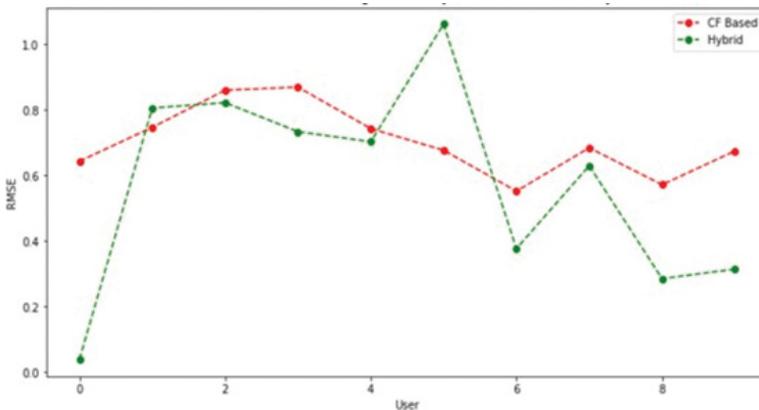


**Fig. 3.** RMSE of collaborative filtering and hybrid recommendation system.
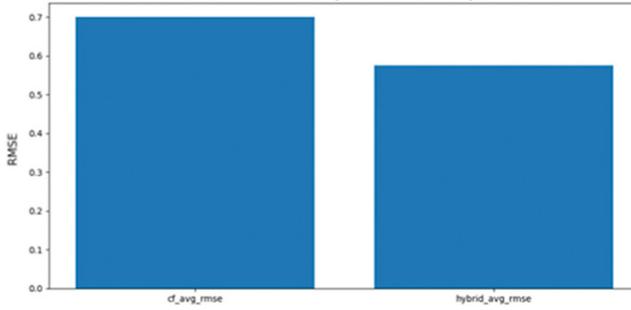
**Fig. 4.** Average RMSE of collaborative filtering and hybrid recommendation system.

Next, we consider 5 batches of users with each batch containing 5 users for whom we do the same test. We calculated the MAE of these sets of users that is shown in Fig. 5 and the comparison shows Hybrid system performs comparatively better. Fig. 6 shows the average MAE of Collaborative Filtering and Hybrid Recommendation System.
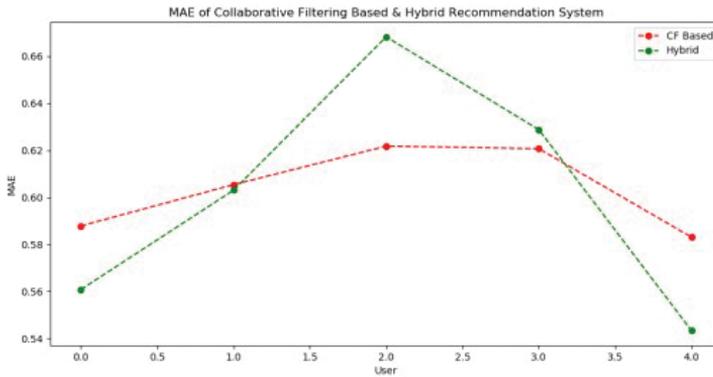


**Fig. 5.** MAE of collaborative filtering and hybrid recommendation system for 5 sets of users.
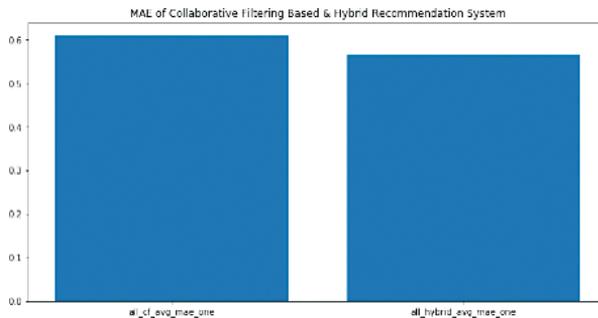


**Fig. 6.** Average MAE of collaborative filtering based and hybrid recommendation system for 5 sets of users.

Fig. 7 shows that Collaborative Filtering will predict which films a client is likely to score higher. However and therefore has no possibility of suggesting related movies to a specific one suited to the consumer? The genres are all around the place, as shown by the genre section. In this segment, we assume User 1 and propose the top 20 movies that he is likely to appreciate high.

| movie_id | estimated_rating | title | actual_rating | genres |
|---|---|---|---|---|
| 527 | 4.995416 | [Schindler's List (1993)] | [5] | [Drama\|War] |
| 318 | 4.958150 | [Shawshank Redemption, The (1994)] | [] | [Drama] |
| 1172 | 4.894460 | [Cinema Paradiso (1988)] | [] | [Comedy\|Drama\|Romance] |
| 2905 | 4.797475 | [Sanjuro (1962)] | [] | [Action\|Adventure] |
| 1269 | 4.727696 | [Arsenic and Old Lace (1944)] | [] | [Comedy\|Mystery\|Thriller] |
| 920 | 4.695179 | [Gone with the Wind (1939)] | [] | [Drama\|Romance\|War] |
| 2019 | 4.693533 | [Seven Samurai (The Magnificent Seven) (Shichi... | [] | [Action\|Drama] |
| 904 | 4.686336 | [Rear Window (1954)] | [] | [Mystery\|Thriller] |
| 922 | 4.654332 | [Sunset Blvd. (a.k.a. Sunset Boulevard) (1950)] | [] | [Film-Noir] |
| 1234 | 4.651489 | [Sting, The (1973)] | [] | [Comedy\|Crime] |
| 1203 | 4.645870 | [12 Angry Men (1957)] | [] | [Drama] |
| 905 | 4.644219 | [It Happened One Night (1934)] | [] | [Comedy] |
| 858 | 4.639405 | [Godfather, The (1972)] | [] | [Action\|Crime\|Drama] |
| 1197 | 4.635691 | [Princess Bride, The (1987)] | [3] | [Action\|Adventure\|Comedy\|Romance] |
| 1233 | 4.632157 | [Boat, The (Das Boot) (1981)] | [] | [Action\|Drama\|War] |
| 356 | 4.629706 | [Forrest Gump (1994)] | [] | [Comedy\|Romance\|War] |
| 1198 | 4.626178 | [Raiders of the Lost Ark (1981)] | [] | [Action\|Adventure] |
| 953 | 4.613348 | [It's a Wonderful Life (1946)] | [] | [Drama] |
| 912 | 4.610552 | [Casablanca (1942)] | [] | [Drama\|Romance\|War] |
| 1242 | 4.608209 | [Glory (1989)] | [] | [Action\|Drama\|War] |

**Fig. 7.** Collaborative filtering based recommendation system's top 20 suggested films for a particular user.

A Content-Based Filtering recommendation framework, from the other side, seems to have the opportunity to relate us so much similar movies to a specified one, as seen in Figure 8, it has very little insight into whether a client will like that or not. In this part, we select Movie Name: Toy Story 39 (1995) with Movie ID 1 and propose the top 20 films that are close to the film, Toy Story.

| movie_index | similarity_score | title | movie_id | genres |
|---|---|---|---|---|
| 1050 | 1.000000 | Aladdin and the King of Thieves (1996) | 1064 | [['Animation', "Children's", 'Comedy']] |
| 2072 | 1.000000 | American Tail, An (1986) | 2141 | [['Animation', "Children's", 'Comedy']] |
| 2073 | 1.000000 | American Tail: Fievel Goes West, An (1991) | 2142 | [['Animation', "Children's", 'Comedy']] |
| 2285 | 1.000000 | Rugrats Movie, The (1998) | 2354 | [['Animation', "Children's", 'Comedy']] |
| 2286 | 1.000000 | Bug's Life, A (1998) | 2355 | [['Animation', "Children's", 'Comedy']] |
| 3045 | 1.000000 | Toy Story 2 (1999) | 3114 | [['Animation', "Children's", 'Comedy']] |
| 3542 | 1.000000 | Saludos Amigos (1943) | 3611 | [['Animation', "Children's", 'Comedy']] |
| 3682 | 1.000000 | Chicken Run (2000) | 3751 | [['Animation', "Children's", 'Comedy']] |
| 3685 | 1.000000 | Adventures of Rocky and Bullwinkle, The (2000) | 3754 | [['Animation', "Children's", 'Comedy']] |
| 236 | 0.869805 | Goofy Movie, A (1995) | 239 | [['Animation', "Children's", 'Comedy', 'Romanc... |
| 12 | 0.826811 | Balto (1995) | 13 | [['Animation', "Children's"]] |
| 241 | 0.826811 | Gumby: The Movie (1995) | 244 | [['Animation', "Children's"]] |
| 310 | 0.826811 | Swan Princess, The (1994) | 313 | [['Animation', "Children's"]] |
| 592 | 0.826811 | Pinocchio (1940) | 596 | [['Animation', "Children's"]] |
| 612 | 0.826811 | Aristocats, The (1970) | 616 | [['Animation', "Children's"]] |
| 700 | 0.826811 | Oliver & Company (1988) | 709 | [['Animation', "Children's"]] |
| 876 | 0.826811 | Land Before Time III: The Time of the Great Gi... | 888 | [['Animation', "Children's"]] |
| 1010 | 0.826811 | Winnie the Pooh and the Blustery Day (1968) | 1023 | [['Animation', "Children's"]] |
| 1012 | 0.826811 | Sword in the Stone, The (1963) | 1025 | [['Animation', "Children's"]] |
| 1020 | 0.826811 | Fox and the Hound, The (1981) | 1033 | [['Animation', "Children's"]] |

**Fig. 8.** Top 20 content-based filtering recommendation system recommendations for a specific film.

We get the best possible outcome in a hybrid system. In this section, we identify User ID 1, Movie Toy Story (1995) with Movie ID 1, and suggest the top 20 films that are close to Toy Story and are probably to still be ranked highly by the User 1. As a result, we may infer that perhaps a hybrid recommendation system outperforms a separate Collaborative Filtering or Content-Based Filtering recommendation system from both qualitative and quantitative terms.

## 6 Conclusion

Within the same dataset, three techniques were applied in the analysis to build a recommendation method. By using possibly the best MovieLens dataset, we examined various recommendation mechanisms such as Collaborative Filtering, Content – Based Filtering and Hybrid recommendation systems. We contrasted all three-suggestion mechanisms using a descriptive and analytical assessment of the dataset. The need for a combined quantitative and qualitative analysis reflects the fact that Content-Based Filtering processes cannot be easily evaluated. Furthermore, for any recommender system, the qualitative analysis is vital. In addition, that is why, in addition to the conventional methodology, we developed our unique assessment process. We discovered that a hybrid recommendation system outperforms a traditional recommendation system in all scenarios. Following the example of the whole study, there have been possibilities for additional research. In the suggestion method, for instance, we did not take into

account any demographic details about the client. Even so, considering this will bring more dimension of complexity to the hybrid recommendation framework. Furthermore, we just addressed genre in our Content-Based Filtering suggestion, but one should check at production team as well as movie ratings for any further similarities. A correlation of various Collaborative Filtering-based approaches and consistency tests can also be of concern.

## References

1. Kalitin D.V. Artificial neural networks [Electronic resource]: tutorial / Kalitin DV – Electron. Text data. — Moscow: Misis Publishing House, 2018. — 88 p
2. Francesco R., Lior R. and Bracha Sh. Introduction to Recommender Systems Handbook. Springer, 2011, pp. 1-35
3. Markovsky I. Low-Rank Approximation: Algorithms, Implementation, Applications, Springer, 2012, ISBN 978-1-4471-2226-5.
4. Takacs G., Pilaszy I., Nemeth B., Tikk D. (March 2009). Scalable Collaborative Filtering Approaches for Large Recommender Systems (PDF). Journal of Machine Learning Research 10: 623–656.
5. Brusilovsky P. (2007). The Adaptive Web. p. 325. ISBN 978-3-540-72078-2.
6. MovieLens dataset, https://grouplens.org/datasets/movielens
7. Konstan J.A. and Riedl J. (2012). Recommender systems: from algorithms to user experience. User Model. User-Adapt. Interact., 22(1-2):101–123.
8. Adomavicius G. and Tuzhilin A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Trans. Knowl. Data Eng., 17(6):734–749.
9. Pu P., Chen L., and Hu R. (2012). Evaluating recommender systems from the user's perspective: survey of the state of the art. User Model. User-Adapt. Inter act., 22(4- 5):317–355.
10. Lu L., Medo M., Yeung C. H., Zhang Y.-C., Zhang Z.-K., and Zhou T. (2012). Recommender systems. Physics Reports, 519(1):1–49.
11. Ning X., Desrosiers C. and Karypis, G. (2015). A comprehensive survey of neighborhood-based recommendation methods. In Recommender Systems Handbook, pages 37–76.