Building A Hybrid Recommendation System For E-Commerce
E-Ticaret için Hibrit Öneri Sistemleri

Hakkı Şimşek[1,*], Mustafa Yeniad[1]
[1]Department of Computer Engineering, Ankara Yildirim Beyazit University, Ankara, Turkey
ORCIDs: 0000-0002-9146-2512, 0000-0002-9422-4974
E-mails: hakkisimsek92@gmail.com, myeniad@gmail.com
*Corresponding author.

***Abstract*—With the technology occupying a large place in human life, our shopping habits have also changed drastically. Increasing product variety and alternatives have made it difficult for people to reach products that suit their own tastes. Thanks to the Suggestion Systems, it has made internet shopping easier by using the evaluation, comment and scoring criteria of other users who have used the same or similar products to the products they are interested in. By learning the tendencies of the users while choosing the products, the presentation of the appropriate products enabled the users to reach their requests easily and quickly. At the same time, it has become mandatory for e-commercial companies to learn the preferences of users in order to compete with other companies. In this study, the most popular recommendation systems and algorithms used in e-commerce platforms are compared.**

Keywords—recommendation systems; hybrid recommendation systems for e-commerce; e-commerce shoppings

***Özetçe*—Teknolojinin insan hayatında büyük bir yer tutmasıyla birlikte alışveriş alışkanlıklarımız da büyük ölçüde değişti. Artan ürün çeşitliliği ve alternatifleri insanların kendi zevklerine uygun ürünlere ulaşmasını zorlaştırmıştır. Öneri Sistemleri sayesinde, ilgilendikleri ürünlerle aynı veya benzer ürünleri kullanmış diğer kullanıcıların değerlendirme, yorum ve puanlama kriterlerini kullanarak internet alışverişini kolaylaştırmıştır. ürünler, uygun ürünlerin sunumu kullanıcıların isteklerine kolay ve hızlı bir şekilde ulaşmasını sağlamıştır. Aynı zamanda e-ticaret firmalarının diğer firmalarla rekabet edebilmesi için kullanıcıların tercihlerini öğrenmesi zorunlu hale gelmiştir. Bu çalışmada, e-ticaret platformlarında kullanılan en popüler öneri sistemleri ve algoritmaları karşılaştırılmıştır.**

Anahtar Kelimeler—öneri sistemleri; e-ticaret için hibrit öneri sistemleri; e-ticaret alışverişleri

I. Introduction

The reshaping of business practices and consumer habits together with the high development of technology has led to a rapid growth in e-commerce applications. One of the most important purposes of e-commerce is to ensure that products or services are delivered to users quickly and without any problems. Factors such as lower costs, savings in time, and reduction in storage costs have made e-commerce activities widespread. The efforts of e-commerce sites to increase their product and service sales potential have led to the emergence of the Recommendation System. Recommendation Systems is a system that aims to provide personalized products and services to users and provides easy purchasing opportunities by using various machine learning algorithms [1]. The increasing variety of products and the fact that each user has different experiences and tastes has created the question of which product will be recommended to whom. Suggestion Systems solve this problem by predicting the interests of the users and recommending the most suitable product for them.

Recommendation systems learn from users' choices by taking into account the data provided by users and their preferences, interests and experiences. In this information, it finds the match between the user and the product-service, imposes the similarity between the user and the items, and offers them the most appropriate suggestion in line with the needs of the users [2].

Recommendation systems have also become very important as many large companies such as Youtube, Spotify, Netflix, Amazon have improved their service delivery practices. In 2009, Netflix offered a $1 million prize by holding a competition because it planned to produce a recommendation system that outperformed the algorithm it used [3]. The knowledge that another big company, Amazon, also generates 35% of its revenue from recommendation systems, reveals how important recommendation systems are in being able to compete with other companies and increase their sales [4].

Recommendation systems work with two basic types of information. These are characteristic information and user-item interactions. The characteristic information is information about the item and users. While the item consists of elements such as category and keyword, the information about the user includes the interests and likes of the users. The results reflecting the preferences are reached by taking into account the information provided by the users during registration and the actions and transactions they have made within the site afterwards. User-item interactions include elements such as the number of likes, ratings, comments and purchases on products by users. Ratings can be either 1-5 or 1-10 stars, or two, such as like-dislike [1].

There are four different recommendation systems: Collaborative Recommendation Systems, Content-Based Recommendation Systems, Popularity-Based Recommender Systems, and Hybrid Recommendation Systems. In this study, the use of these four different recommendation systems in e-commerce sites was examined and the algorithms were compared with each other.

Collaborative Recommendation Systems is one of the most widely used recommendation systems. It is purely based on past experience and the similarity between two users' preferences and tastes [5]. This recommendation system collects and analyzes information about users' interests and tries to predict which products users will like, taking into account the similarity with other users. When it finds similar users, it analyzes how much the likes overlap with each other and offers products accordingly [6].

In Content-Based Recommendation Systems, not only the user-item interaction but also the item information that the user has interacted with in the past is utilized [3]. This recommendation system is based on the idea of presenting items that are similar to users' past tastes. For example, it is recommended to the user by checking the movies that are similar to the content of the movie the user is watching [7].

Popularity-Based Recommendation Systems reveal popular or trending items among users and recommend them directly to the user. Unlike other recommendation systems, it is not affected by the user's preferences [2].

Hybrid Recommender Systems are based on the combined use of Collaborative Recommender Systems and Content-Based Recommendation Systems [7]. In order to achieve the best performance, a model can be created and used by adding content-based features to Collaborative Suggestion Systems or by applying vice versa scenarios [4]. By using these two systems together, it is aimed to get more efficient results by preventing the weaknesses of each of the systems. Big companies such as Netflix, Amazon, Spotify benefit greatly from this system. For example, Netflix offers suggestions both by checking users' past actions and taking into account the ratings made by users [8].

A. *Problems in Recommendation Systems*

The inclusion of new users and products in the system, the increase in user evaluations about the product, and the increased data may cause some problems. Data sparsity is one of these problems. It occurs when similarity recognition between users is insufficient despite appropriate data [9]. Not having enough information about users who are new to the system creates an uncertain situation about the products to be offered to them [10]. In this study, we tried to reveal the right recommendation system to avoid certain problems encountered in the system and to make appropriate recommendations to the right users.

II. Materials & Methods

Python libraries (pandas, NumPy and sci-kit learn) were used separately for Collaborative Recommendation Systems, Content-Based Recommendation Systems and Hybrid Recommendation Systems.

The NumPy library is like Python but more functional and used for scientific calculations. The Pandas library is a library that can be effectively used in operations with non-homogeneous sequences. Sci-kit learn is a convenient library for machine learning-oriented operations such as regression, matrices, and vectors. Cosine similarity is calculated by formulating the similarity between two different texts with the cosine function. SVD allows operations to be made with a complex matrix as the product of simpler matrices. In TfidfVectorizer, the processed words are converted to TF / IDF matrix and the similarity of the products is measured by considering the product features.

The live database of an e-commerce website was used as the dataset which includes a total of 5012 different orders and 7190 products. Firstly, no recommendation system was used for the first 15 days. Next, only the collaborative recommendation system was used. After that, a content-based recommendation system was used. Lastly, the hybrid recommendation system was used in 15-day periods.

In the first period, 344 orders were received without using any recommendation system, and a total of 501 products were sold in these orders.

In the second period, the collaborative recommendation system was used, 498 orders and 776 products were sold. When the whole dataset is examined, 36 orders and 45 products are matched with the collaborative recommendation system.

In the third period, the content-based recommendation system was used, 730 orders and 1105 products were sold. When the entire dataset is analyzed, 126 orders and 247 products match the content-based recommendation system.

In the third (last) period, the hybrid recommendation system was used, 1399 orders and 2241 products were sold. When the entire dataset is examined, 248 orders and 420 products match the hybrid recommendation system.

A. *Comparison of Recommendation Systems*

Users with at least 1 order were evaluated. Accordingly, in the dataset with 4694 users, the number of users with at least 1 order was 4260, and the number of user interactions with at least 3 orders was 1254, with a total of 5012 orders. The total number of products is 378. In this study, TF/IDF technique (Term Frequency / Inverse Document Frequency) was used. However, for every product that the user interacts with;

- 30 products that the user does not interact with are sampled.
- The user is not aware of the existence of these products, because they have nothing to do with them.
- A list of 1 interactive and 30 non-interactive products is requested from the suggestion system.
- "Top N" accuracy measurements are made for the user and the interacted product from the list.
- All "first N grains" measurements are summed.

III. Results & Discussion

For Collaborative Recommendation Systems;

- Table I shows the number and proportions of products in orders and orders every other day, among 776 products out of 498 orders.
- In the cooperation suggestion system, the order product ratio is between 1 and 2. This means that at least 1 and at most 2 products are ordered in 1 order.

In the collaborative recommendation system, the rate of purchasing a product in an order is 1.5582.

Order	Number of orders	Number of Products	Products/Orders
Order 1	2	2	1
Order 2	1	1	1
Order 3	1	2	2
Order 4	3	4	1.33
Order 5	1	2	2
Order 6	4	4	1
Order 7	2	3	1.5
Order 8	2	2	1
Order 9	2	2	1
Order 10	3	5	1.66

Table I: Order and Products in Collaborative Recommendation System

For Content Based Recommendation Systems;

- Table II shows the number and proportions of the products in the order and every two days among 1105 products placed in 730 orders.
- In the content-based recommendation system, the order product ratio is between 1 and 3. This means that at least 1 and at most 3 products are ordered in 1 order.

In the content-based recommendation system, the rate of purchasing a product in an order is 1.5136.

Order	Number of orders	Number of Products	Products/Orders
Order 1	3	5	1.66
Order 2	3	4	1.33
Order 3	2	2	1
Order 4	1	3	3
Order 5	3	6	2
Order 6	2	3	1.5
Order 7	3	3	1
Order 8	3	4	1.33
Order 9	5	6	1.2
Order 10	4	5	1.25

Table II: Order and Products in Content Based Recommendation System

For Hybrid Recommendation Systems;

- Table III shows the number and proportions of products ordered every other day, among 2241 products placed in 1399 orders.
- In the content-based recommendation system, the order product ratio is between 1 and 8. This means that at least 1 and at most 8 products are ordered in 1 order.

In the hybrid recommendation system, the rate of purchasing a product in an order is 1.6018.

IV. Conclusions

The percentage of orders that do not use recommendation systems is 11.57 out of a total of 2971 orders in Table IV. The percentage of collaborative recommendation systems is 16.76. The percentage of content-based recommendation systems is 24.57. The percentage of hybrid recommendation systems is 47.08. The highest number of orders were received when using hybrid recommendation systems. In addition to the highest number of orders, the highest order/product ratio was also seen in hybrid recommendation systems. As future work, all e-commerce companies will be able to use recommendation systems. Currently, large companies (Amazon, Netflix) are successfully using recommendation systems. Medium-sized and small companies cannot use the recommendation systems due to the large size of the data and the high operating costs. Our renewed recommendation algorithms will make recommendation systems more accessible by managing big data with less cost.

Order	Number of orders	Number of Products	Products/Orders
Order 1	3	24	8
Order 2	4	11	2.25
Order 3	2	2	1
Order 4	1	1	1
Order 5	2	3	1.5
Order 6	2	3	1.5
Order 7	2	5	2.5
Order 8	3	4	1.33
Order 9	2	2	3
Order 10	1	2	2

Table III: Order and Products in Hybrid Recommendation System

Recommendation Systems	Products/Orders
Collaborative Recommendation Systems	1.5582
Content-Based Recommendation Systems	1.5136
Hybrid Recommendation Systems	1.6018

Table IV: Comparison of Methods with Each Other

Appendices

Author Contributions

This paper is a part of *H. Simsek*'s MSc thesis and *M. Yeniad* is the advisor of the thesis.

Acknowledgments

None declared.

Conflicts

None declared.

Ethical Declaration

This article does not contain any studies involving human participants and/or animals performed by any of the authors.

REFERENCES

[1] Rodriguez G. Introduction to recommender systems. 2018, Retrieved from https://tryolabs.com/blog/introduction-to-recommender-systems/

[2] Dwivedi R. What are recommendation systems in machine learning? 16 April 2020. [Online]. Retrieved from https://www.analyticssteps.com/blogs/what-are-recommendation-systems-machine-learning

[3] Rocca B. Introduction to Recommender Systems. June 3, 2019. Retrieved from https://towardsdatascience.com/introduction-to-recommender-systems-6c66cf15ada

[4] Cilingir I. Recommendation Systems. June 24, 2019. Retrieved from https://medium.com/@irmcilingir/%C3%B6neri-sistemleri-recommendation-systems-28a3f341c0a9

[5] Le J. The 4 Recommendation Engines That Can Predict Your Movie Tastes. May 1, 2018. Retrieved from https://towardsdatascience.com/the-4-recommendation-engines-that-can-predict-your-movie-tastes-109dc4e10c52

[6] Allen RB. User models: Theory, method, and practice. International Journal of Man-Machine Studies 1990; 32(5): 511-543.

[7] Ozkok H. Recommendation Engine. April 17, 2020. Retrieved from https://www.datasciencearth.com/recommendation-engine-tavsiye-oneri-sistemleri/

[8] Guler S. Recommendation Systems and the Usage of Recommendation Systems in E-commerce. MSc Thesis, Sakarya University, 2019.

[9] Papagelis M, Plexousakis D, Kutsuras T. Alleviating the sparsity problem of collaborative filtering using trust inferences. In iTrust'05 Proceeding of Third International Conference on Trust Management, May 2005, pp. 224-239.

[10] Bobadilla J, Ortega F, Hernando A, Bernal J. A collaborative filtering approach to mitigate the new user cold start problem. Knowledge-based systems 2012; 26: 225-238.