UDC 004.93(045)
DOI:10.18372/1990-5548.68.16088

[1]V. M. Sineglazov,
[2]Y. I. Oliinyk

ALGORITHMS FOR THE FORMATION OF RECOMMENDATIONS IN THE INFORMATION SYSTEM

Aviation Computer-Integrated Complexes Department, Faculty of Air Navigation Electronics and Telecommunications, National Aviation University, Kyiv, Ukraine
[2]Faculty of Program Engineering, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute," Kyiv, Ukraine
E-mails: [1]svm@nau.edu.ua ORCID 0000-0002-3297-9060, [2]onyfrak1999@gmail.com

Abstract**—***The article deals with the problem of scalability and dimension reduction of data in the algorithms of recommendations. It is proposed to improve the item-to-item algorithm by excluding from the user-item matrix elements that that do not have enough estimates. Thus more denser data are used that allows to receive more exact results. Also due to the fact that the dimension of the user-item matrix decreases, the execution time of the algorithm decreases. To solve the problem, the Tachimoto coefficient, the cosine measure, the Pearson correlation coefficient and the Euclidean distance are used to calculate the degree of similarity of the elements. The efficiency of the usual item-to-item algorithm and the algorithm were compared using only the active values in the user-item matrix. The obtained results confirm the efficiency of the item-to-item algorithm based on a dense matrix. The obtained results can be used to optimize the operation of any recommendation system.*

Index terms—Algorithm; filtration; matrix dimensios; recommendation system; sparsity.

I. INTRODUCTION

E-commerce markets have actively introduced an automated personalization service to analyze the customer's behavior and patterns as purchase factors. E-commerce sites try to collect various users' interests, such as purchase history, product information in the cart, product ratings, and product reviews in order to recommend new relevant products to customers. Collaborative filtering is the most commonly used algorithm to build personalized recommendations on the website including Amazon, CDNOW, Ebay, Moviefinder, and Netflix beyond academic interest.

II. PROBLEM STATEMENT

There are three common approaches to solving the recommendation problem: traditional collaborative filtering, cluster models, and search-based methods.

Need resolve data sparsity problem in collaborative filtering, that allows scales independently of the number of customers and number of items in the product catalog and algorithm produces recommendations in realtime, scales to massive data sets, and generates highquality recommendations.

III. EXIST SOLUTIONS

Recommendation systems in various applications have tried to provide users with an accurate recommendation to meet the needs of the user and to bring higher benefits to companies. Collaborative filtering is an effective and well known technology in recommendation systems. Many web sites, particularly Ecommerce sites, have used collaborative filtering technology in their recommendation systems to personalize the browsing experience for each user as seen. As successful use cases of collaborative filtering, Amazon increased sales by 29% [1], Netflix increased movie rentals by 60% [2], and Google news increased click-through rates by 30.9% [3]. Collaborative filtering (CF) can be categorized into two main methods as user-based collaborative filtering (memory-based) and item-based collaborative filtering (model-based) [4].

User-based collaborative filtering (UBCF) approach is to predict items to the target user that are already items of interest for other users who are similar to the target user. For example, as seen Fig. 1 [5], let User 1 and User 3 have very similar preference behavior. If User 1 likes Item A, UBCF can recommend Item A to User 3. UBCF needs the explicit rating scores of items rated by users to calculate similarities between users and exploits *k*-nearest neighbor algorithms to find the nearest neighbors based on user similarities. And then, it generates prediction in terms of items by combining the neighbor user's rating scores based on similarity weighted averaging.

Item-based collaborative filtering (IBCF) approach is to predict items by inquiring into similarities between the items and other items that are already associated with the user. For example, as seen in Fig. 2 [5], let's say Item *A* and Item *C* are very similar. If a User likes Item *A*, IBCF can recommend Item *C* to the User. IBCF needs a set of items that the target user has already rated to calculate similarities between items and a target item. And then, it generates prediction in terms of the target item by combining the target user's previous preferences based on these item similarities [4]. In IBCF, users' preference data can be collected in two ways. One is that user explicitly gives rating score to item within a certain numerical scale. The other is that it implicitly analyzes user's purchase records or click-through rate [6].

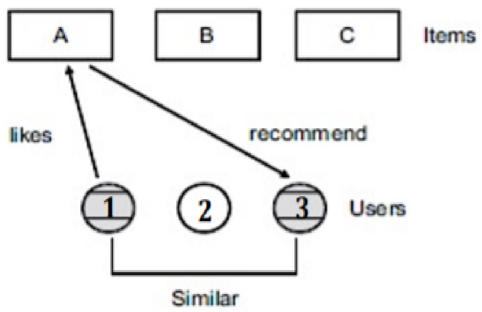

Fig. 1. User-based collaborative filtering (4)

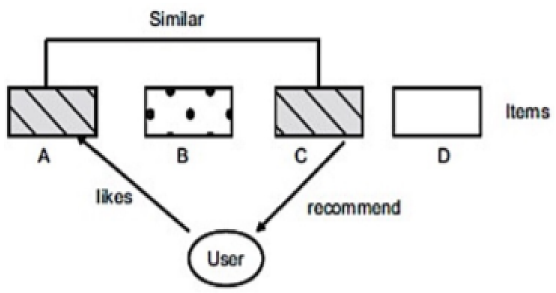

Fig. 2. Item-based collaborative filtering

IV. Solution Problem

User-based collaborative filtering is easy to implement and good to scale correlated items [7]. However, as stated previously above, it comes up against a couple of problems: data sparsity and data scalability. Data sparsity problem could lead to a skewed prediction and low reliability of predictions. Besides, data scalability requires low operation time and high memory feature to scale with all users and items in the database. To address these issues in UBCF, this paper proposes IBCF approach applying dimension reduction.

Enormous users and products have been added at E-commerce domains. A typical example is Amazon. Amazon added 30 million new customers in 2013 and had had over 244 million active customers as Geekwire reported in 2014 [8]. Also, Amazon had sold over 200 million products as ReportX reported in 2013 [9]. Currently in 2015, it is expected that Amazon would have more than these numbers of users and products. If the recommendation system using UBCF at 22 Amazon should look into all datasets similar to a 244 million × 200 million matrix, it will encounter data scalability and data sparsity issues. In UBCF, more the number of users and items increase, more the number of matrix dimensions increase and runtime takes long to find nearest neighbor of users. Therefore, it is assumed that using denser data having much more preference information given by users with IBCF effectively addresses data scalability and data sparsity problems. To focus on active items assuming that they have many ratings given by users, matrix is required to reduce dimension in IBCF without regard to passive items.

Item / *User*	*I1*	*I2*	*I3*	*I4*	*I5*
U1	2.0	4.0	3.0	3.0	3.0
U2			1.0	3.0	
U3	5.0	1.0	5.0	5.0	
U4	4.0		3.0		

Fig. 3. User-Item matrix before dimension reduction

Item / *User*	*I1*	*I3*	*I4*
U1	2.0	3.0	3.0
U2		1.0	3.0
U3	5.0	5.0	5.0
U4	4.0	3.0	

Fig. 4. User-Item matrix after dimension reduction

For instance, as seen in Fig. 3, each item can get up to a maximum of 4 ratings by users. Item I2 has 2 ratings and Item I5 has 1 rating, which means the number of ratings for Item I2 and Item I5 is not bigger than half of the total number of ratings. We can assume that Item I2 and Item I5 do not carry much weight with this matrix. Hence, when matrix has impactful items like Item I1, Item I3, and Item I4 as seen in Fig. 4, running time of the recommendation system in computing similarity between items and to provide more accurate prediction is expected to reduce.

V. Performance Test

In this chapter, I implemented item-based collaborative filtering applying dimension reduction (R-IBCF). The goal of the proposed R-IBCF is to provide better quality of prediction in terms of the

MAE measure and to make faster execution time. I compared the R-IBCF algorithm to IBCF in order to find an optimal similarity algorithm and training/test ratio of the dataset. Also, I selected an optimal value of the number of ratings per item on R-IBCF as I varied the value of it. In addition, I implemented UBCF as a benchmark to compare runtime of R-IBCF and IBCF to UBCF and the quality of prediction with optimal parameters.

A. Optimum Similarity Measurement

It was implemented four different similarity measurements: Cosine vector similarity, Pearson correlation coefficient, Euclidean distance, and Tanimoto coefficient. For each similarity algorithms, was measured MAE to find an optimal similarity on IBCF and R-IBCF for this dataset.

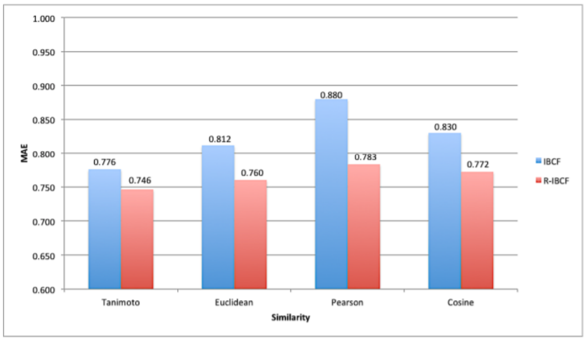

Fig. 5. The impact of the similarity computation on IBCF and R-IBCF

Figure 5 shows the experiments results. I observed that IBCF applying dimension reduction generally produced better quality of predictions more than IBCF with four similarity measurements. In particular, Tanimoto coefficient has a clear advantage, as MAE is the lowest on IBCF and R-IBCF. Therefore, I select Tanimoto coefficient similarity for the rest of my experiments.

B. Comparison of Runtime with Benchmark

Runtime of performance is also an important point in terms of data scalability. I implemented R-IBCF consuming memory. I ran each experiment with four similarity algorithms 30 times and got the average of their runtime excluding the first 5 times. These results are shown in Fig. 6. Even though it takes more time to filter data based on the number of ratings per item, I observed that it is faster 39 than computing similarity between all co-rated items or all users. Therefore, reduction of dimension on IBCF has considerable impact on runtime being fast in terms of data scalability. In addition, because IBCF and R-IBCF only consider corated items to compute similarity, they do not take finding the nearest neighbors step. Therefore, it generally influences on better runtime of IBCF and R-IBCF by comparison with UBCF.

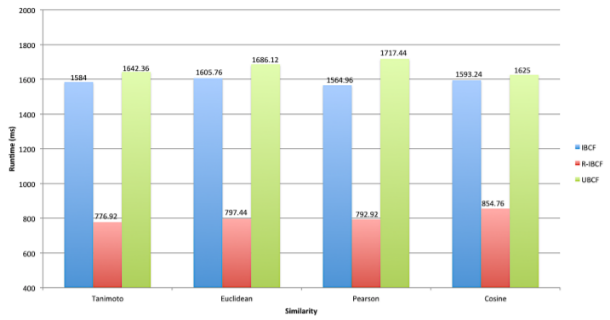

Fig. 6. Comparison of runtime of IBCF, R-IBCF, and UBCF

VI. CONCLUSIONS

Recommendation systems have been an important in E-commerce on the web for the customer to suggest items what they would be interested. With the increasing number of users and items, recommendation systems encounter the main shortcoming: data sparsity and data scalability problems, which bring out the poor quality of prediction and the inefficient time consuming. In this paper, I have proposed item-based collaborative filtering approach applying dimension reduction to improve the predictive accuracy and recommendation quality in overcoming the existing limitations. By reducing the noise of dimensional data, it focuses on typical and popular items to compute the similarity between them and to predict the most similar items to users. The experimental results show that this approach makes a considerable impact on providing better accuracy of prediction and much faster execution time in comparison with traditional UBCF and IBCF. It results in improving the quality of recommendation system using collaborative filtering. The potential limitation would use this approach with dataset widely consisting of not enough ratings by users, expecting less accuracy. Therefore, to overcome this challenge, I propose an approach to mix both explicit and implicit ratings to alleviate the data sparsity problem further in this aspect.

REFERENCES

[1] J. P. Mangalindan, "Amazon'S Recommendation Secret," Fortune. 2012. http://fortune.com/2012/07/30/amazons-recommendation-secret/.

[2] Yehuda Koren, "Collaborative Filtering with Temporal Dynamics," *15th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD 09)*, ACM, 2009, pp. 447–455. https://doi.org/10.1145/1557019.1557072

[3] Jiahui Liu, Peter Dolan, and Elin Ronby Pedersen, "Personalized news recommendation based on click behavior," in: Rich, et al. (eds.) *In the 14th Int. Conf. on Intelligent User Interfaces (IUI)*, *ACM*, 2010: 31–40. https://doi.org/10.1145/1719970.1719976

[4] George Kaypi Sarwar, Konstan Joseph, and John Riedl, "Item-based Collaborative Filtering Recommendation Algorithms," *in the 10th International World Wide Web Conference*, 2001, pp. 285–295. https://doi.org/10.1145/371920.372071

[5] Sachin Walunj and Kishor Sadafale, "An online recommendation system for e-commerce based on apache mahout framework," *Proceedings of the 2013 annual conference on Computers and people research, ACM*. 2013, pp. 153–158. https://doi.org/10.1145/2487294.2487328

[6] Songjie Gong, "A Collaborative Filtering Recommendation Algorithm Based on User Clustering and Item Clustering," *JSW* 5(7), 2010. https://doi.org/10.4304/jsw.5.7.745-752

[7] Xiaoyuan Su, and Taghi M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques," Advances in Artificial Intelligence, 2009: 1–19. https://doi.org/10.1155/2009/421425

[8] Tricia Duryee, "Amazon Adds 30 Million Customers In the Past Year-Geekwire," *Geekwire*. 2014. http://www.geekwire.com/2014/amazon-adds-30-millioncustomers-past-year/

Received: March 16, 2021

Sineglazov Victor. ORCID 0000-0002-3297-9060. Doctor of Engineering Science. Professor. Head of the Department.
Aviation Computer-Integrated Complexes Department, Faculty of Air Navigation Electronics and Telecommunications, National Aviation University, Kyiv, Ukraine.
Education: Kyiv Polytechnic Institute, Kyiv, Ukraine, (1973).
Research area: Air Navigation, Air Traffic Control, Identification of Complex Systems, Wind/Solar power plant, artificial intelligence.
Publications: more than 660 papers.
E-mail: svm@nau.edu.ua

Oliinuk Yuriy. Magister.
Faculty of Program Engineering, National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute," Kyiv, Ukraine.
Education: National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute," Kyiv, Ukraine, (2020).
Research area: artificial intelligence.
E-mail: onyfrak1999@gmail.com

В. М. Синеглазов, Ю. І. Олійник. Алгоритми формування рекомендацій в інформаційній системі

У статті розглядається проблема масштабованості та розрідженості даних в алгоритмах рекомедацій. Запропоновано вдосконалення алгоритму item-to-item за допомогою виключення з user-item матриці елементів які мають мало оцінок. Таким чином використовуються більш щільні, що дозволяє отримати більш точні результати. Також за рахунок того, що зменшується розмірність user-item матриці зменшується час виконання алгоритму. Для вирішення задачі використовується коефіцієнт Тахімото, косинусна міра, коефіцієнт кореляції Пірсона та Евклідова відстань для обрахування міри схожості елементів. Було здійснено порівняння ефективності роботи звичайного алгоритму item-to-item і алгоритму з використанням лише активних значень в user-item матриці. Отримані результати підтверджують ефективність item-to-item алгоритму основаному на щільній матриці. Отримані результати можуть бути використані для оптимізації роботи будь-якої рекомендаційної системи.

Ключові слова: алгоритм; рекомендаційна система; розмірність матриці; розрідженість; фільтрація.

Синєглазов Віктор Михайлович. ORCID 0000-0002-3297-9060.
Доктор технічних наук. Професор. Завідувач кафедрою.
Кафедра авіаційних комп'ютерно-інтегрованих комплексів, Факультет аеронавігації електроніки і телекомунікацій, Національний авіаційний університет, Київ, Україна.
Освіта: Київський політехнічний інститут, Київ, Україна, (1973).
Напрям наукової діяльності: аеронавігація, управління повітряним рухом, ідентифікація складних систем, вітроенергетичні установки, штучний інтелект.
Кількість публікацій: більше 660 наукових робіт.
E-mail: svm@nau.edu.ua

Олійник Юрій Іванович. Магістр.
Факультет програмної інженерії, Національний технічний університет України «Київський політехнічний інститут ім. Ігоря Сікорського», Київ, Україна.
Освіта: Національний технічний університет України «Київський політехнічний інститут ім. Ігоря Сікорського», Київ, Україна, (2020).

Напрям наукової діяльності: штучний інтелект.
E-mail: onyfrak1999@gmail.com

В. М. Синеглазов, Ю. І. Олейник. Алгоритмы формирования рекомендаций в информационной системе

В статье рассматривается проблема масштабируемости и разреженности данных в алгоритмах рекомендаций. Предложено усовершенствование алгоритма item-to-item с помощью исключения из user-item матрицы элементов, имеющих мало оценок. Таким образом используются более плотные, что позволяет получить более точные результаты. Также за счет того, что уменьшается размерность user-item матрицы, уменьшается время выполнения алгоритма. Для решения задачи используется коэффициент Тахимото, косинусная мера, коэффициент корреляции Пирсона и Евклидова расстояние для вычисления степени сходства элементов. Было произведено сравнение эффективности работы обычного алгоритма item-to-item и алгоритма с использованием только активных значений в user-item матрице. Полученные результаты подтверждают эффективность item-to-item алгоритма, основанного на плотной матрице. Полученные результаты могут использоваться для оптимизации работы любой рекомендательной системы.

Ключевые слова: алгоритм; рекомендательная система; размерность матрицы; разреженность; фильтрация.

Синеглазов Виктор Михайлович. ORCID 0000-0002-3297-9060.
Доктор технических наук. Профессор. Заведующий кафедрой.
Кафедра авиационных компьютерно-интегрированных комплексов, Факультет аэронавигации электроники и телекоммуникаций, Национальный авиационный университет, Киев, Украина.
Образование: Киевский политехнический институт, Киев, Украина, (1973).
Направление научной деятельности: аэронавигация, управление воздушным движением, идентификация сложных систем, ветроэнергетические установки, искусственный интеллект.
Количество публикаций: более 660 научных работ.
E-mail: svm@nau.edu.ua

Олейник Юрий Иванович. Магистр.
Факультет программной инженерии, Национальный технический университет Украины «Киевский политехнический институт им. Игоря Сикорского», Киев, Украина.
Образование: Национальный технический университет Украины «Киевский политехнический институт им. Игоря Сикорского», Киев, Украина, (2020).
Направление научной деятельности: искусственный интеллект.
E-mail: onyfrak1999@gmail.com