Proceedings of the 7th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2023)
IEEE Xplore Part Number: CFP23OSV-ART; ISBN: 979-8-3503-4148-5

2023 7th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) | 979-8-3503-4148-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/I-SMAC58438.2023.10290568

Model-based Filtering Techniques for Recommendation Systems in Healthcare Domain

Sinchana N M[1], Prasanna Kumar K R[2], Santhosh B J[3]

[1] Software Engineer, Blackline Technologies, Banglore, Karnataka, India,
sinchananm19@gmail.com
[2] Assistant Professor, Department of Computer Science and Engineering, Siddaganga Institute of Technology, Tumkur, Karnataka, India,
prasanna.kghatta@sit.ac.in
[3] Assistant Professor, Department of Computer Science and Engineering, BGS Institute of Technology, Adichunchanagiri University, Mandya, 571 448, Karnataka, India
Santhoshbj221@gmail.com

Abstract— **A drug recommendation system is a technology-based solution that assists healthcare professionals in suggesting appropriate medications for patients based on various factors such as medical history, symptoms, demographics, and drug effectiveness. The system utilizes advanced algorithms and techniques to analyse large datasets, including patient information, drug profiles, clinical studies, and drug-drug interactions, among others. This research study presents a drug recommendation system that utilizes model-based filtering techniques, specifically Singular Value Decomposition (SVD) and Non-Negative Matrix Factorization (NMF). The system aims to enhance the accuracy and effectiveness of drug recommendations by analysing large diabetes patient datasets containing patient information, drug profiles, and clinical studies. To evaluate the system's performance, three standard evaluation metrics are employed. Furthermore, a visual comparison of the evaluation metrics was presented using a bar graph, which clearly demonstrates the superiority of SVD over NMF. The findings contribute to the advancement of personalized healthcare and emphasize the importance of employing suitable algorithms for accurate medication suggestions.**

Keywords— Recommendation System, Drug, Collaborative Filter, Error, Accuracy, UCI Repository.

I. INTRODUCTION

Recommendation systems in healthcare analyse patient data and utilize algorithms to provide personalized recommendations. They suggest suitable drugs based on patient characteristics and medical history. Treatment recommendation systems offer guidance on appropriate treatments considering patient symptoms and evidence-based medicine. Clinical decision support systems assist in diagnostic and treatment decisions. Personalized health monitoring systems offer recommendations for preventive care and lifestyle modifications. Health insurance plan recommendation systems help select appropriate insurance plans, and telemedicine recommendation systems aid in finding suitable healthcare providers [1]. These systems enhance healthcare decision-making by tailoring recommendations to individual needs and improving patient outcomes. In the healthcare domain, there are various recommendation systems available, one of which is the drug recommendation system. The need for drug recommendation arises due to several reasons, including [2]:

- Each patient may have unique medical conditions, treatment history, and genetic factors. A drug recommendation system can help healthcare professionals identify suitable medications tailored to an individual's specific needs, improving treatment outcomes.
- Certain medications can interact with each other, leading to potential adverse effects or reduced efficacy. A drug recommendation system can assist in avoiding harmful drug combinations by providing insights into potential interactions and suggesting alternative options.
- With a vast range of drugs available, healthcare providers may find it challenging to stay updated on all the latest medications. A recommendation system can offer suggestions based on patient characteristics, medical history, and treatment guidelines, thereby streamlining the decision-making process and ensuring optimal treatment efficiency.

Collaborative Filtering (CF) is a popular approach employed in recommendation systems for healthcare, including drug recommendation. CF leverages the collective knowledge and experiences of a group of patients to make recommendations [3]. Here are some reasons why CF is effective in healthcare recommendation systems: 1) Healthcare systems often contain large volumes of patient data, including treatment records, medical histories, and patient outcomes. CF utilizes this abundant data to identify patterns and similarities among patients, enabling accurate drug recommendations. 2) CF takes into account the preferences and experiences of similar patients. By analysing historical data, it identifies patients with similar characteristics and recommends drugs that have been effective for others with comparable profiles. This personalized approach enhances the likelihood of successful treatment outcomes. 3) CF compensates for limited or incomplete information about individual patients by drawing insights from similar patients. It can bridge knowledge gaps, especially when specific patient data may be missing or insufficient for precise recommendations.

In summary, drug recommendation systems are crucial in the healthcare domain to provide personalized treatment options, prevent drug interactions, and optimize treatment efficiency. CF is a valuable technique employed in recommendation systems, leveraging patient data to offer accurate and personalized drug recommendations based on the experiences and outcomes of similar patients. Few research on recommendation systems in the healthcare system is detailed below.

The paper [4] provides a methodical, innovative, and collaborative strategy based on patient-centred technology, including a recommender system structure, to improve health service levels in accordance with medical specializations. Based on each individual's unique user profile, the system assesses local medical facilities and gives recommendations. To semantically compute the similarity of medical specializations and provide healthcare services with response-ability, medical service kind, and near-user location, they suggest a health attention parameter. To address the difficulties of mobile recommendations and the wide variety of objects seen in healthcare, the method integrates semantic and spatial processing. The Web-GIS application used to display the filtered healthcare facilities underwent evaluation in numerous areas of Mexico City. Journal [5,] provides a systematic review of healthcare systems for recommendation. The study differs from others in the sector in that it dives further into unique recommendation scenarios and methodologies. Advice on what to eat and take, how to anticipate health, where to go for medical care, and which doctors to see are some examples. They also build functioning examples to provide a full understanding of recommendation systems. Finally, some of the challenges are discussed that may arise during future healthcare recommender system research and development.

The study [6] looks into how to incorporate various recommendation techniques, including CF, content, and knowledge-based filtering, within a hybrid framework to take benefit of each technique's capabilities for personalized mental health recommendations. Factors like data integration and algorithm selection are discussed in this study as they pertain to developing a Hybrid Recommender System. The research also delves into the key data sources utilized by mental health hybrid recommender systems, as well as the evaluation criteria utilized to ascertain the effectiveness of the system. This study's results help advance the growth of individualized mental health care and the creation of effective, individualized recommendation systems. In the paper [7], the author presents a cutting-edge health recommendation framework constructed with the Internet of Things, Deep Learning (DL), and Restricted Boltzmann Machine (RBM). This DL + RBM model can recognize and evaluate data from a patient's medical decision-making tool, enabling doctors to identify if the patient is experiencing a serious health problem and provide appropriate treatment. Cross-validation tests are increasingly being utilized to investigate the proposed system's behaviour, as they establish a number of normative measures that are relevant across populations. This architecture can intelligently recognize and evaluate patient data with the help of a healthcare DSS. The experimental findings support the proposed system's efficiency and intelligence in providing medical care. The findings of this investigation provide empirical evidence for the concept. This device is an inexpensive choice for those who live in distant areas; everyone can utilize it to assess the severity of their medical issue and, if necessary, call local hospitals.

In work [8], an ML method is suggested as a way to give patients with multiple diseases effective treatment suggestions. This method provides useful counsel for anyone suffering from ophthalmic, cold, obesity, cardiac, fever, or ortho difficulties. Supervised ML techniques were used to make recommendations for patients. Testing and analysis were performed using a dataset developed specifically for this reason and gathered from medical practitioners. This experimental evaluation reveals that the Random Forest categorization technique outperforms other methods in terms of suggestion accuracy. As a result, the proposed method is regarded as a valuable resource for offering trustworthy patient recommendations in the healthcare sector. In the publication [9], a hybrid AI-based technique for processing massive amounts of data was given. They made advantage of ML to provide people with dietary recommendations. Hybrid Recommender System is the name of the system being suggested. In this approach, NLP and ML are brought together. To analyse data and provide nutritional advice, a unique algorithm called Intelligent Recommender for Healthy Diet (IR-HD) is developed. IR-HD has the ability to provide better healthy eating recommendations than current models. The Python data science framework serves as the foundation for the recommender system. The trial results confirmed the system's capacity to make solid recommendations as well as its superior performance when compared to the state of the art.

II. PROCESS OF RECOMMENDATION SYSTEM

This section provides a detailed explanation of the drug recommendation system's process, which utilizes CF techniques.

A. Data collection and preparation

The employed data set contains information on patients with diabetes-related disorders, and it was collected from the UCI ML Repository [10]. There are over 50 features in the raw data set representing results for patients across 130 hospitals in the US. This dataset contains over a hundred thousand individual patient records spanning a decade of medical history (from 1999 to 2008). The collected data from the UCI repository is given in Figure 1.

Out[5]:

	encounter_id	patient_nbr	race	gender	age	weight	admission_type_id	discharge_disposition_id	admission_source_id
0	2278392	8222157	Caucasian	Female	[0-10)	?	6	25	1
1	149190	55629189	Caucasian	Female	[10-20)	?	1	1	7
2	64410	86047875	AfricanAmerican	Female	[20-30)	?	1	1	7
3	500364	82442376	Caucasian	Male	[30-40)	?	1	1	7
4	16680	42519267	Caucasian	Male	[40-50)	?	1	1	7

Figure 1. Data Sample

Data preparation is an important first process in correctly analysing the supplied data. Before the original data set was filtered, all duplicate patient records were removed, and those who had not filled any prescriptions were crossed off the list [11]. Then we converted the data and chose which variables to investigate.

Variables are frequently transformed to a new data type and new variables are produced during data transformation. We accomplished this by converting our categorical variables to binary variables. New variables were created for the following conditions based on the first three numbers of the disease classification: cardiovascular, diabetic, gastrointestinal, genitourinary, musculoskeletal, neoplastic, respiratory, and other. New variables were created to represent the various races. Following that, racial and diagnostic groups were removed from the dataset. Furthermore, the median of the age ranges was used rather than the ranges themselves.

Non-informative features were removed from the data set due to a high number of missing values, features that are useless for classifying the data (such as patient identity), and imbalanced features. Furthermore, persons who were using numerous medicines were highlighted . The final dataset had 5177 individual patient records as well as 42 variables categorised as demographics, clinical history, and medicines. Drugs with a small proportion of administration or those that have never been administered to a patient were eliminated. After the data was processed, the drugs like Metformin, Pioglitazone, Nateglinide, Glyburide-metformin, Glimepiride, Repaglinide, Rosiglitazone, Glyburide, Glipizide, and Insulin were finalized.

B. *Recommendation Model*

SVD: The SVD model is a popular technique used in drug recommendation systems. It leverages matrix factorization to analyse the relationships between patients, drugs, and their preferences. Here's an explanation of the SVD model along with the equations involved [12]:

1. Creating the User-Drug Preference Matrix: The first step is to create a user-drug preference matrix. Each cell in the matrix represents a user's preference for a particular drug. The matrix dimensions are $(m\, x\, n)$, where m is the number of patients and n is the number of drugs. The entry in cell (i, j) represents the preference of patient i for drug j.

2. Applying SVD: SVD is applied to decompose the user-drug preference matrix into three matrices: $U, \Sigma,$ and V^T [13].
   - $U\,(m * k)$: The left singular matrix represents the patient's latent features. Each row of U represents a patient, and the columns (k) capture the patient's preferences across latent factors.
   - $\Sigma\,(k * k)$: The diagonal matrix Σ holds the singular values, which quantify the importance of each latent factor. The singular values are arranged in decreasing order.
   - $V^T\,(k * n)$: The right singular matrix represents the drug's latent features. Each row of V^T corresponds to a drug, and the columns capture the drug's characteristics across latent factors.

3. Approximating the User-Drug Preference Matrix: To recommend drugs to a specific patient, the original matrix is approximated using a reduced number of latent factors (k). The approximation is given by the equation: $M \approx U\Sigma V^T$

4. Generating Drug Recommendations: To generate drug recommendations for a patient, the SVD model calculates the predicted preference score for drugs that the patient has not used or rated. The predicted preference score is determined using the equation: $Score\,(i, j) = \sum \left(U(i, k) * \Sigma(k, k) * V^T(k, j)\right)$

5. Selecting Top Recommended Drugs: The model selects the top-recommended drugs based on the highest predicted preference scores for the patient. These drugs are then suggested as potential recommendations.

By employing the SVD model in drug recommendation systems, healthcare providers can leverage patient preferences and latent features to make personalized drug recommendations. This approach considers the relationships between patients and drugs, allowing for more accurate and targeted recommendations.

NMF: NMF can also be applied to recommendation systems, where it can be used to factorize a user-item matrix to uncover latent factors or features that capture user preferences and item characteristics. The NMF-based recommendation system aims to predict user ratings or preferences for items based on these latent factors [14].

Let's consider a user-item matrix R, where each entry $R[i, j]$ represents the rating or preference of the user i for item j. This matrix has dimensions $(m * n)$, where m is the number of users and n is the number of items. The goal of NMF in recommendation systems is to factorize this matrix into the product of two lower-rank non-negative matrices, $W\,(m * r)$ and $H\,(r * n)$, such that their product approximates the user-item matrix:

$$R \approx W * H \qquad [1]$$

Here, W represents the user matrix, and H represents the item matrix. The rank r determines the level of approximation and the number of latent factors to consider [15].

To find the optimal values for W and H, the NMF algorithm can be formulated as an optimization problem that minimizes the reconstruction error while considering additional regularization terms. One common objective function used in NMF-based recommendation systems is the squared error loss with regularization terms for W and H:

$$\text{minimize } 0.5 * ||R - W * H||^2 + \lambda_1 * ||W||^2 + \lambda_2 * ||H||^2 \quad [2]$$

Here, $|||^2$ denotes the Frobenius norm, which calculates the sum of squared differences between corresponding elements of the matrices. The regularization terms λ_1 and λ_2 are regularization parameters that control the trade-off between fitting the data and preventing over-fitting.

To solve this optimization problem, iterative algorithms like the multiplicative update rule can be applied. The update equations for NMF-based recommendation systems are similar to the standard NMF update equations, with the addition of the regularization terms:

$$H_{new} = H * \frac{W^T * R}{(W^T * W * H + \lambda_2 * H)} \quad [3]$$

$$W_{new} = W * \frac{W^T * H_{new}{}^T}{(W * H_{new}{}^T + \lambda_1 * W)} \quad [4]$$

The iterations continue until convergence is reached, and the resulting matrices W and H capture the latent factors or features in the user-item matrix. These factors represent underlying user preferences and item characteristics. Once the NMF algorithm converges, the predicted user-item ratings or preferences can be calculated by taking the dot product of the learned matrices W and H:

$$R_{pred} = W * H \quad [5]$$

The predicted ratings can then be used to generate recommendations for users by suggesting items with the highest predicted ratings that the users have not yet interacted with.

NMF-based recommendation systems provide a CF approach that can handle sparsity and scalability issues often encountered in recommendation tasks. By uncovering latent factors, NMF enables personalized recommendations based on user preferences and item characteristics, even in the absence of explicit feature information.

III. EXPERIMENTAL RESULT

The obtained results from the drug recommendation system, which was designed using model-based filtering techniques such as SVD and NMF, are discussed below. The system was evaluated using three different metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), and accuracy. To evaluate the system's performance, the first metric chosen was MSE. Table 1 presents the results of SVD and NMF on various folds. The mean MSE score achieved by SVD is 0.8484, while NMF obtained a mean MSE score of 0.90882. This suggests that SVD performs better in terms of minimizing the squared differences between predicted and actual drug recommendations. The second metric chosen for evaluation was MAE.

Table 1. Comparison of recommendation model using MSE

MODEL		SVD	NMF
Fold	1	0.838	0.917
	2	0.8427	0.9142
	3	0.8489	0.9068
	4	0.8534	0.9061
	5	0.859	0.9
Mean		0.8484	0.90882

Table 2 displays the outcomes of SVD and NMF on different folds. The mean MAE score obtained by SVD is 0.7565, whereas NMF achieved a mean MAE score of 0.8133. Similar to MSE, SVD outperforms NMF in terms of minimizing the absolute differences between predicted and actual drug recommendations. For the third evaluation metric, accuracy was selected.

Table 2. Comparison of recommendation model using MAE

MODEL		SVD	NMF
Fold	1	0.768	0.807
	2	0.747	0.8147
	3	0.7583	0.8214
	4	0.7486	0.819
	5	0.761	0.8048
Mean		0.75658	0.81338

Table 3 illustrates the results of SVD and NMF on various folds. The mean accuracy score achieved by SVD is 0.8579, while NMF obtained a mean accuracy score of 0.78686. In terms of accuracy, SVD again performs better than NMF, indicating that SVD provides more precise drug recommendations.

Table 3. Comparison of recommendation model using Accuracy

MODEL		SVD	NMF
Fold	1	0.842	0.795
	2	0.855	0.7858
	3	0.8604	0.7924
	4	0.8641	0.7812
	5	0.868	0.7799
Mean		0.8579	0.78686

To provide a visual comparison of all the mentioned metrics, Figure 2 presents a bar graph. Based on the graph, it is evident that SVD consistently outperforms NMF in terms of MSE, MAE, and accuracy, making it the best model among the two.

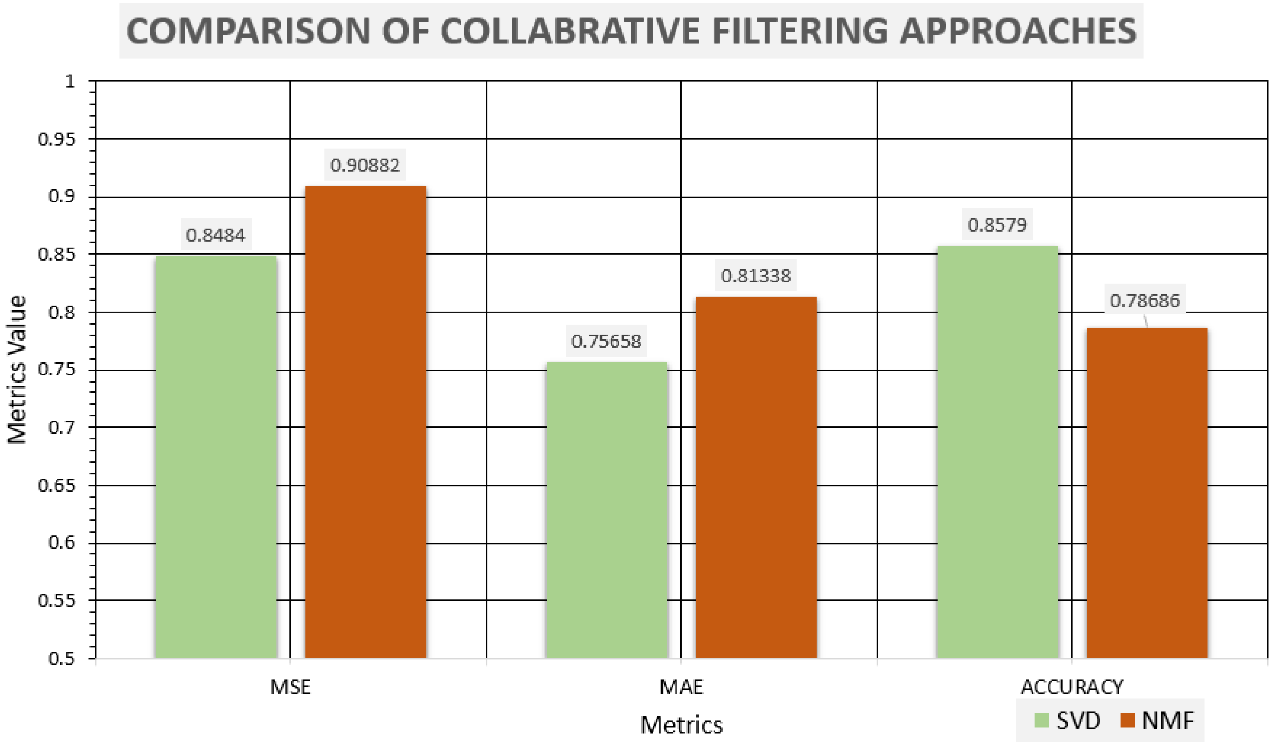

Fig. 2. Comparison of model-based filtering system.

Furthermore, Figure 3 showcases the outcome of drug recommendations by SVD for various patients. This provides a visual representation of the actual drug recommendations made by the system using SVD.

Patient ID	Drug Suggestion		
	Drug-1	Drug-2	Drug-3
32	Nateglinide	Insulin	Glyburide
8	Glyburide	Glimepiride	Metformin
11	Insulin	Repaglinide	Nateglinide

Fig. 3. Drugs recommended by SVD model.

IV. CONCLUSION

The designed drug recommendation systems employing model-based filtering methods like SVD and NMF were the primary focus of this investigation. MSE, MAE, and accuracy were used to assess the system's efficacy. Consistently, SVD surpassed NMF in terms of MSE, MAE, and accuracy, demonstrating its superiority in making more precise and accurate medicine recommendations. By reducing the variance between anticipated and actual drug recommendations, SVD has proven to be a superior method to NMF in drug recommendation systems. Despite the fact that the results of this study suggest that SVD is the better drug recommendation system, there are still many avenues for research and development to be pursued. Here are some potential paths for future study:

- Integration of contextual information: Considering contextual factors such as patient preferences, cost-effectiveness, and regional regulations can provide more comprehensive and tailored drug recommendations, taking into account individual patient needs and constraints.
- Real-time recommendation updates: Developing mechanisms to incorporate new patient data and update drug recommendations in real-time can ensure the system provides the most up-to-date and accurate suggestions.
- User feedback and validation: Collecting user feedback and conducting validation studies involving healthcare professionals and patients can help evaluate the practicality and effectiveness of the drug recommendation system using SVD. This feedback can be utilized to refine and optimize the system.

REFERENCES

[1]. Tran, Thi Ngoc Trang, Alexander Felfernig, Christoph Trattner, and Andreas Holzinger. "Recommender systems in the healthcare domain: state-of-the-art and research issues." Journal of Intelligent Information Systems 57 (2021): 171-201.
[2]. Garg, Satvik. "Drug recommendation system based on sentiment analysis of drug reviews using machine learning." In 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), pp. 175-181. IEEE, 2021.
[3]. Melville, Prem, Raymond J. Mooney, and Ramadass Nagarajan. "Content-boosted collaborative filtering for improved recommendations." Aaai/iaai 23 (2002): 187-192.
[4]. Torres-Ruiz, Miguel, Rolando Quintero, Giovanni Guzman, and Kwok Tai Chui. "Healthcare Recommender System Based on

Medical Specialties, Patient Profiles, and Geospatial Information." Sustainability 15, no. 1 (2022): 499.

[5]. Tran, Thi Ngoc Trang, Alexander Felfernig, Christoph Trattner, and Andreas Holzinger. "Recommender systems in the healthcare domain: state-of-the-art and research issues." Journal of Intelligent Information Systems 57 (2021): 171-201.

[6]. Idayati Mazlan, Noraswaliza Abdullah, Norashikin Ahmad, “Exploring the Impact of Hybrid Recommender Systems on Personalized Mental Health Recommendations”, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 14, No. 6, 2023

[7]. Kannan, Sharmilee, R. U. Anitha, M. Divayapushpalakshmi, and K. S. Kalaivani. "IoT-Deep Learning Based Activity Recommendation System." Computer Systems Science & Engineering 45, no. 2 (2023).

[8]. Komal Kumar, N., and D. Vigneswari. "A drug recommendation system for multi-disease in health care using machine learning." In International Conference on Advanced Communication and Computational Technology, pp. 1-12. Singapore: Springer Nature Singapore, 2019.

[9]. Lambay, Muhib Anwar, and S. Pakkir Mohideen. "A Hybrid Approach Based Diet Recommendation System using ML and Big Data Analytics." (2022).

[10]. Strack, Beata, Jonathan P. DeShazo, Chris Gennings, Juan L. Olmo, Sebastian Ventura, Krzysztof J. Cios, and John N. Clore. "Impact of HbA1c measurement on hospital readmission rates: analysis of 70,000 clinical database patient records." BioMed research international 2014 (2014).

[11]. Granda Morales, Luis Fernando, Priscila Valdiviezo-Diaz, Ruth Reátegui, and Luis Barba-Guaman. "Drug recommendation system for diabetes using a collaborative filtering and clustering approach: development and performance evaluation." Journal of Medical Internet Research 24, no. 7 (2022): e37233.

[12]. Kalman, Dan. "A singularly valuable decomposition: the SVD of a matrix." The college mathematics journal 27, no. 1 (1996): 2-23.

[13]. Ba, Qilong, Xiaoyong Li, and Zhongying Bai. "Clustering collaborative filtering recommendation system based on SVD algorithm." In 2013 IEEE 4th International Conference on Software Engineering and Service Science, pp. 963-967. IEEE, 2013.

[14]. Lee, Daniel, and H. Sebastian Seung. "Algorithms for non-negative matrix factorization." Advances in neural information processing systems 13 (2000).

[15]. Zhang, Sheng, Weihong Wang, James Ford, and Fillia Makedon. "Learning from incomplete ratings using non-negative matrix factorization." In Proceedings of the 2006 SIAM international conference on data mining, pp. 549-553. Society for Industrial and Applied Mathematics, 2006.