Journal of Information Systems Engineering and Management
2025, 10(2)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

An Improved Hybrid Recommendation System Algorithm for Resolving the Cold-Start Issues

Dr. A. Noble Mary Juliet[1], Dr. D. Sivaganesan[2], Dr. J. Bhavithra[3], Dr. N. Suba Rani[4], Dr. N. Senthil Madasamy[5]

[1]*Associate Professor, Department of CSE, Dr.Mahalingam College of Engineering and Technology, Pollachi, Tamil Nadu, India.*
cse.julie@gmail.com

[2]*Professor, Department of CSE, Dr.Mahalingam College of Engineering and Technology, Pollachi, Tamil Nadu, India.*
sivaganesand@gmail.com

[3]*Assistant Professor (SG), Department of CSE, Dr.Mahalingam College of Engineering and Technology, Pollachi, Tamil Nadu, India.*
bavi.rr@gmail.com

[4]*Associate Professor, Department of CSE, Dr.Mahalingam College of Engineering and Technology, Pollachi, Tamil Nadu, India.*
suba.drmcet@gmail.com

[5]*Associate Professor, Department of CSE, Dr.Mahalingam College of Engineering and Technology, Pollachi, Tamil Nadu, India.*
senkav1293@gmail.com

ARTICLE INFO

Received: 24 Oct 2024

Revised: 30 Dec 2024

Accepted: 24 Jan 2025

ABSTRACT

Online shopping has turned out to be very popular nowadays. Recommendation systems are decision aids to analyze customer's purchase sequences and their product information to provide customer preferences A sequential pattern mining method called the Prefix Span algorithm is used to find common sub-sequences that are longer than the minimal support requirements. Rules are constructed using frequent sequences to improve the performance for identify top-N prediction. A significant challenge faced by recommendation systems is the cold-start problem. The issue arises when the system does not have enough information to propose new users. This work tries to solve the issue of cold starting by incorporating sequential rules with the Bi-clustering approach. The recommendation system is evaluated using Precision, Recall, F1 measure and accuracy. Our investigation revealed that incorporating bi-clustering enhances performance and effectively resolves the cold-start problem.

Keywords: Recommendation, Cold-start problem, Sequential pattern mining, Prefix span algorithm, Bi-clustering.

INTRODUCTION

Online shopping has gained massive recognition these days. Customers can access a large volume of online products with just one click. It has changed the conventional method of shopping. It has been better for them to meet their wants by shopping online. E-commerce businesses work to balance the demands of customers and producers. Customers require trustworthy recommendation systems because of the explosive expansion of online information and e-commerce firms [6]. Recommendation is an information filtering system used to analyze user's past behavior and product ratings to recommend new user for purchasing products. The recommendation system is used to reduce the information overhead problems in variety of industries, including online shopping, e-learning, movies, books, news, and research publication [7].

A data mining method called sequential pattern mining is used to uncover common sequences and patterns in a sequential dataset [23]. [1] was the first to address about Sequential mining. This emerging field of data mining utilizes sequential data analysis to derive meaningful models from large databases. This method also used to locate sub-sequences in a given sequence database where the frequency of recurrence exceeds a minimum threshold. Sequence data mining has multitude applications in diverse discipline, from maintaining safety to health care [22] and from student management [21] to consumer behavior [24]. [2] Addressed Prefix Span technique to detect common sequential patterns. The concept behind this algorithm is, the sequence databaseis projected based on the prefixes by exploring the locally frequent sequences [15]. It gives better overall performance than FreeSpan, GSP [20], and SPADE algorithms [3]. Rule-Based Systems (RBS) arealso referred as Expert systems (ES) is a way of encoding human knowledge into an automated systems. RBS represent knowledge as a set of rules and provides users with analysis and recommendation based on knowledge base. Rules are represented as a set of IF-THEN statements [4].

More problems were handled and solved by the recommendation system. Cold-start is one of the most critical challenges. The lack of information in the system is a factor in the cold-start issue. This problem does not constantly occur. This problem arises if the system must promote new users or new goods to existing users without enough information. Finding and grouping comparable people and items into clusters might help solve the cold start issue. In every single group, the top-N suggestions are made for new users. Traditional clustering techniques, such as k-means, spectral clustering, and min-max cut, take just one feature into consideration [19]. Bi-Clustering is clustering technique of co-grouping two types of entities simultaneously. Bi-clustering can also be called as co-clustering or two-mode clustering or block clustering [16]. The bi-clustering approach and sequential rules are utilized in this research for addressing the issue of cold start.

The rest of the section of the work is organized as follows: Section 2 is a brief overview of the literature on sequential pattern mining and recommendation systems. Section 3 deals with the proposed research work and the suggested technique. Section 4 presented implementation and results. The conclusion and further work may be found in Section 5.

RELATED WORK

In the early days, Sequential pattern mining uses Apriori algorithms to find intra-transactionalassociations and to generate rules for associations [5]. The apriori style sequential pattern mining generates too many candidate sequence sets. To check whether the candidate sets meet the minimum threshold, the database must be scanned multiple times which results in expensive costs for real-time application. As a result, the projected based databases are developed to reduce the cost of creating andidentifying the candidates [14]. The PrefixSpan is a prefix projected sequential mining algorithm that employs the divide and conquer strategy for projecting frequent prefixes. It avoids large candidate generation thus improvising the execution time and memory storage [3].

The recommender system has grown as an individual research field from the mid-1990s. From the enormous undiscovered dataset, recommendation systems are used to produce customized recommendations [8]. Collaborative filtering system personalizes item recommendations for product users based on their previous history [10]. A content-based system evaluates the attributes of the product and makes recommendations. Knowledge-based systems make product recommendations based on user interests and domain knowledge about people and items [9]. The recommendation system is used to reduce the information overhead problems in variety of industries, including online shopping, e- learning, movies, books, news, and research publication [13]. The recommendation process is used to predict and model web navigation behaviors, as well as to identify the top frequently visited websites [11]. Anwar [12] proposed a cross-domain recommendation system combining Topseq rule mining for finding common rules and sequential pattern mining methods to locate patterns. The most frequent issues the recommendation system encounters include scaling, limited data, over-specialization, and cold-start issues [27,28,29].

The cold start issue has been approached from a variety of angles. By posing a series of queries to users, some study seeks to prevent the cold-start issue. Some employ hybrid methods that combine several data sources, like user profiles and tags, to determine how similar two objects are [18]. The item cold start problem in sequential recommendation is addressed by a cold start recommendation system based on meta learning [17,30]. From the item rating patterns, Matrix Factorization describes the items and consumers. Cold start problems can be solved using deep learning techniques, however this involves analyzing very huge amounts of data [25,26]. The cold start problem is resolved by combining association rule mining and the k-means clustering technique. The absence of duality between samples and features is the primary flaw in the conventional clustering approach [31].

PROPOSED METHOD

In this work, the cold start issue is overcome by adopting a bi-clustering strategy. The foundation of the recommendation system is the analysis of each customer's purchasing history, which is used to produce common sequences and patterns using the Prefix-span algorithm. Rules are framed using frequent sequences and their support. Rules facilitate appropriate product recommendations. Figure 1 depicts the proposed work's flowchart.

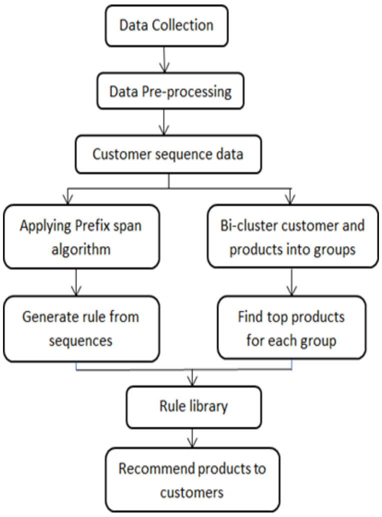

Data Collection
Data Pre-processing
Customer sequence data
Applying Prefix span algorithm
Bi-cluster customer and products into groups
Generate rule from sequences
Find top products for each group
Rule library
Recommend products to customers

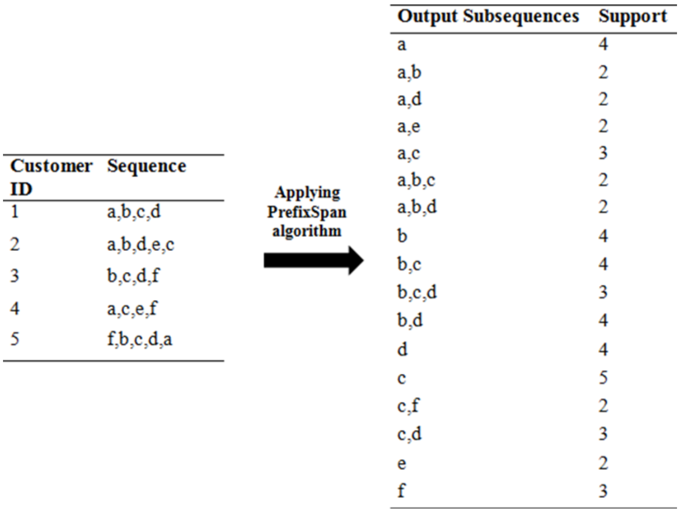
Figure 1: The proposed work's flowchart

1.1 Data Cleaning

Getting the relevant information from an online retailer is the first step. They provide information about every customer and the things they purchased during a given time frame. The pre-processing stage of data standardizes the data's format and verifies its accuracy. If a tuple has any empty values or data of different types, it is considered to be invalid. This procedure turns the incorrect data into a standard format, simplifying the mining process. Each customer's purchase sequences are taken from the transactional database once the data has been pre-processed. The Prefix Span technique is then used to mine the customer's sequential data.

1.2 Prefix Span Algorithm

[2] Introduced the PrefixSpan algorithm for mining sequential patterns. PrefixSpan algorithm recursively projects the sequential database based on prefixes and produces the sequential patterns based on exploring the locally common sequences. The prefix span technique displays the common sub-sequences bigger than minimal support when combined with a minimum support criterion and a sequence database. Figure 2 displays the outcomes of the prefix span technique for elementary sequences.



Customer ID	Sequence
1	a,b,c,d
2	a,b,d,e,c
3	b,c,d,f
4	a,c,e,f
5	f,b,c,d,a

Applying PrefixSpan algorithm

Output Subsequences	Support
a	4
a,b	2
a,d	2
a,e	2
a,c	3
a,b,c	2
a,b,d	2
b	4
b,c	4
b,c,d	3
b,d	4
d	4
c	5
c,f	2
c,d	3
e	2
f	3

Figure 2: Results of PrefixSpan algorithm

Prefix Span Algorithm:

a) Sequence database and minimal support threshold are the inputs.

b) A list of frequently recurring sequential patterns and the evidence backing them are the outputs.

Step 1: Locate sequences of length 1. Search the sequence database for all length-1 sequences that go beyond the required level of threshold.

Step 2: Split the search space. The frequent set of sequences are categorized into subsequences based on the number of prefixes.

Step 3: The search for sequential pattern subsequences. By mining each sequence iteratively in projected databases, the subsequences of the sequential patterns are created.

1.3 Bi-Clustering

In data mining, the phrase "Bi-clustering" refers to the simultaneous clustering of a matrix's rows and columns. A group of rows that behave similarly across a subset of columns, or vice versa, are called bi-clusters by the bi-clustering algorithm. A bi-clustering method that seeks out a constant bi-cluster rearranges the matrix's rows and columns to cluster bi-clusters with comparable values by combining similar rows and columns. Based on the resemblance between the client and the product, the customer purchase data is grouped into various categories. With the amount of support for each group, a list of users and frequently purchased things will be provided. The top-N products from the best support are what are retrieved.

Bi-Clustering Algorithm:

Input: Data Matrix A
Output: Bi-clusters containing similar behavior among rows and columns

Step 1: Normalize the matrix A with row and column's diagonal matrix R and C

$$A_n = R^{-1/2} \times AC^{-1/2}$$

Step 2: Perform singular value decomposition to partition the rows and columns of A

$$A_n = U \sum V^T$$

Step 3: The best singular vectors are selected to project the data.

Step 4: Create a matrix Z using the formula

$$Z = \begin{bmatrix} R^{-1/2} & U \\ C^{-1/2} & V \end{bmatrix}$$

Step 5: Create a cluster using the k-means algorithm from each row of the matrix.

Step 6: The first n_rows provides row partitioning and the first n_columns provides the column partitioning.

1.4 Tackling cold start problem

When the system lacks enough data to promote new users and items, a cold start problem occurs. Bi-clustering involves grouping both the users and the things they have purchased. We believe that the individuals in the group have comparable tastes. Upon joining the system, a new user will be allocated to one of the groups based on their easily available traits and get the group's suggestions or recommendation. If there are no features available about the new user, he will be recommended with top-N products and if the user find interest in any of the top-N products, then he will be put into the product group and recommends other similar product from that group. When a new product enters the system, its similar products are identified using cosine similarity and the new product is recommendedin the similar product's group.

1.5 Recommendation

Recommendation strategy plays a major role in e-commerce system. It can influence the reaction time and recommendation time directly. Rules are framed from the sequential pattern which forms the basis for recommendation. When a customer buys a product, the server analyses the customer purchase history and run product recommendation engine to find the top n products to be recommended with best support count. If the product sequences have fewer number of products, the next similar one with second highest score is recommended. If the purchase sequence data is not enough, the system recommends top-N products to the customers. The flow of work for the proposed recommendation algorithm is shown in Figure 3.

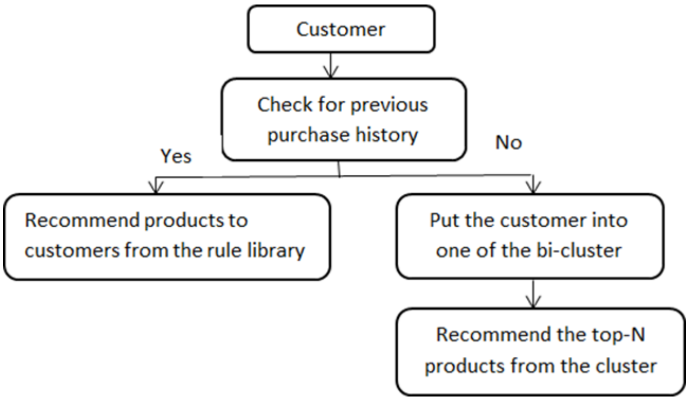

Customer
Check for previous purchase history
Yes
No
Recommend products to customers from the rule library
Put the customer into one of the bi-cluster
Recommend the top-N products from the cluster

Figure 3: Flow of work for the proposed Recommendation system

EXPERIMENTAL EVALUATION

1.5 Data Set

The data set was gathered from an online retailer that sold 541909 products to 4373 different consumers. Invoice No, Stock Code, Description, Quantity, Invoice Date, Unit Price, Customer ID, and Country are the eight domains for each record. Product code and description are referred to as stock codes and descriptions. Each client purchase sequence is created using the client ID to identify the common sequences.

1.6 Evaluation metrics

The training set and the test set are two separate sets of data that are separated from one another. Only the training set is used by the algorithm to build rules for product recommendations. The test set is employed to gauge how well the recommendation system is working. Recall and precision, which are the most widely used assessment metrics in information retrieval systems, are used to calculate the performance of the recommendation system, followed by F1 measure and accuracy.

Precision is defined as the ratio of the number of pertinent things chosen to the total number of items chosen. The proportion of pertinent things chosen to the total number of retrieved items is known as recall. The single metric known as the F1 measure is the harmonic mean and weighted average of recall and accuracy. The percentage of accurate forecasts to all predictions is known as accuracy.

1.7 Results

Table 1 shows the precision, recall, F1 measure, and accuracy performance results for various groups of customers. The optimal and unfavorable F1 measures are 1, respectively. High precision levels indicate that only pertinent things should be suggested. Not all of the goods are advised since recall values are in the middle. The recommendation system's average accuracy while utilizing the bi-clustering approach is 0.93. When employing user-based clustering to solve a cold start problem, the current system's average accuracy is 0.90. Compared to user-based clustering recommendation system, the suggested approach is more accurate. The performance comparison between user-based clustering and bi-clustering is shown in Figure 4.

Table 1: Performance results for various groups of customers

	Precision	**Recall**	**F1 measure**	**Accuracy**
Group1	0.899	0.571	0.695	0.923
Group 2	0.894	0.684	0.747	0.917
Group 3	0.923	0.715	0.738	0.908
Group 4	0.869	0.665	0.743	0.921

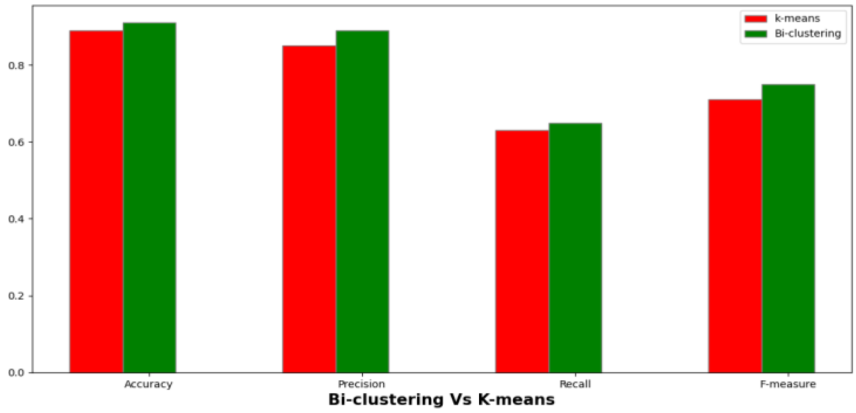


Figure 4: Performance Comparison of Bi-clustering and User based clustering

Instead of forecasting a single item, the suggested Bi-clustering based recommendation system gives consumers a list of the Top-N recommended things. The average prediction performance of the suggested system has proven to be good through trials. Consider 50 consumers at random and compare their actual sequence of purchased things with the anticipated items to gauge how well the algorithm predicts future purchases.

Analysis of the result shown in Figure 5 proves the following:

(1) More than 60% of the 50 items have the predicting accuracy of 0.9.

(2) The similarity between actual sequence and predicted sequence is 0.927.

(3) The predicting accuracy of rating with our method reflects well.

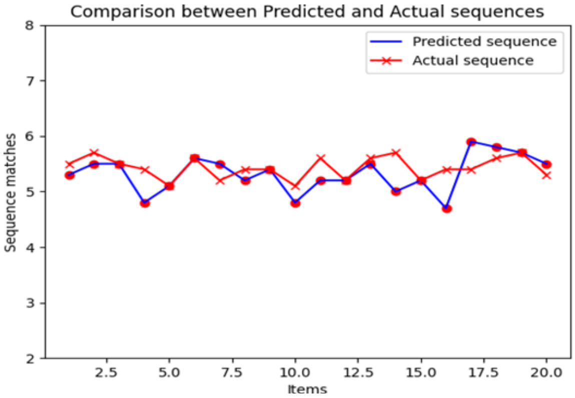


Figure 5: Comparison between predicted and actual sequence

CONCLUSION

An essential element of the recommendation system is played by sequential pattern mining. In this study, bi-clustering is employed to tackle the cold start problem while sequential pattern mining is used to identify common sequences from consumer transaction data. Prefix Span algorithm recursively projects the database by examining the locally frequent sequences when given a sequence database and little support. The common sequences that serve as the foundation for recommendations are used to frame the rules. Utilizing bi-clustering, the problem of product and user cold start is resolved by grouping customers and goods into distinct groups. Precision, recall, F1 score, and accuracy metrics are used to gauge the system's effectiveness. Future research may take into account unfavorable sequential patterns for recommendations.

REFERENCES

[1] Agrawal, R. and Srikant, R., 1995. Mining sequential patterns. In Proceedings of the eleventh international conference on data engineering (pp. 3-14). IEEE.

[2] Pei, J., Han, J., Mortazavi-Asl, B., Wang, J., Pinto, H., Chen, Q., Dayal, U. and Hsu, M.C., 2004. Mining sequential patterns by pattern-growth: The prefixspan approach. IEEE Transactions on knowledge and data engineering, 16(11), pp.1424-1440.

[3] Saraf, P., Sedamkar, R. and Rathi, S., 2015. Prefixspan algorithm for finding sequential pattern with various constraints. International Journal of Applied Information Systems (IJAIS), pp.2249-0868.

[4] Grosan, C. and Abraham, A., 2011. Rule-based expert systems. Intelligent systems. Springer, Berlin, Heidelberg pp. 149- 185.

[5] Mooney, C.H. and Roddick, J.F., 2013. Sequential pattern mining--approaches and algorithms. ACM Computing Surveys (CSUR), 45(2), pp.1-39.

[6] Trivonanda, R., Mahendra, R., Budi, I. and Hidayat, R.A., 2020. Sequential Pattern Mining for e-Commerce Recommender System. In 2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS) (pp. 393-398).

[7] Lopes, P. and Roy, B., 2014. Recommendation System using Web Usage Mining for users of E- commerce site. Int. J. Eng. Res. Technol, 3, pp.1714-1720.

[8] Suganeshwari, G. and Ibrahim, S.S., 2020. Rule-based effective collaborative recommendation using unfavorable preference. IEEE Access, 8, pp.128116-128123.

[9] Anwar, T. and Uma, V., 2020, Book Recommendation for eLearning Using Collaborative Filtering and Sequential Pattern Mining. In 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI) (pp. 1-6). IEEE.

[10] Yap, G.E., Li, X.L. and Yu, P.S., 2012, Effective next-items recommendation via personalized sequential pattern mining. In International conference on database systems for advanced applications Springer, Berlin, Heidelberg, pp. 48-64.

[11] Khorgade, S., Sambhare, P.,2017, Web Recommendation System Based on Approach of Mining Frequent Sequential Patterns, International Journal of Latest Engineering Research and Applications pp. 25-31

[12] Anwar, T. and Uma, V., 2019. CD-SPM: cross-domain book recommendation using sequential pattern mining and rule mining. Journal of King Saud University-Computer and Information Sciences. pp. 793-800

[13] Hameed, M.A., Al Jadaan, O. and Ramachandram, S., 2012. Collaborative filtering based recommendation system: A survey. International Journal on Computer Science and Engineering, 4(5), p.859.

[14] Fournier-Viger, P., Lin, J.C.W., Kiran, R.U., Koh, Y.S. and Thomas, R., 2017. A survey of sequential pattern mining. Data Science and Pattern Recognition, 1(1), pp.54-77.

[15] Zaki, M.J., 2001. SPADE: An efficient algorithm for mining frequent sequences. Springer Nature Link, Machine learning, 2001 Kluwer Academic Publishers, 42(1), pp.31-60.

[16] Anagnostopoulos, A., Dasgupta, A. and Kumar, R., 2008, Approximation algorithms for co- clustering. In Proceedings of the twenty-seventh ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, pp:201-210.

[17] Zheng, Y., Liu, S., Li, Z. and Wu, S., 2020. Cold-start sequential recommendation via meta learner. The Thirty-Fifth AAAI Conference on Artificial Intelligence,pp.4706-4713

[18] Son, L.H., 2016. Dealing with the new user cold-start problem in recommender systems: A comparative review. Information Systems,58,pp: 87-104.

[19] Yanxiang, L., Deke, G., Fei, C. and Honghui, C., 2013. User-based clustering with top-n recommendation on cold-start problem. In 2013 Third International Conference On Intelligent System Design And Engineering Applications, pp:1585-1589.

[20] Kang, J.S., Baek, J.W., Chung, K. 2020, PrefixSpan based Pattern Mining using Time Sliding Weight from Streaming Data,IEEE Access, 8,pp.124833-124844.

[21] Lien, Y.C., Wu, W.J., Lu, Y.L. 2020, How well do Teachers Predict Students' actions in Solving an Ill-defined Problem in STEM Education: A Solution using Sequential Pattern Mining. IEEE Access ,8,pp.134976-134986.

[22] Matloob, I., Khan, S.A., Rahman ,H.U. 2020, Sequence Mining and Prediction-Based Healthcare Fraud Detection Methodology. IEEE Access,8,pp.143256-143273.

[23] Guyet, T., Besnard, P. 2019, Semantics of Negative Sequential Patterns. European Conference on Artificial Intelligence,pp 1009-1015.

[24] Kim, B., Yi, G. 2019, Location-based Parallel Sequential Pattern Mining Algorithm,IEEE Access, 7,pp.128651-128658.

[25] Heidari, N., Moradi, P., & Koochari, A., 2022, An attention-based deep learning method for solving the cold-start and sparsity issues of recommender systems. Elsevier -Knowledge-Based Systems, 256, pp.109835 1-13.

[26] Wei, J., He, J., Chen, K., Zhou, Y., & Tang, Z., 2017, Collaborative filtering and deep learning based recommendation system for cold start items. Science direct-Expert Systems with Applications, 69, pp.29-39.

[27] Sejwal, V. K., & Abulaish, M., 2022, A hybrid recommendation technique using topic embedding for rating prediction and to handle cold-start problem. Science direct-Expert Systems with Applications, 209, p.118307 - 1-12.

[28] Briand, L., Salha-Galvan, G., Bendada, W., Morlon, M., & Tran, V. A., 2021, A Semi- Personalized System for User Cold Start Recommendation on Music Streaming Apps. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, pp. 2601-2609.

[29] Tey, F. J., Wu, T. Y., Lin, C. L., & Chen, J. L., 2021, Accuracy improvements for cold- start recommendation problem using indirect relations in social networks. Journal of Big Data, 8(1),pp.1-18.

[30] Huang, X., Sang, J., Yu, J., & Xu, C., 2022, Learning to learn a cold-start sequential recommender, 5th International Conference on Data Science and Information Technology

[31] Renjie, L., Shiping,W., & Wenzhong,G., 2019, "An Overview of Co-Clustering via Matrix Factorization," in IEEE Access,7,pp.33481-33493.