International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7177
***Article Received**: 24 April 2023 **Revised:** 08 June 2023 **Accepted:** 26 June 2023*

A Deep Learning Approach towards Cold Start Problem in Movie Recommendation System

Jinal Panchal[1], Dr. Sandeep Vanjale[2]
[1]M.Tech, Computer Engineering Department,
Bharati Vidyapeeth (Deemed to be University) College of Engineering
Pune, India
panchaljins@gmail.com
[2]HoD and Professor, Computer Engineering Department
Bharati Vidyapeeth (Deemed to be University) College of Engineering
Pune, India
sbvanjale@bvucoep.edu.in

Abstract— Recommendation systems play an important role for e-commerce websites to make profits. It has a variety of applications in different domains. There are three types of categories in which recommendation systems are classified i.e. content based, collaborative and hybrid systems. These systems suffer when a redundant amount of information is not available to provide recommendations. This problem is known as the cold start problem. In this digital era, it is possible to collect meta information about a user and provide rich recommendations. Various approaches such as social media analysis, graph networks have been proposed to solve this problem. But they lack personalization and generate irrelevant recommendations affecting the system performance. The objective of this work is to resolve new user cold start problem in movie recommendation systems using a deep learning approach that utilizes demographic attributes to cluster similar users. This embedding is given to the deep neural network to generate the recommendations. From the analysis done, we verify the effectiveness of our approach..

Keywords- Recommendation Systems, Cold start problem, Deep Learning, Artificial Neural Networks and Demographic attributes.

I. INTRODUCTION

In this competitive era, it becomes essential for any business to attract the user community to increase their sales that can lead to attractive profits. To increase engagement of users, recommendation systems play an important role. It helps customers to find relevant products among thousands of products available online. Moreover based on the user's past preferences these intelligent systems are able to provide personalized recommendations as well. Most popular applications such as YouTube, Netflix, Spotify, Amazon make use of recommendation systems to attract the users.

The recommendation systems are classified into three main categories: (1) Content based filtering: It recommends the new items to the user based on his/her history. (2) Collaborative filtering: It finds similar users to provide recommendations. (3) Hybrid systems: It combines both types of systems to overcome the limitations.

These systems suffer from Data scarcity, Scalability, Grey Sheep and Shilling attack and Cold Start problem. Among all these problems, the cold start problem is a long-standing problem. There are two kinds of cold start problems. One is a user cold start problem in which no information is available about the new user. Second is the item cold start problem in which there are no target users to recommend new items. The item cold start problem is comparatively easier to solve than the new user cold start problem.

With the advancement of machine learning technologies for different domains, it has been also applied to provide accurate recommendations. The self-driven and feature extraction ability of these models have attracted the research community. Various machine learning approaches have been proposed to solve the cold start problem, but they lack personalization. To alleviate this, we first cluster the similar using demographic attributes. This cluster embedding is fed into the deep neural network along with the user and item embeddings to generate the recommendations. To prove the effectiveness of our approach, we have compared it with the existing demographic approach.

This article is organized as follows: Section 2 reviews the existing literature. In Section 3 Proposed Framework is explained in detail. Experiment details are presented in Section 4. The result and analysis are described in Section 5 followed by the Conclusion.

II. RELATED WORK

In this section, we review various approaches employed to solve cold start problem in recommendation systems.

Gupta et al. [1] proposed a K-means clustering algorithm to solve the cold start problem for recommendation systems. They have utilized the demographic attributes such as Age, Gender to cluster the similar users and then using MySQL recommendations are generated for a new user. Xing et al. [2] modified traditional Singular Value Decomposition (SVD) method by fusing attribute information of items with the past rating data. The proposed framework resolves the new item recommendation problem. Sridevi [3] proposed a DECORS framework for movie recommendation system that considers demographic attributes of users to make partitions initially. In addition to that, K-means clustering is performed to cluster similar users based on movie ratings. This framework optimizes the computational complexity for cold start problem. Zamanzadeh et al. [4] proposed a graph based GHES model that alleviates the cold start problem by finding similar users utilizing graph networks in addition to the demographic and location attributes. This framework outperformed the other baseline models for MovieLens dataset.

Another way to generate recommendations is to explicitly ask users and collect data for a new user. Active learning approaches are Personalized and Non-personalized. Non-personalized approaches are simple to implement and require no feedback from the user. These approaches provide the same list of items to the user. To alleviate the need for extra information, a probabilistic ZeroMat [5] model is proposed that requires no input from the user. But it assumes that the user item ranking is distributed according to the Zipf distribution. To remove this assumption Wang et al. [6] proposed a DotMat framework that adopts the idea of improved matrix factorization RankMat framework [7] and optimized the loss function. In addition to that, RankMat considers various kinds of probability distribution functions. DotMat outperforms ZeroMat and other heuristic approaches.

Alabdulrahman et al. [8] proposed the active learning-based PUPP (Popular User Personalized Prediction) framework to recommend items to new users. They have employed K-nearest Neighbour (KNN) to initially assign a group to a new user and then popular users are identified to generate the recommendation. Zhu et al. [9] combined an active learning approach and item's attribute information to solve the new item cold start problem. Specifically, a new framework is created to select the users based on the user's history and item's attributes.

Relying on user input to gather information about an user or item leads to privacy concerns. To alleviate that Wahab et al. [10] have proposed trust based Double Deep-Q learning algorithm to solve item cold start problem for MovieLens dataset. Moreover, Federated learning is utilized to address privacy concerns. Ifada et al. [11] proposed a popularity-based hybrid model that considers item as well as user popularity to provide accurate recommendations.

Tey et al. [12] proposed a social network-based approach to improve the recommendation accuracy. They utilized the “Friend” and “Friend of Friend” relationship in order to provide recommendations to new users. The experimental results show that their proposed method outperformed traditional methods on YELP dataset [13]. This approach is non-personalized and requires further modifications. Zhang et al. [14] proposed a social media-based approach that considers publicly available tweets to improve upon the existing models. An ensemble model is proposed that combines Generalized Matrix Factorization (GMF) and Multi-Layer Perceptron (MLP) model to make predictions. Natarajan et al. [15] proposed RS-LOD (Recommendation System with Linked Open Data) that uses a public interface cloud to solve cold start problem in collaborative filtering systems. To identify similar users Pearson's Correlation Coefficient is used.

From the literature, we have found that all the proposed work lacks the personalization attribute while providing recommendations. To alleviate that we have considered a personalized deep learning paradigm that utilizes demographic attributes.

III. PROPOSED FRAMEWORK

In this section, we describe our proposed framework. It has two modules: Clustering module and Recommendation module. In subsequent sections, each module is discussed in detail.

A. Clustering Module

This module aims to cluster similar users based on the demographic attributes. We have used fanny clustering to partition users. In this we have considered the number of clusters to be 4. The reason behind choosing fanny clustering is that it can assign more than one cluster to a data point. The fanny clustering tries to minimize the objective function defined as:

$$\sum_{v=1}^{k} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} u_{iv}^{r} u_{jv}^{r} d(i,j)}{2\sum_{j=1}^{n} u_{jv}^{r}} \quad (1)$$

Where k,n, and r represent the number of clusters, data points and membership coefficient respectively. u(i,v) denotes the membership of observation i to the cluster v and d(i,j) represents the dissimilarity between the observations. The clusters created are embedded in the next module.

B. Recommendation Module

The objective of this module is to predict the ranking for an item. We have employed an Artificial Neural Network to generate the recommendations. We have proposed a deep ANN

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7177
***Article Received**: 24 April 2023 **Revised:** 08 June 2023 **Accepted:** 26 June 2023*

model that consists of two fully connected hidden layers. The architecture of the proposed network is presented in Figure 1.

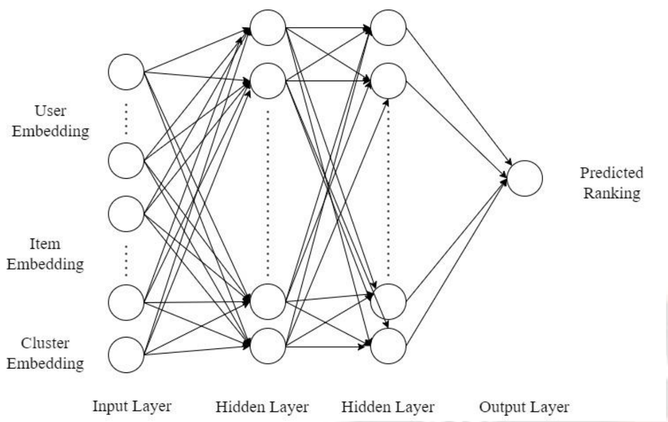

Figure. 1. Proposed Modified ANN Architecture.

Table 1. Summarizes the parameters considered for the proposed model.

Table 1. Experimental parameters

Parameter	Value
No. of clusters	4
User Embedding	5
Item Embedding	5
Cluster Embedding	2
Hidden layer sizes	128,32
Optimizer	Adam

We have utilized python language to develop our model. To implement ANN module deep learning library Tensorflow is used.

The proposed methodology is explained in Algorithm 1.

Algorithm 1: Recommendation system for new user cold start problem

1. Cluster users using demographic attributes.
2. Initialize User Embedding.
3. Initialize Item Embedding.
4. Initialize Cluster Embedding.
5. Train ANN model.
6. Predict the ranking.

We have utilized python language to develop our model. To implement ANN module deep learning library Tensorflow is used.

IV. EXPERIMENTAL DETAILS

A. Dataset

We have experimented with the MovieLens 100K dataset. It is available from the website (http://movielens.org). It contains 1,00,000 ratings (1-5) from 943 users for 1682 movies. Each user has rated at least 20 movies. It contains the following Detailed Description of Data files:

- **u.data**: - The full data set, 100000 ratings by 943 users on 1682 items. Each user has rated at least 20 movies.
- **u.item** :- It contains the Information about the items (movies); this is a tab separated list of movie id | movie title | release date | video release date | IMDb URL | unknown | Action | Adventure | Animation | Children's | Comedy | Crime | Documentary | Drama | Fantasy | Film-Noir | Horror | Musical | Mystery | Romance | Sci-Fi | Thriller | War | Western |
- **u.user**: - It has demographic information about the users; this is a tab separated list of user id | age | gender | occupation | zip code.

B. Evaluation Parameters

We have used the following evaluation parameters.

- **Silhouette Factor**: - It is used to verify the effectiveness of the clustering algorithm. It ranges from [-1,1]. Higher value is preferred.
- **MAE (Mean Absolute Error):** - It is absolute mean error between actual and predicted ranking of users. It is calculated as:

$$MAE = \frac{1}{N}\sum_i \left|a_{i,r} - p_{i,r}\right| \qquad (2)$$

- **RMSE (Root Mean Square Error): -** It is calculated as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_i \left|a_{i,r} - p_{i,r}\right|^2} \qquad (3)$$

- **MSE (Mean Square Error):** - It is calculated as follows:

$$MSE = \frac{1}{N}\sum_i \left|a_{i,r} - p_{i,r}\right|^2 \qquad (4)$$

Where $a_{i,r}$ and $p_{i,r}$ represents the actual and predicted ranking for item i and N represents the total number of items. The lower value of error metrics is preferred.

V. RESULT AND ANALYSIS

In this section, we represent the results obtained for our proposed approach. To prove the efficiency of fanny clustering over k-means clustering, we have performed analysis on silhouette factor. Table 2. Summarizes the results obtained for two different clustering mechanisms.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7177
***Article Received**: 24 April 2023 **Revised:** 08 June 2023 **Accepted:** 26 June 2023*

Table 2. Analysis on Silhouette factor

Evaluation Parameter		K-means	Fanny
Silhouette	Cluster 1	-0.6940	-0.0940
	Cluster 2	-0.3054	-0.0563
	Cluster 3	-0.7232	0.2941
	Cluster 4	-0.6182	-0.1130

From the Table, we can observe that the fanny clustering has outperformed the k-means clustering approach.

We have compared our method with existing User based, Item based collaborative filtering, SVD and non-personalized popular item recommendation approaches. Table 3. Summarizes the results.

Table 3. Analysis on Error Metrics

Method	RMSE	MSE	MAE
UBCF	1.1719	1.3735	0.9160
IBCF	1.0917	1.1917	0.7534
SVD	1.0571	1.1176	0.8222
POPULAR	1.0483	1.0991	0.8186
Proposed approach	**0.7470**	**0.8643**	**0.7267**

From the above Table, we can observe that our approach outperforms the other existing approaches. Figure 2. Compares the performance of baseline models and proposed approach for evaluation parameters.

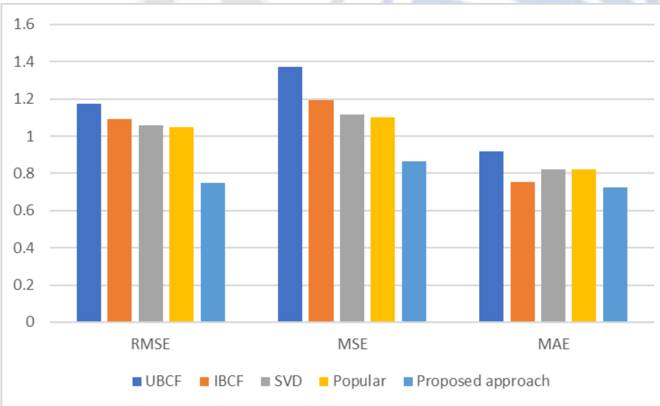

Figure. 2. Performance analysis.

VI. CONCLUSION

Recommendation systems play an important role in boosting performance of e-commerce websites. These systems suffer from cold start problem, that arise when there is no available information for a new user. Various social media analysis, machine learning approaches have been proposed although they lack personalization aspects. To overcome this, we have proposed a deep learning approach that utilizes demographic attributes to build a user profile. We have modified ANN that considers clusters created from the demographic attributes and generates personalized recommendations for the movie domain. Our proposed approach has shown its robustness on various evaluation parameters considered.

REFERENCES

[1] Gupta, Sugandha and Shefali Arora. “Handling Cold Start Problem in Recommender Systems by Clustering Demographic Attribute.” (2016).J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Guo, Xing & Yin, Shi-chao & Zhang, Yi-wen & Li, Wei & He, Qiang. (2019). Cold Start Recommendation Based on Attribute-Fused Singular Value Decomposition. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2891544.

[3] Sridevi, M.. “DECORS : A Simple and Efficient Demographic Collaborative Recommender System for Movie Recommendation.” (2017).

[4] Aziz Mutlaq, K. A. ., Kareem, H. R. ., & Abbass, G. Y. . (2023). Modeling and Prediction LTE 4G NW according to Memory Algorithm of Long-Short Term . International Journal of Intelligent Systems and Applications in Engineering, 11(1s), 172–178. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2490

[5] Zahra Zamanzadeh Darban, Mohammad Hadi Valipour, GHRS: Graph-based hybrid recommendation system with application to movie recommendation, Expert Systems with Applications, Volume 200, 2022, 116850, ISSN 0957-4174.

[6] H. Wang, “ZeroMat: Solving Cold-start Problem of Recommender System with No Input Data”, IEEE 4th International Conference on Information Systems and Computer Aided Education, 2021.

[7] Makarand L, M. . (2021). Earlier Detection of Gastric Cancer Using Augmented Deep Learning Techniques in Big Data with Medical Iot (Miot). Research Journal of Computer Systems and Engineering, 2(2), 22:26. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/28

[8] Wang, Hao. (2022). DotMat: Solving Cold-start Problem and Alleviating Sparsity Problem for Recommender Systems. 10.48550/arXiv.2206.00151.

[9] H. Wang, “RankMat: Matrix Factorization with Calibrated Distributed Embedding and Fairness Enhancement”, ICCIP, 2021.

[10] Alabdulrahman, Rabaa & Viktor, Herna & Paquet, Eric. (2019). Active Learning and User Segmentation for the Cold-start Problem in Recommendation Systems. 113-123. 10.5220/0008162901130123.

[11] Zhu, Y., Lin, J., He, S., Wang, B., Guan, Z., Liu, H., & Cai, D. (2018). Addressing the Item Cold-Start Problem by Attribute-Driven Active Learning. IEEE Transactions on Knowledge and Data Engineering, 32, 631-644.

[12] Alexei Ivanov, Machine Learning for Traffic Prediction and Optimization in Smart Cities , Machine Learning Applications Conference Proceedings, Vol 3 2023.

[13] Andrew Hernandez, Stephen Wright, Yosef Ben-David, Rodrigo Costa, David Botha. Predictive Analytics for

Decision-Making: Leveraging Machine Learning Techniques. Kuwait Journal of Machine Learning, 2(3). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/193

[14] Omar Abdel Wahab, Gaith Rjoub, Jamal Bentahar, Robin Cohen, Federated against the cold: A trust-based federated learning approach to counter the cold start problem in recommendation systems, Information Sciences, Volume 601, 2022, Pages 189-206, ISSN 0020-0255.

[15] Noor Ifada, Ummamah, and Mochammad Kautsar. 2021. Hybrid popularity model for solving cold-start problem in recommendation system. In Proceedings of the 5th International Conference on Sustainable Information Engineering and Technology (SIET '20). Association for Computing Machinery, New York, NY, USA, 40–44.

[16] Tey, F.J., Wu, TY., Lin, CL. et al. Accuracy improvements for cold-start recommendation problem using indirect relations in social networks. J Big Data 8, 98 (2021).

[17] YELP Dataset : https://www.yelp.com/dataset. Last accessed : January, 2023.

[18] Zhang, Yihong & Maekawa, Takuya & Hara, Takahiro. (2021). Using Social Media Background to Improve Cold-start Recommendation Deep Models.

[19] Senthilselvan Natarajan, Subramaniyaswamy Vairavasundaram, Sivaramakrishnan Natarajan, Amir H. Gandomi, Resolving data sparsity and cold start problem in collaborative filtering recommender system using Linked Open Data, Expert Systems with Applications, Volume 149, 2020, 113248, ISSN 0957-4174.