International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 7 | ISSN : 2456-3307

Optimization of Cold Start Problem in Recommendation Systems : A Review

Manali Patel*[1], Pratik A. Barot[2]

*[1]PG student, Department of Computer Engineering, Government Engineering College Gandhinagar, Gandhinagar, Gujarat, India

[2]Assistant Professor, Department of Computer Engineering, Government Engineering College Gandhinagar, Gandhinagar, Gujarat, India

ABSTRACT

A major issue in collaborative filtering based recommendation systems is providing recommendations for a new user or to find a target user for a new item. This is referred to as a cold start problem in recommendation systems. This cold start problem is caused by lack of information about the said entity and it is very important issue to deal with. Many solutions have been suggested for a cold start problem. These solutions have not focused on the issue of personalization. Using one's demographic attributes (age, gender, occupation) more accurate and interesting recommendations can be provided to the users. It makes recommender systems more personalized and leads to user satisfaction that is very important aspect for any e-commerce websites. Personality information can address cold start problem. We did literature survey and from our finding active learning can overcome cold start problem effectively.

Keywords : Recommendation Systems, Cold Start, Demographic Attributes, Active Learning.

I. INTRODUCTION

We are living in an internet era, where there is a load of information available. It is important to find or extract information which is useful. Data mining is a process of discovering patterns in large datasets. In today's era it is all about the competition. There are so many companies and firms available to sell their product. Marketing becomes important factor to maximize the profit. User satisfaction also plays crucial part and user has to be convinced to buy products. Here recommendation systems become important and crucial part. Recommender systems are tools which provide recommendations to the user based on his/her previous preferences, browsing history [1]. Many big companies like Amazon, Facebook use this to provide user better experience.

All recommendation systems use available information to make recommendations. These systems use filtering techniques. There are two widely used filtering techniques. (a) Content based filtering and (b) collaborative filtering. Content based filtering technique filters items through past experience of user, browsing history. Collaborative filtering techniques find similar users who share same interest to provide recommendations [2]. Both the techniques have their own advantages and disadvantages. One of the problems in content based filtering technique is the vast size of the items; each item has to be examined to provide recommendations [3]. There are hybrid filtering techniques which combine both content based filtering and collaborative filtering techniques. Collaborative filtering techniques are widely used methods to provide recommendations.

Collaborative filtering techniques cause following problems in recommendation systems:

(a) **Data sparsity:** This problem arises when ratings provided by user is sparse. It can lead to inefficient recommendations.

(b) **Scalability:** when there is a huge amount of data available scalability problem occurs.

(c) **Gray sheep:** This problem arises when a user does not belong to any other group.

(d) **Shilling attacks:** These are categorized in push attacks and nuke attacks. When rival party tries to increase ratings of their own items by means of unfair ways or they try to affect the ratings of other competitors.

(e) **Cold start:** when new user enters a system. There is no information available for that user and it becomes difficult to provide recommendations with no rating information available [4].

Among all these problems Cold start problem is a long standing problem in recommendation systems. There are two kinds of cold start problems one is new user cold start problem in which no information is available about the user. Second is new item cold start problem in which there are no target users to recommend new item [5]. The new item cold start problem has nominal effect on recommender systems. They can be handled by finding motivated users who are responsible for rating items.

Cold start problem arises due to lack of information about the new user. Many solutions are given to resolve new user and new item cold start problem in recommendation systems. Following table gives comparative study of various techniques used to solve cold start problem.

Table 1 : Comparative Study of Techniques Used To Resolve Cold Start Problem In Various Domains

Author	Domain	Techniques
Punam Bedi et al. [7].	Book Domain	IBSP Factor
Ivica Obadi´c et al. [8].	Business	SVD++
Zhenzhen Xu et al. [9].	Movies, Books	Trust and Usage
Szu-Yu Chou et al. [11].	Music	Deep learning methods
Anand Kishor Pandey et al. [14].	Movie	Demographic Approach
Zhenzhen Xu et al. [15].	Movie	Demographic approach and personality traits
Rasoul Karimi et al. [16].	Movie	Active Learning

II. ACTIVE LEARNING

Active learning is a semi-supervised learning. It interactively queries the user to get the desired output. It is very expensive to label the data it becomes easier with Active learning techniques. Real world examples of Active learning techniques are Natural Language Processing (NLP), deep learning, Twitter comment detection.

Active learning in recommendation systems is an explicit way of collecting information [5]. From given set of items active learning chooses relevant items to ask users to rate. Active learning strategies are of two kinds: personalized and non-personalized. In non-personalized users previous preferences is not considered and all users are provided same set of items to rate based on popularity or some other measures. In personalized active learning it asks different users to rate different set of items [6].

Following are various Active Learning strategies:

Non-Personalized strategies:

A) Single-Heuristic Strategies

- **Uncertainty Based**

  Variance considers rating's variance. It asks users to rate items with the highest variance.

  Entropy of ratings is computed by using the relative frequency of each of the five possible rating values.

  Entropy0 techniques solve the limitation of entropy method by considering the unknown rating as a new rating value equal to 0.

- **Error Based**

  Greedy Extend obtains item list whose ratings should be elicited from user, such that RMSE is minimized.

- **Attention Based**

  Popularity selects the items that have received the highest number of ratings.

  Coverage selects the items that are highly co-rated by the users.

B) Combined-Heuristic Strategies

- **Static**

  Log (pop)*Entropy strategy scores an item by computing the logarithm of the popularity multiplied by the entropy of the ratings given to the item.

  HELF stands for Harmonic mean of Entropy and Logarithm of Frequency. In this method Entropy and Popularity scores are combined to select items.

Personalized strategies:

A) Single-Heuristic Strategies

- **Acquisition Probability Based**

  Item-Item Item Strategy selects the item with the highest similarity between previously rated items. Hence, the similarity between previously rated item and unrated item is computed.

  Personality Based strategy transforms matrix into new rating matrix with same number of rows and columns and maps non null entries to 1 and null entries to 0. Then new matrix is used to train a Matrix Factorization algorithm.

- **Influence Based**

  Influence Based strategy estimates the influence of item's ratings on the rating prediction of other items and selects the items with the largest influence

  Impact Analysis .strategy selects the items that have highest impact on the prediction of other item's ratings.

- **Prediction Based**

  Aspect Model describes latent space model for each user.

  Min rating provides user with the items that have minimum predicted ratings.

- **User Partitioning**

  IGCN strategy constructs a decision tree where each leaf node represents a cluster of users and each internal node stores a test for a specific item that is proposed to a user to rate. Users are clustered according to their similarity values that are measured using a neighbourhood based collaborative filtering approach.

B) Combined-Heuristic Strategies

- **Static**

  Non-myopic combines two personalized strategies Min rating and Min norm.

  Voting strategy scores an item with the number of votes given by a committee of strategies. Each strategy produces its top candidate items for rating elicitation and then the items appearing more often in these lists are selected.

- **Adaptive**

  Switching strategy is applied, a certain percentage of the users (called exploration group) are randomly selected for choosing the best performing individual strategy, and this winning strategy is applied on the remaining users. Each individual strategy is tested to an equal number of random users in the exploration group: it selects items to be rated by these users, and acquires their ratings [6].

Active learning techniques generally make use of introductory survey to gather information about users. Sometimes survey becomes tiresome for users and leads to providing weak recommendations. Demographic approach and active learning techniques can be combined to enhance the accuracy.

III. DETAILED LITERATURE SURVEY

Many solutions have been proposed to resolve new user and new item cold start problem in different domains. One of the solutions to solve cold start problem in book domain is given by using interaction based social proximity factor which takes into account the likes and comments of the user's friend. The choices made by the friends have impact on user's choice. The proposed factor 'IBSP' is used to rank the user's friends on the basis of likes and comments made by them. To further sort these friends of the user, the IBSP calculation is made for user's response on the various status of the friend [7]. One of the drawbacks is that people do not make friends on social sites only on the basis of having same interests. This can lead to recommendations that do not interest new user.

Deep learning methods are used to overcome new item cold start problem. Here model based approach is used. To overcome this issue in YELP dataset they have used matrix factorization method SVD++ (Single Value Decomposition) to decompose user and business rating matrix into user feature vector and business feature vector. But we do not have any prior information about new business we cannot describe its latent features. To resolve this problem there is a textual description available for each business and deep convolution network is used to find latent features from given textual information. It maps business reviews given in text file to latent factor vectors. After having two feature vectors, they are concatenated and predicted ratings are obtained for new business [8].

The cold start issue in recommendation systems is addressed using trust and usage context. The knowledge learned from one domain could be transferred to other domains to improve recommendation efficiency. For an example if a user likes singer's songs, some movies in which the singer starred could be recommended to user. It utilizes trust relation and usage context to build association between different domains. They have applied a traditional trust-aware recommendation method to a cross-domain for predicting ratings of source domain's user on target domain's items [9]. One of the limitations is that it is not always the case that two domains are related to each other.

There are many approaches which deal with both new user and new item cold start problem, one approach is decoupled completion and transduction. It follows two steps: 1) the completion of a rating sub-matrix, which is generated by excluding cold start users and items because for matrix factorization it is not possible to fill out row or column this is entirely empty. 2) Transduction of knowledge from existing ratings to cold start items/users [10].

Deep learning method is used to resolve both new user and new item cold start problem in music recommendation system. Deep learning method is used which is a conditional preference net. This is compared with deep learning method D+IF which uses matrix factorization. To solve this problem an introductory survey is conducted for new imaginary user, in survey user has been given items to select. An items represented in survey are top rated. From survey preference vector for a new user is recorded on which neural network is conditioned. To address new item cold start problem CPN considers item features given in dataset. From the information gathered from preference vectors of new user new items are suggested [11].

Cold start problem present in tourism destination recommendation systems, is addressed using Opinion mining and hybrid collaborative filtering techniques. It is very common that more users are writing text reviews for tourism destinations. So the method proposed here integrates reviewer's ratings with text reviews to improve accuracy. This includes opinion mining then by mining text reviews user preference and tourism destination ratings are obtained [12].

Combination of association rule and clustering technique is used to solve problem. Association rules are used to create and expand user profile so it will contain more ratings. So using association rule new user cold start problem is tackled. User profile is expanded so that more ratings are available for that particular user. Clustering is used to group items based on similarity measures. This handles new item cold start problem in recommendation systems [13].

Most recently to make recommendations more accurate and personalized demographic attributes are used. Demographic attributes can be (age, occupation, gender, location). In the proposed method clusters are created based on demographic information. After clustering classification is performed to check accuracy. When new user enters the system he is assigned to a particular cluster using his/her demographic information. After that similarity between new user and old users is calculated and recommendations are provided [14]. Demographic attributes are considered hard skills and to provide accurate recommendations soft skills should also be considered [15].

Demographic attributes are considered as hard skills, it is not enough to provide recommendations. Because two people having same age and gender can have different preferences. Personality traits also should be taken into considerations. Personality trait is based on Big 5 model that includes Extroversion, Agreeableness, Conscientiousness, Emotional stability and openness [15].

Active learning is used for enhancing accuracy of system and to reduce waiting time that is the time user has to wait before being asked a new query. In this paper they have modified aspect model approach and Bayesian approach to reduce the waiting time [16].

Recently there has been a survey paper on cold start problem. It defines Implicit and Explicit methods to gather Information of a new user. In this survey paper Active learning methods are defined. [17]

IV. CONCLUSION

In this competition era recommender systems are very useful and crucial tools for e-commerce websites. There is a plethora of information is available over the internet and recommender systems extracts information which interests a users giving them personalized and satisfactory choices. Recommender systems use various techniques: (1) Content based filtering, (2) Collaborative filtering, (3) Hybrid techniques. Among this Collaborative filtering is widely used to provide recommendations. It suffers from various problems like sparsity, scalability, cold start, and Gray sheep. Cold start problem is one of them and crucial one. In this paper various solutions are discussed which have been proposed to resolve this problem.

Active Learning for recommendation systems is a new research area. In future work, Active learning techniques can be used to enhance accuracy and provide more personalized recommendations to a cold user.

V. REFERENCES

[1]. Charu c. Aggarwal "Recommender Systems: The textbook" Springer ch-1

[2]. Iateilang Ryngksai, L. Chameikho "Recommender systems: Types of filtering techniques" International journal of Engineering research and technology" vol. 3 issue 11 November 2014

[3]. Harry Zisopoulos, Savvas Karagiannidis, Georgios Demirtsoglou, Stefanos Antaris" Content-Based Recommendation systems" Researchgate November 2008

[4]. Gizaw, Zewengel Tilahun, Huang Dong Jun, Ammar Oad "Solving Cold-Start Problem by Combining Personality Traits and Demographic Attributes in a User Based Recommender System" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 7, Issue 5, May 2017

[5]. Jyotirmoy Gope, Sanjay Kumar Jain "A survey on solving cold start problem in recommender systems" IEEE 2017

[6]. Mehdi Elahi, Francesco Ricci, and Neil Rubens "Active learning in collaborative filtering recommender systems" Springer pp 113-124

[7]. Punam Bedi, Chhavi Sharma, Pooja Vashisth, Deepika Goel, Muskan Dhanda "Handling Cold Start Problem in Recommender Systems by using Interaction Based Social Proximity Factor" IEEE 28 September 2015

[8]. Ivica Obadi'c, Gjorgji Madjarov, Ivica Dimitrovski, and Dejan Gjorgjevikj ": Addressing Item-Cold Start Problem in Recommendation Systems Using Model Based Approach and Deep Learning" Springer,07 September 2017

[9]. Zhenzhen Xu, Fuli Zhang, Wei Wang, Haifeng Liu, Xiangjie Kong "Exploiting Trust and Usage Context for Cross-Domain Recommendation" IEEE, 11 May 2016

[10]. Iman Barjasteh, Rana Forsati, Farzan Masrour, Abdol-Hossein Esfahanian, Hayder Radha "Cold-Start Item and User Recommendation with Decoupled Completion and Transduction" ACM 16th September 2015

[11]. Szu-Yu Chou, Li-Chia Yang, Yi-Hsuan Yang, Jyh-Shing Roger Jang "Conditional Preference Nets for User and Item cold start problems in music recommendation" IEEE 31st August 2017

[12]. Xiaoyao Zheng, Yonglong Luo, Zhiyun Xu, Qingying Yu and Lin Lu "Tourism destination recommender system for the cold start problem" Researchgate 31st July 2016

[13]. Hridya Sobhanam, A.K.Mariappan ": A Hybrid Approach to Solve Cold Start Problem in Recommender Systems using Association Rules and Clustering Technique" International journal of computer applications 4th July 2013

[14]. Anand Kishor Pandey, Dharmveer Singh Rajpoot "Resolving Cold Start problem in recommendation system using demographic approach", IEEE 17th July 2017.

[15]. Zhenzhen Xu, Fuli Zhang, Wei Wang, Haifeng Liu, and Xiangjie Kong "Solving Cold-Start Problem by Combining Personality Traits and Demographic Attributes in a User Based Recommender System" Researchgate 5th May 2017

[16]. Rasoul Karimi, Christoph Freudenthaler, Alexandros Nanopoulos, Lars Schmidt-Thieme "Active learning for aspect model in recommendation system" IEEE 12th July 2011

[17]. Jyotirmoy Gope, Sanjay Kumar Jain "A Survey on Solving Cold Start Problem in Recommender Systems" IEEE, 21st December 2017