15th International Conference on Electronics Computer and Computation (ICECCO 2019)

Improving the Result of Personalized Questionnaire Towards Solving Cold User Problem.

Muhammad Abubakar
Faculty of Computer Science and Information Technology
Bayero University Kano.
Kano, Nigeria.
Muhammadabubakar08155@gmail.com

Kabiru Umar
Faculty of Computer Science and Information Technology
Bayero University Kano.
Kano, Nigeria.
Ukabir.se@buk.edu.ng

Abstract—Collaborative filtering techniques is among the popular approaches used in addressing product recommender systems, which uses ratings and predictions to make new suggestions. However the major weakness of collaborative filtering approaches is cold user problem. Literature investigation has shown that cold user problem could be effectively addressed using active learning technique of administering personalized questionnaire. Unfortunately, the result of personalized questionnaire technique could contain some user preference uncertainties where the product database is too large (as in Amazon.com). This research work tends to address the weakness of personalized questionnaire technique by applying the active learning technique of uncertainty reduction over the result obtained from administering personalized questionnaire. This strategy has the tendency of resolving user preference uncertainties that could be associated with the result of personalized questionnaire. This research work is in progress. Preliminary result is encouraging.

Keywords—cluster; cold user; collaborative filtering;recommendation system; active learning.

I. Introduction

People have being making decision when trying to buy an item, and they usually link their opinion with that of others. They ask their friends who are trustworthy about which product would they suggest for them to buy [31]. In its early stage product recommender systems did not create a separate research area and their roots can be traced back to the cognitive science, information retrieval, forecasting theories, and approximation theory and management science. However the growth of internet and the rise of e-commerce solution lead to the development of the online product recommender systems, and since the mid 1990's product recommender system have become an important research domain [32]. Product recommender system is an information system filtering technique that uses ratings and predictions to make new suggestions. It is an assistive model for users with the interest of suggesting a set of new products for users to view [32]. Product recommender system is a subclass of information filtering system that seek to predict preference that a user would give to an item [33]. Cold user problem could be effectively addressed using active learning technique of administering personalized questionnaire in which set of personalized questions are presented to new user, and conversely the result of the questionnaire is used to determine which cluster to place the user[1]. The techniques of personalized questionnaires could be effective where database of product is not too large [1]. Another active learning technique of uncertainty reduction has also been applied in addressing cold user problem [2, 3]. This involve presenting several items (one at a time) to a new user to rate. Then the resulting user ratings are used in associating new user to a given cluster of existing user. However this technique is not very effective when items database is too large [2] and may involve presentation too many items to the user [3]. Combining the strength of personalized questionnaire strategy and uncertainty reduction strategy is worthy of exploration. By applying uncertainty reduction to the result obtained from administering personalized questionnaire prior to user cluster assignment, it is optimistic that performance of the recommender system could be greatly improved towards achieving better recommendation to a new user. This research work aims at developing a system for improving solution to cold user problem of online recommendation system by reducing uncertainty in the result of personalized questionnaire.

The remaining of this paper is organized as follows. Section 2 review of related works. Section 3 presents the popular algorithms used in product recommendation system. Section 4 present the proposed model of the system. Section 5 present the expected result of the proposed work. The conclusions are presented in Section 6.

II. Review of Related Work

In this section we have presented a review of related work. Pozo et al. [1] worked on exploitation of past users interest and predictions in an active learning method for dealing with cold start in recommender systems. The problem they identified was when a new user does not receive pertinent recommendation, he/she may abandoned the system. They address cold user problem by presenting the new user a questionnaire in order to know his preference. Decision tree was used for implementation. The weakness of the research is result gotten from questionnaires could be too large and might contain some elements of uncertainty and many questions might challenge the willingness of the user to fill the questionnaire.

Nadimi-Shahraki [2] worked on cold-start problem in collaborative recommender systems by presenting an efficient methods based on ask-to-rate Technique. The researcher identified that a major challenge of the collaborative filtering approach can be how to make recommendation for a new user. Pearson's correlation was used to compute similarity of target users and previous users. The drawback of his research was new user need to rate many items before items of preference will be recommended to him. Karimi[3] worked on active learning approach for recommender systems. The problem the researcher identified that new user are not recommended an item. Items were presented for the new user to rate. Active learning techniques were used to acquire the most informative ratings. The drawback of this research was user preferences are gotten after rating many items. Kohrs and Merialdo[4] worked on improving collaborative filtering for new users by smart object selection. The researchers identified that recommending an item with collaborative filtering technique is difficult when few \zthe new user to rate. Pearson correlation was used. The drawback with this research was new users need to rate many items before items of preference will be recommended. Zlu[5] worked on addressing the item based cold start problem by attribute-driven active learning . The researcher identified that attributes of an item are not considered in active learning approach. The researcher leverage both active learning approach and attributes of an item in solving cold start item. Matrix factorization technique was used to compute the similarity of new item with other existing items. The weakness of the research was items might be wrongly rated by selected users due to lack of item experience. The summary of related work is presented in table I.

Table I. Summary of Related work

REF.	SUMMARY
[1]	Pozo et al. [1] worked on exploitation of past users interest and predictions in an active learning method for dealing with cold start in recommender systems. They address cold user problem by presenting the new user a questionnaire in order to know his preference. The weakness of the research is result gotten from questionnaires could be too large and might contain some elements of uncertainty and Many questions might challenge the willingness of the user to fill the questionnaire.
[2]	Mohammad-Hossein[2] worked on cold-start problem in collaborative recommender systems by presenting an efficient methods based on ask-to-rate Technique. The researcher identified that a major challenge of the collaborative filtering approach can be how to make recommendation for a new user. Pearson's correlation was used to compute similarity of target users and previous users. The drawback of his research ws new user need to rate many items before items of preference will be recommended to him.
[3]	Karimi[3] worked on active learning approach for recommender systems. The problem the researcher identified that new user are not recommended an item. Items were presented for the new user to rate. Active learning techniques were used to acquire the most imformative ratings. The drawback of this research was user preferences are gotten after rating many items.
[4]	Kohrs and Merialdo[4] worked on improving collaborative filtering for new users by smart object selection. The researchers identified that recommending an item with collaborative filtering technique is difficult when few information is known about the user. Objectswere presented for the new user to rate. Pearcon correlation was used. The drawback with this research ws new users need to rate many items before items of preference will be recommended.
[5]	Zlu[5] worked on addressing the item based coldstatrt problem by attribute-driven active learning . The researcher identified that attributes of an item are not considered in active learning approach. The researcher leverage both active learning approach and attributes of an item in solving cold start item. Matrix factorization technique was used to compute the similarity of new item with other existing items. The weakness of the research was items might be wrongly rated by selected users due to lack of item experience

III. Popular Algorithms Used in Product Recommendation System

In this section we present discussion of some of the algorithms used in product recommendation system. Among the popular algorithms used in product recommendation systems are Matrix factorization algorithm, decision tree algorithm, K-means algorithm and genetic algorithm. Decision tree algorithm uses decision based strategies in order to identify informative items to be selected for active learning. Matrix factorization algorithm has been recognized as the basic yet most effective model in recommendation systems. It represent each user and the item as an embedding vector. The mainly known as a subclass of collaborative filtering algorithm. K-means is undoubtedly one of the workhorses of machine learning faced with the exponential growth of data. Researchers have recently started to study strategies to speed up k-means and related clustering algorithms [34]. Clustering analysis is one of the primary data analysis methods and k-means is one of the most well-known popular clustering algorithms. The k-means algorithm is one of the frequently used clustering methods in data mining, due to its performance in clustering massive data sets [35].Genetic algorithm is a variant of stochastic beam search in which successor states are generated by combining two parent states rather than modifying a single state. The analogy to natural selection is the same as in stochastic beam search, except that it deals with sexual rather than asexual reproduction [36]. We investigate about 33 articles for extraction of algorithms used in recommendation system. Table II present the commonly used algorithms and their frequency of usage in the articles investigated.

Table II. Frequency of Algorithm usage

Algorithm	Frequency
Matrix Factorization Algorithm[5,7,8,9,15,18,20,23,27]	9
K-means Algorithm[14,16,24]	3
Genetic Algorithm[10,13]	2
Fuzzy Logic Algorithm[19]	1
RBF Neural network Algorithm[26]	1
Random Walk Algorithm[30]	1
Pearson correlation[2,4,6,22]	4
Decision tree[1,3]	2
Euclidean distance[11,12,25]	3
SGD[17]	1
UUB [21]	1
K nearest neighbor[28]	1
SVV[29]	1

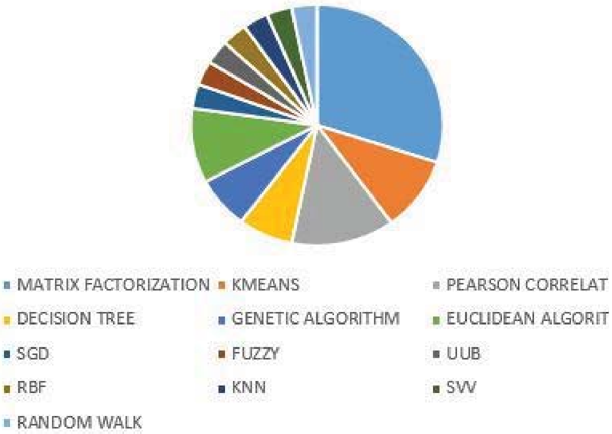

Fig. 1. Algorithms for Recommender Systems

Result of table II is plotted in a pie chart shown in figure 1. From the figure it can be seen that the most used algorithms used in product recommendation systems are matrix factorization and Pearson correlation algorithm with the percentage of 30% and 13.25% respectively. Active learning technique are usually implemented using decision tree[1]The basic steps in decision tree algorithm are presented in table III, followed by the basic steps of Pearson correlation algorithm.

Table III. Basic Decision tree Algorithm

```
functionBUILDDECISIONTREE(Ut,Rt-train,Rt-validation,Pt, curr TreeL)
    for user u Є of Ut do
        compute RMSEu1 on Rt-validation(u) and Pt(u)
    end for
    for candidate item j from Rt-train do
        split Ut into 3 child nodes based on j
        for user u Є Ut do
            find the child node where u has moved into
            compute RMSEu2 on Rt-validation(u) and Pt(u)
            Mu,i= RMSEu1 - RMSEu
end for
end for
δ = aggregate all Δu,i
discriminative item i∗ = argmaxiΔi
compute pi∗ by using item prediction average
if currentTreeLevel < maxTreeLevel and Mi∗≥ 0 then
create 3 child nodes Ut-child based on based on i∗ ratings
for child in child nodes do
exclude i∗ from Rt-child
BuildDecisionTree(Ut-child, Rt-child-train, Rt-child-validation, Pt-child, currentTreeLevel+1)
end for
end if
return i∗
end function
```

Table IV. Basic Pearson Algorithm

```
ρ =Correlation(X) /*x is features and data */
for 0≤i<len(ρ)do
    Wi=0
    for 0≤len(ρi) do
        ki=aabs(ρij)/*Absolute value*/
        auxi+=ki/*sample addition*/
    end
    wi=V(i)/auxi/*calculate weights*/
end
r=sort(w,byhighervalues
```

IV. Proposed Model

This section present our proposed model. The section begins by presenting its principles followed by architectural design. The proposed system is a personalized questionnaire approach of solving sold start problem tailed with uncertainty reduction of presenting an item for a user to rate.

A. *Principles*

The technique of personalized questionnaire is trailed by uncertainty reduction technique of asking a new user to rate an item before recommendation. The dominant problem of collaborative filtering effectiveness is cold start i.e. failure to make recommendation for new user, because collaborative filtering require historical information of user or product before recommendation. In this research our exertion will be on cold user. Even though personalized questionnaire technique is a good attempt of solving cold start user problem, the technique has the tendency of recommending items to the new user that does not satisfy his taste especially when the database is large. We employ a technique of uncertainty reduction to reduce user preference uncertainty inherent in the result of personalized questionnaire by presenting highly rated items from the result of the questionnaire for the user to rate. This principle is the basic upon which our proposed model is built.

B. *System Architecture*

A most important component of recommender system is the database. All personalized and non-personalized recommendation techniques are manipulating the data in the database. The technique used in this system is personalized questionnaire mended by uncertainty reduction technique. The personalized questionnaire present personalized question to the new user after highly rated item from the result of the questionnaire are presented to the user to rate based on his/ her preference.

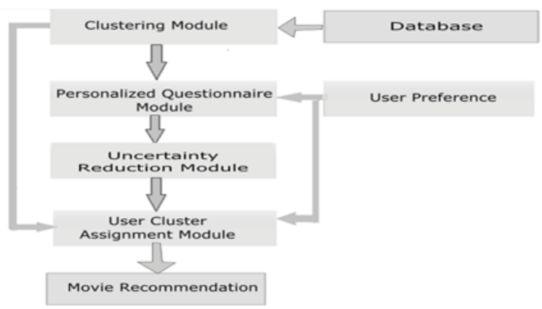

Fig. 2. System Architecture.

The system is composed of four major components, namely clustering module, personalized questionnaire module, uncertainty reduction module, and user cluster assignment module. These modules are described in the following sub sections:

i. Clustering Module

This module group existing users having similar movie interest. Users that happens to be in a cluster will be recommended similar movies. Kmeans clustering technique were employed to achieve the clustering.

ii. Personalized Questionnaire Module

This module present a new some personalized questions in order to know the users preferences. The new user is asked genre of his choice after which highly rated items form selected genre are presented to the user to select. Decision tree was used for implementation.

iii. Uncertainty Reduction Module

This module select highly rated items from the result of questionnaire for the new user to rate in order to know the users preference.

iv. User Cluster Assignment Module

This module assign a new user to a cluster so that the new user will be recommended items recommended to the existing users in the cluster.

V. Proposed Experiment

In this section we evaluate the proposed recommendation system. It containd three different components.

A. Dataset

Movielens dataset exists in three different sizes which are 100 KB, 1 MB, 10 MB and ratings. The movielens dataset is available and can be downloaded from grouplens website. The dataset contains 10 million ratings, 100,000 tag application applied to 10,000 movies by 72,000 users released in 1/2009.

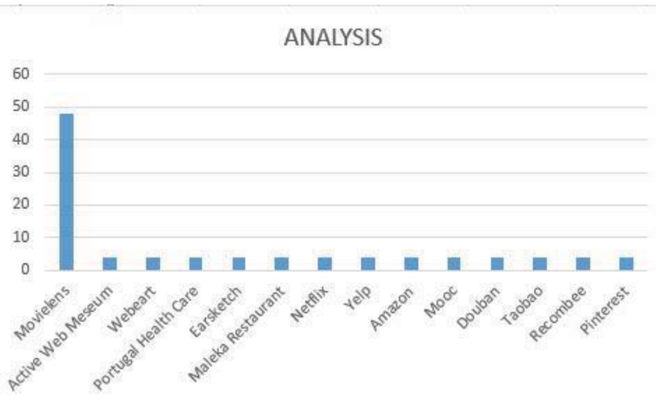

Fig. 3. Dataset

B. Performance Evaluation Metrics

Movielens dataset is the most used dataset in movie recommendation system[1,2,3,10,15,19,21,22,28,29,5,13]. Most of the researchers uses RMSE and Precision to evaluate the performance of their system. Based on the analysis below we will use the top two most used metrics.

RMSE: The RMSE is calculated using by computing the mean value of all the differences square between the actual and predicted ratings and then find the square root out of the result. 2

1. RMSE= $\sqrt{\frac{1}{n}\sum_{i=1}^{n}\ (ar-pr)^2}$

n is the total number of ratings

ar is the actual ratings

pr is the predicted ratings

2. Precision:

The precision is calculated as the total number of correctly recommended item over total number of recommended items.

Precision = Number of correctly recommended items/ Total recommended items.

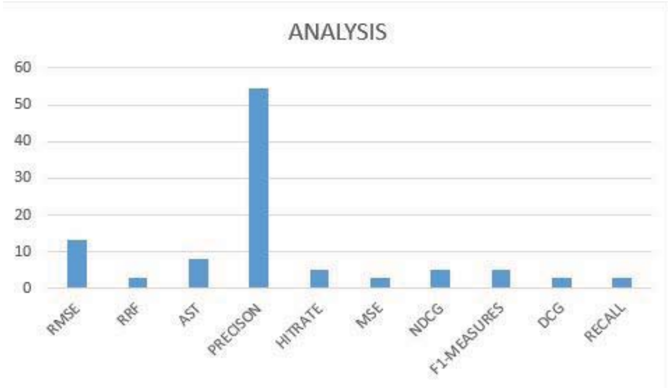

Fig. 4. Performance Metrics

C. Experimental design

In our experimental design we intend to perform four different experiment. The experiments are as follows;

- Personalized questionnaire approach of solving user cold start implemented using Movielens dataset of 1M size.
- Personalized questionnaire implemented using Movielens dataset of 10M size.
- Personalized questionnaire with uncertainty reduction implemented using Movielens dataset of 1M size.
- Personalized questionnaire with uncertainty reduction implemented using Movielens dataset of 10M size.

Movie lens dataset will be used in the four experiments above. We will compare our proposed approach with personalized questionnaire approach to solving cold start developed by Pozo[1] using Movielens dataset. We have to perform evaluation of the system in order to evaluate the recommendation by measuring the RMSE recall and Precision.

VI. Conclusion and Future Work

We proposed using active learning technique of presenting items for a user to rate in order to reduce user preference uncertainty inherent in personalized questionnaire. Decision tree will be used will be used for implementation. The proposed system will be evaluated using movie lens dataset with respect to root mean square error, precision and recall. In the future we propose implementation of the proposed technique.

References

[1] Pozo, M., Chiky, R., Meziane, F., & Métais, E. (2018, September). Exploiting past users' interests and predictions in an active learning method for dealing with cold start in recommender systems. In Informatics (Vol. 5, No. 3, p. 35). Multidisciplinary Digital Publishing Institute.

[2] Nadimi-Shahraki, M. H., & Bahadorpour, M. (2014). Cold-start problem in collaborative recommender systems: efficient methods based on ask-to-rate technique. Journal of computing and information technology, 22(2), 105-113.

[3] Karimi, R., Freudenthaler, C., Nanopoulos, A., & Schmidt-Thieme, L. (2011, April). Active learning for aspect model in recommender systems. In 2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM) (pp. 162-167). IEEE.

[4] Merialdo, A. K. B. (2001). Improving collaborative filtering for new-users by smart object selection.

[5] Zhu, Y., Lin, J., He, S., Wang, B., Guan, Z., Liu, H., & Cai, D. (2019). Addressing the item cold-start problem by attribute-driven active learning. IEEE Transactions on Knowledge and Data Engineering.

[6] Pazzani, M. J. (1999). A framework for collaborative, content-based and demographic filtering. Artificial intelligence review, 13(5-6), 393-408.

[7] Patil, P., Sata, M., Fancy, C. and Saranya, D. (2018).Demographic collaborative and content filtering technique based hybrid recommender system, 10(2-6).

[8] Vishwajith, V., Kaviraj, S., & Vasanth, R.(2019). Hybrid Recommender System for Therapy Recommendation, 13(5-6), 393-408.

[9] Han, Q., Ji, M., de Troya, I. M. D. R., Gaur, M., & Zejnilovic, L. (2018, October). A Hybrid Recommender System for Patient-Doctor Matchmaking in Primary Care. In 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA) (pp. 481-490). IEEE.

[10] Neysiani,B.,Soltani,N., Mofidi, R., & Nadimi-Shahraki,M.(2019). Improving Performance of Association Rule-Based Collaborative Filtering Recommendation Systems using Genetic Algorithm, 02(06).

[11] Smith, J., Weeks, D., Jacob, M., Freeman, J., & Magerko, B. (2019, March). Towards a Hybrid Recommendation System for a Sound Library. In IUI Workshops.

[12] Nouh, R. M., Lee, H. H., Lee, W. J., & Lee, J. D. (2019). A Smart Recommender Based on Hybrid Learning Methods for Personal Well-Being Services. Sensors, 19(2), 431.

[13] Alhijawi, B. J. M. (2017). The Use of the Genetic Algorithms in the Recommender Systems (Doctoral dissertation, Hashemite University).

[14] Mahadi, M. I., Zainuddin, N., Shah, N. B. A., Naziron, N. A., & Rum, S. F. M. E-Halal Restaurant Recommender System Using Collaborative Filtering Algorithm.

[15] Vozalis, Manolis, and Konstantinos G. Margaritis. "Collaborative filtering enhanced by demographic correlation." AIAI symposium on professional practice in AI, of the 18th world computer congress. 2004.

[16] Zarzour, H., Al-Sharif, Z., Al-Ayyoub, M., & Jararweh, Y. (2018, April). A new collaborative filtering recommendation algorithm based on dimensionality reduction and clustering techniques. In 2018 9th International Conference on Information and Communication Systems (ICICS) (pp. 102-106). IEEE.

[17] He, X., He, Z., Du, X., & Chua, T. S. (2018, June). Adversarial personalized ranking for recommendation. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (pp. 355-364). ACM.

[18] Sun, M., Li, F., & Zhang, J. (2018). A multi-modality deep network for cold-start recommendation. Big Data and Cognitive Computing, 2(1), 7.

[19] Ayyaz, S., Qamar, U., & Nawaz, R. (2018). HCF-CRS: A Hybrid Content based Fuzzy Conformal Recommender System for providing recommendations with confidence. PloS one, 13(10), e0204849.

[20] Brigui-Chtioui,I., Caillou,P., & Negre, E.(2018). Intelligent Digital Learning: Agent-Based Recommender System.

[21] Ekstrand, M. D., Tian, M., Azpiazu, I. M., Ekstrand, J. D., Anuyah, O., McNeill, D., & Pera, M. S. (2018, January). All the cool kids, how do they fit in?: Popularity and demographic biases in recommender evaluation and effectiveness. In Conference on Fairness, Accountability and Transparency (pp. 172-186).

[22] Gohari, F. S., & Tarokh, M. J. (2016). A New Hybrid Collaborative Recommender Using Semantic Web Technology and Demographic data. International Journal of Information and Communication Technology Research, 8(2), 51-61.

[23] Divya, N., Sandhiya, S., Liz, D. A. S., & GNANAOLI, P. (2018). A collaborative filtering based recommender system using rating prediction. International Journal of Pure and Applied Mathematics, 119(10), 1-7.

[24] Geetha, G., Safa, M., Fancy, C., & Saranya, D. (2018, April). A hybrid approach using collaborative filtering and content based filtering for recommender system. In Journal of Physics: Conference Series (Vol. 1000, No. 1, p. 012101). IOP Publishing.

[25] Wang, Y., Wang, M., & Xu, W. (2018). A sentiment-enhanced hybrid recommender system for movie recommendation: A big data analytics framework. Wireless Communications and Mobile Computing, 2018.

[26] Guo, Y., Yin, C., Li, M., Ren, X., & Liu, P. (2018). Mobile e-commerce recommendation system based on multi-source information fusion for sustainable e-business. Sustainability, 10(1), 147.

[27] Drdák, J. (2018). Optimalizace doporučovacích algoritmů.

Jazayeriy, H., Mohammadi, S., & Shamshirband, S. (2018). A fast recommender system for cold user using categorized items. Mathematical and Computational Applications, 23(1), 1.

[28] Jazayeriy, Hamid, Saghi Mohammadi, and Shahaboddin Shamshirband. "A fast recommender system for cold user using categorized items." Mathematical and Computational Applications 23.1 (2018): 1.Uçar, Ahmet, Karl E. Lonngren, and Er-Wei Bai. "Synchronization of the unified chaotic systems via active control." Chaos, Solitons & Fractals 27.5 (2006): 1292-1297.

[29] Zarzour, H., Al-Sharif, Z., Al-Ayyoub, M., & Jararweh, Y. (2018, April). A new collaborative filtering recommendation algorithm based on dimensionality reduction and clustering techniques. In 2018 9th International Conference on Information and Communication Systems (ICICS) (pp. 102-106). IEEE.

[30] Eksombatchai, C., Jindal, P., Liu, J. Z., Liu, Y., Sharma, R., Sugnet, C., ... & Leskovec, J. (2018, April). Pixie: A system for recommending 3+ billion items to 200+ million users in real-time. In Proceedings of the 2018 World Wide Web Conference (pp. 1775-1784). International World Wide Web Conferences Steering Committee.

[31] Kangas, Sonja. "Collaborative filtering and recommendation systems." VTT information technology (2002).

[32] T. Adovicius., "A hidden Markov model for collaborative filtering." Management Information Systems Quarterly, Forthcoming (2005).

[33] Y. Taysuzoglu, "Collaborative filtering enhanced by demographic data", Memorial University of Newfoundland(2018).

[34] Kurihara, Kenichi, Max Welling, and Yee Whye Teh. "Collapsed Variational Dirichlet Process Mixture Models." IJCAI. Vol. 7. 2007.

[35] Gunaruwan, T. Lalithasiri, and W. Neluka Gunasekara. " A Modular Framework for Extensible and Adaptable Recommendation Algorithms." NSBM Journal of Management 2.1 (2007).

[36] Russel, Stuart, and Peter Norvig. "Artificial Intelligence: A Modern Approach Thrid Edition." Person Education, Boston Munich (2010)