IMPROVING PERSONALIZED QUESTIONNAIRE WITH REDUNDANCY REDUCTION FOR ADDRESSING COLD USER PROBLEM

[1]Kabiru Umar and [2]Muhammad Abubakar

[1]Department of Software Engineering, Faculty of Computer Science and Information Technology, Bayero University Kano.
[2]Department of Computer Science, Faculty of Computer Science and Information Technology, Bayero University Kano.

[*]Corresponding Author's Email: ukabir.se@buk.edu.ng

ABSTRACT

User-based and item-based collaborative filtering techniques are among most explored strategies of making products' recommendations to Users on online shopping platforms. However, a notable weakness of the collaborative filtering techniques is the cold start problem. Which include cold user problem, cold item problem and cold system problem – i.e., the failure of collaborative filtering to make recommendation of products to a new user, failure of an item to be recommended, or combination of the two respectively. Literature investigation has shown that cold user problem could be effectively addressed using technique of personalized questionnaire. Unfortunately, where the products' database is too large (as in Amazon.com), results obtained from personalized questionnaire technique could contain some user preference uncertainties. This paper presents technique of improving personalized questionnaire with uncertainty reduction technique. In addition, the paper presents classification of product recommendation systems. In this work we will be limited to user-based cold start. Experimentation was conducted using Movielens dataset, where the proposed technique achieved significant performance improvement over personalized questionnaire technique with RMSE, Precision, Recall, F1 and NDCG of 0.200, 0.227, 0.261, 0.174 and 0.249 respectively.

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

INTRODUCTION

Earlier to a product been purchased by a buyer some decisions have to be made by the buyer. Friends and relatives do assist in making such decisions because of commonality of interest between the buyer and friends and relatives (Sonya, 2002). At first separate research area where not created for product recommendation system. Early product recommendation systems effort done by researchers can be seen in forecasting theories, cognitive science, approximation theories, information retrieval, and management science. However the tremendous improvement seen in internet today gave birth to e-commerce solutions. Today products were bought and recommended based on user taste and since then product recommendation system become one of the important research domain (Adovicius, 2005).

Product recommender system attempt to learn the interest of a user so that new suggestions can be made to him or her. It is an assistive model for users to prevent them from skimming over billions of products online by predicting product of interest to the user (Adovicius, 2005). Product recommender system can be seen as a subclass of information filtering system that suggest user inclination towards a product (Taysuzoglu, 2018). Literature investigation showed that Cold user problem (failure of a new user to be recommended an item) could be addressed using active learning technique. A new user is presented personalized questionnaire so that his/ her interest could be acquired, and then the result of the personalized questionnaire is used to determine which cluster to place the new user so that recommendations are provided to him as though he/she is an existing user (Smith et al., 2019).

Unfortunately, the techniques of administering personalized questionnaires could be effective in situations where database of product is not too large (Pozo et al., 2018). Uncertainty reduction technique is also another technique applied in addressing cold user problem (Nadimi-Shahraki and Bahadorpour, 2014; Karimi et al., 2011). It involve presenting several items (one at a time) to a new user to rate so that such ratings could be used as user interest to make appropriate suggestions. The resulting user ratings are used in assigning a new user to a given cluster of existing user that similar rating pattern as the new user. However this technique is not very effective in situations where items database is too large (Nadimi-Shahraki and Bahadorpour, 2014) and this may involve presentation of several items to the new user(Karimi et al., 2011).

By applying uncertainty reduction strategy to the result obtained from administering personalized questionnaire to a new user prior to assigning the new user to a cluster, it is optimistic that performance of the product recommender system could be tremendously improved. This will help towards attaining better recommendations to a new user. This research work aims at improving personalized

questionnaire approach of solving cold user problem of online recommendation system by reducing uncertainty in the result of the 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 classification of product recommendation system. Section 5 presents the proposed model of the system. Section 6 present the expected result of the proposed work. The conclusions are presented in Section 7.

Review of Related Work

We have presented in this section a review to our work. Pozo et al. (2018) worked on a personalized questionnaire approach of solving cold start problem. The researcher realized that when a new user is not recommended items of his choice, the user might choses to restrain the recommendation system. The researcher address cold user problem by personalized questionnaire to a new user of recommendation system purposely capture the new user's interest. Unfortunately result gotten from personalized questionnaire could be too large, which might result into wrong item prediction. A new user that is presented many questions to answer might refrain him/herself from the system. Decision tree was used to implementation implement the personalized questionnaire.

Nadimi-Shahraki and Bahadorpour (2014) attempt to solve cold user problem of collaborative filtering. The new user is asked to rate some item in order to capture his interest. The researcher noted that the major challenge to collaborative filtering is cold user problem. Similarities of the new user

and the previous users is computed using Pearson correlation. The weakness of this technique is that several items need to be rated before recommendation that suits the new user are provided. Karimi et al.(2011) also work on active learning technique of presenting items to a new user to rate. The researcher identified that new users are failed to be given recommendations due to lack of information regarding the new user. The new user is presented items to rate. The main drawback of this approach is that new user have to rate several items before recommendations are made to him.

Medialdo(2001) improved collaborative filtering by a smart object selection strategy. A new user is presented a picture of items for him/her to select. The researchers identified that in collaborative filtering system there is difficulty for the system to recommend products to a new user. The technique that was used by the researcher is Pearson correlation. But the weakness of this techniques is that too many objects has to be presented before recommendations that suits the new user are provided. Zlu(2019) on the other hand addresses item based collaborative filtering. It was noted by the researcher that attributes of an item are not usually considered before recommendation. A new item to a recommendation system is presented to some existing users to rate purposes to know the group of users interested in that item. To address item based cold start the researcher combines both attributes of an item and active learning techniques. The algorithm used for implementation is matrix factorization algorithm. One of the problem of this technique is that items might be wrongly rated by existing users due to lack of users experience. The summary of related work is presented in table I.

Table I. Summary of Related work

Reference	SUMMARY
(Pozo, 2018)	Pozo(2018) worked on a personalized questionnaire approach of solving cold start problem. The researcher address cold user problem by presenting personalized questionnaire to a new user purposely capture the new user's interest. However result gotten from personalized questionnaire could be too large, which might result into wrong item prediction. A new user that is presented many questions to answer might refrain him/her self from the system. The personalized questionnaire implemented using decision tree.
(Nadimi-Shahraki & Bahadorpour,2014)	Nadimi-Shahraki and Bahadorpour(2014)worked on cold-start weakness of collaborative recommender systems and try to improve it. A new user is asked to rate some items. Pearson's correlation was used for implementation. The drawback of their research was new user need to rate many items before items of preference will be recommended to him.
(Karimi et al.,2011)	Karimi et al.(2011) worked on cold start problem of recommendation systems. Active learning technique of "ask to rate" was used. Unfortunately user preferences are gotten after rating many items.
(Merialdo ,2001)	Merialdo (2001) worked on cold start problem by a smart object selection strategy. The researchers identified that in collaborative filtering system there is difficulty for the system to recommend products to a new user. The technique that was used by the researcher is Pearson correlation. But the weakness of this technique is that too many objects have to be presented before recommendation.
(Zlu, 2019)	Zlu(2019) on the other hand addresses item based collaborative filtering. It was noted by the researcher that attributes of an item are not usually considered before recommendation. A new item to a recommendation system is presented to some existing users to rate purposes to know the group of users interested in that item. To address item based cold start the researcher combines both attributes of an item and active learning techniques. The algorithm used for implementation is matrix factorization algorithm. One of the problem of this technique is that items might be wrongly rated by existing users due to lack of users experience.

Popular Algorithms Used in Product Recommendation System

In this section algorithms that are used in product recommendation system are discussed. The most popular algorithms used in implementing product

recommendation systems are Matrix factorization algorithm, decision tree algorithm, Kmeans algorithm and genetic algorithm. The most effective model in recommendation systems is recognized to be matrix factorization. Embedding vector is used to represent both items and users. K-means grows in an exponential

rate, and it is the work force of machine learning. There are numerous effort by researchers to improve kmeans algorithm Kenichi(2007).

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 Gunaruwan(2007). Genetic algorithm is a type of stochastic beam search that successor states where produced by joining two parent states not modifying a single state. The similarity to natural selection is the same as in stochastic beam search only that it deals with sexual not asexual reproduction Stuar,& Norvig(2010). About 33 articles were investigated in getting the algorithms used in product recommendation system.

Below in Table II present the most used algorithms and their frequency of the articles examined followed by figure 1 depicting its pie chart.

Table II. Frequency of Algorithm usage

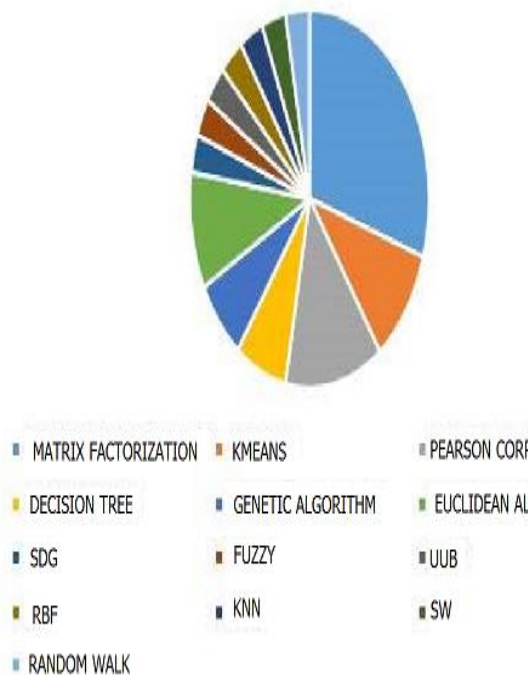| Algorithm	Frequency
Matrix Factorization Algorithm	9
K-means Algorithm	3
Genetic Algorithm	2
Fuzzy Logic Algorithm	1
RBF Neural network Algorithm	1
Random Walk Algorithm	1
Pearson correlation	4
Euclidean distance	3
Decision tree	2
SGD	1
K nearest neighbor	1
SVV	1

Fig. 1. Algorithms for Recommender Systems

IV. Classification of product recommendation systems.

In this section a classification of product recommendation system was made. Jazayeriy(2018) realized that recommender system can be classified broadly into two, Personalized and non-personalized recommender system. The personalized recommender system according to Jazayeriy(2018) are content based filtering, collaborative filtering, demographic technique and knowledge based technique. The non-personalized recommendation techniques are aggregated opinion and product association (Alhijawi, 2017).

Literatures like Han et al.(2018)and Zarzour et al.(2018) embrace utility based technique as an additional technique for personalized recommender system after the first four technique identified by Jazayeriy(2018]. Furthermore knowledge based recommender system are categorized into two, case-based system and constrained-based systems (Zarzour et al., 2018)]. Case-cased system are based on case based reasoning, which relies on the similarities between a current case and the solutions that already exist in a database.

The interaction with a constraint based recommender system consists of four steps cycle: retrieve, reuse, revise and retention (Han et al.,2018). Constraint-based systems is based on a given set of preferences, constraint based systems provide possible solutions including explanations as to why these solutions were selected. We can reach to a conclusion that techniques used for personalized recommender systems are five (5) in number, they are collaborative filtering, content based filtering, demographic approach, utility and knowledge based filtering.

It should be well noted that hybrid systems can be formed as the recommendation accuracy is usually higher in hybrid systems (Geetha et al., 2018). Two to three techniques are normally combined for better recommendation. The combination of techniques leads to common knowledge increase, which contributes to better recommendations (Geetha et al., 2018). Figure 2 below depict the classification of product recommendation systems.
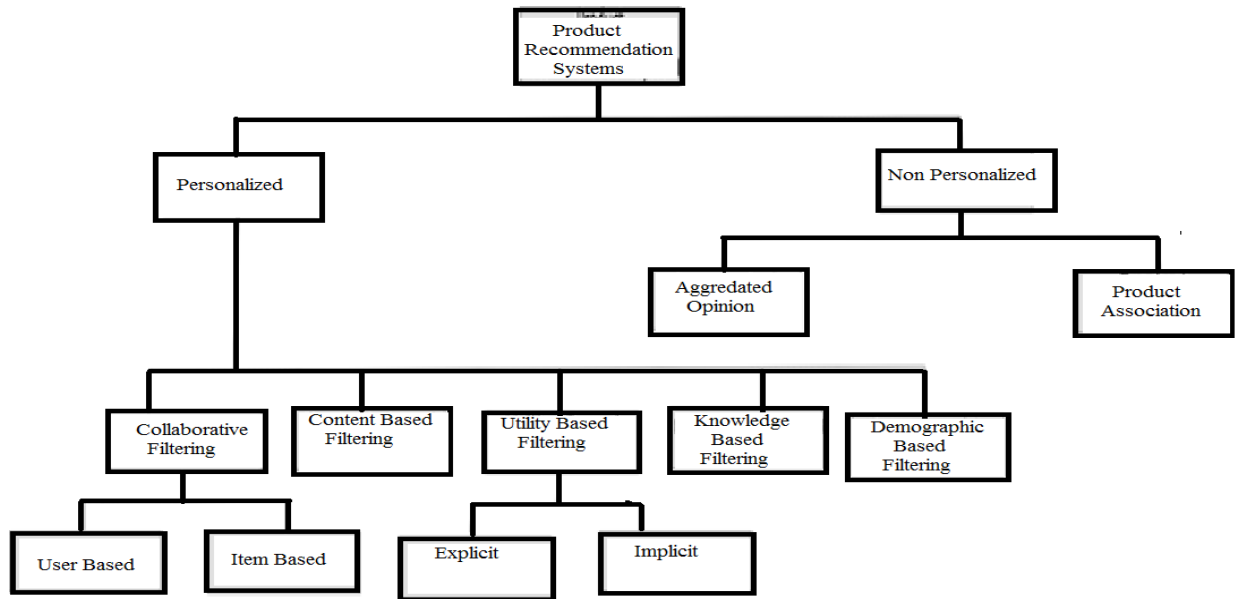
Fig 2. Classification of product recommender systems

The result of table II can be seen plotted in a pie chart shown in figure 1. From figure1 we can clearly see that the most used algorithms are matrix factorization algorithm and Pearson correlation algorithm with the percentage of 30% and 13.25% respectively. Most active learning techniques implementation are implemented using decision tree algorithm (Pozo et al., 2018). We present the basic steps involved in decision tree algorithm in Algorithm1, and below it are the basic steps involved in Pearson correlation algorithm in Algorithm 2.

Algorithm 1: Basic Decision tree

1:functionBUILDDECISIONTREE(Ut,Rt-train,Rt-validation,Pt, curr TreeL)
2:     for user u Є of Ut do
3:         compute RMSEu[1] on $R_{t-validation}$(u) and P_t(u)
4:     end for
5:     for candidate item j from $R_{t-train}$ do
6:         split U_t into 3 child nodes based on j
7:         for user u Є U_t do
8:             find the child node where u has moved into
9:             compute RMSEu2 on Rt-validation(u) and Pt(u)
10:            Mu,i= RMSEu1 - RMSEu
11: end for
12: end for
13: δ = aggregate all Δu,i
14: discriminative item i* = argmaxiΔi
15: compute pi* by using item prediction average
16: if currentTreeLevel < maxTreeLevel and Mi* ≥ 0 then
17: create 3 child nodes Ut-child based on based on i* ratings 18: for child in child nodes do
19: exclude i* from Rt-child
20: BuildDecisionTree(Ut-child, Rt-child-train, Rt-child-validation, Ptchild, currentTreeLevel+1) 21:end for
22: end if
23: return i*
24: end function

Algorithm 2: Basic Pearson Algorithm

1: ρ =Correlation(X) /*x is features
2: for 0≤i<len(ρ)do
3:     Wi=0
4:     for 0≤len(ρi) do
5:         ki=aabs(ρij)/*Absolute value*/
 6:         auxi+=ki/*sample addition*/ 7:   end
8:     wi=V(i)/auxi/*calculate weights*/
9: end
10:r=sort(w,byhighervalues

V.   Proposed Model

In this section the proposed model is presented. It comprises of the principles, system architectural diagram, and description of the components in the architectural diagram.

A. Principles

In this work, we use a technique of uncertainty reduction to reduce user preferences uncertainties inherent in the result of personalized questionnaire approach of solving cold user problem conducted by Pozo (2018). Highly rated items from the result of personalized are presented for the new user to rate before the new user is assigned to a cluster of users having similar interest.

B. System Architecture

Both active and passive learning techniques of product recommendation system get of the database in one way or the other. Data manipulation are usually made by both active and passive learning technique. Personalized questionnaire approach is improved with uncertainty reduction technique. Personalized questions are presented to the new user, after which highly rated items from the result of the personalized questionnaire are presented for the new user to rate. Below in figure 3 is the architectural design of our proposed technique.

The clustering module group existing users into various clusters. Users in the same cluster shares similar
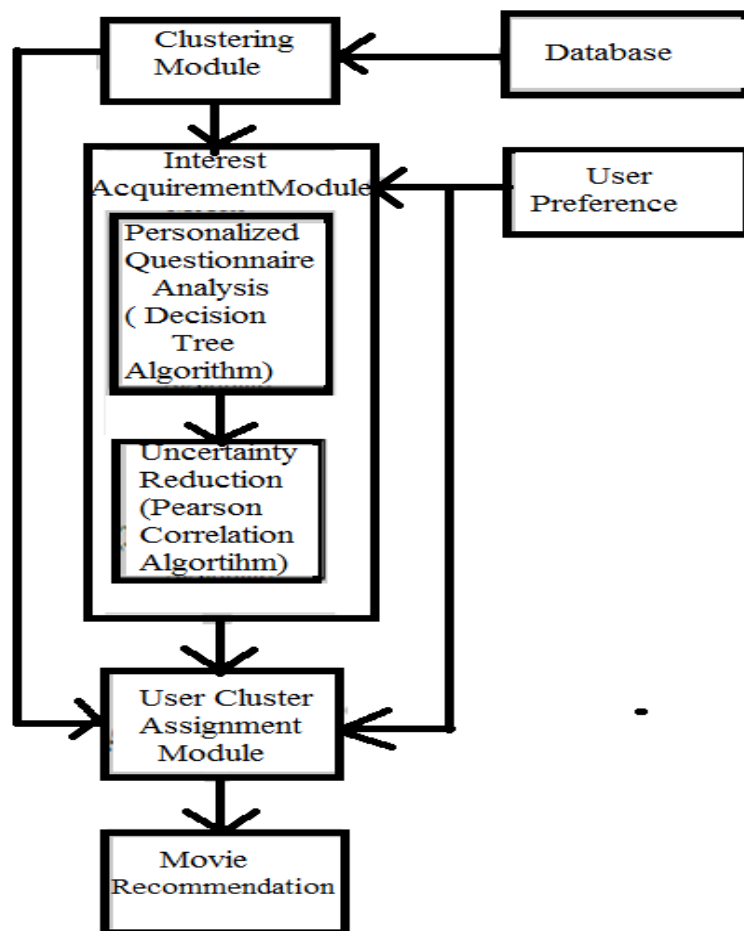
Fig. 3. System Architecture.

Three modules can be seen in figure 3 above, the modules include clustering module, interest acquirement module, and user cluster assignment module. The interest acquirement module is composed of two sub modules which personalized questionnaire module and uncertainty reduction module. These modules are described in the following sub sections:

 i. Clustering Module

interest and so are recommended similar product.

ii. Interest Acquirement Module

In this module a new user is asked some personalized questions in order to acquire the new user's interest. The questions include genre of choice, movie of choice e.t.c. The personalized questionnaire is implemented using decision tree algorithm. To know the new user preference highly rated items from the result

iv.   Uncertainty Reduction Module:

VI. Proposed Experiment

In this section we evaluate the proposed recommendation system.

   A. Dataset

Movielens dataset are available online in three sizes 100 KB, 1 MB, 10 MB and ratings. To download the dataset, you can visit grouplens website. It contains in it 10 million ratings, 100,000 tag application applied to 10,000 movies by 72,000 users released in 1/2009. Below in figure 4 is the dataset frequency used in related works.
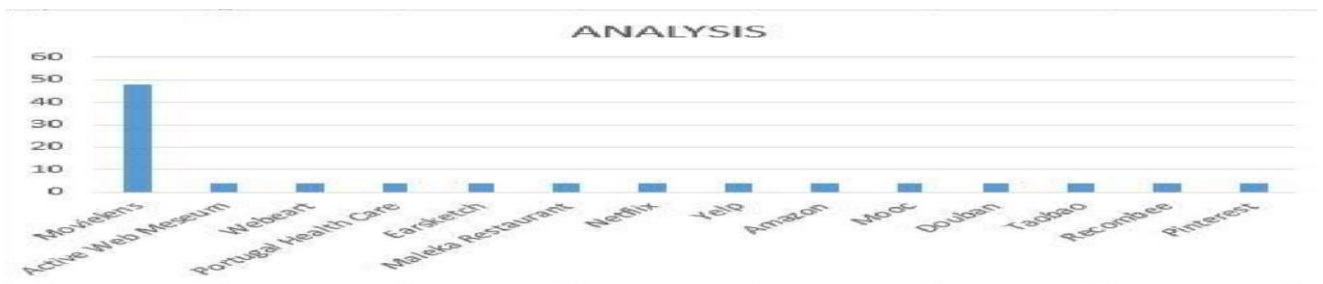
of personalized questionnaire are presented to the new user to rate.

iii.   User Cluster Assignment Module

In this module, a new user is assigned to cluster of users having interest as the new user so that similar recommendation will be made both to the existing user and new user.

Fig. 4. Dataset

   B. Performance Evaluation Metrics

Movielens dataset is the most used dataset in movie recommendation system. In this research the evaluation metrics we will use are RMSE, Precision, Racall, F1-measure, and NDCG. In figure 5 below is the frequency of performance metrics used in related work.
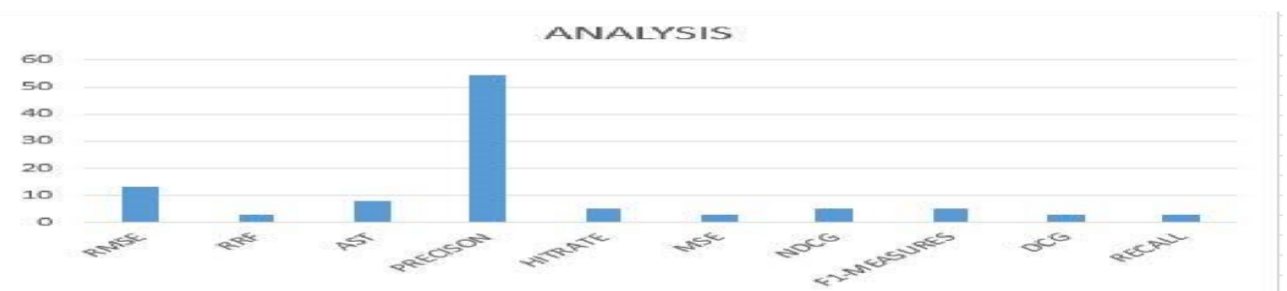
Fig. 5. Performance Metrics

1.   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

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

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.

3. Recall:

Recall is calculated as total number of correctly recommended items over the total items liked by the user.

Recall= total number of correctly recommended items total items liked by

the users.

4. NDCG:

It stands for normalized discounted cumulative gain. The mean of upper and lower bound score is calculated using NDCG. The

NDCG is calculated as discounted cumulative gain of the recommended order over discounted cumulative gain of the ideal

order.

NDCG=DCG

iDCG

- Personalized questionnaire implemented using Movielens dataset of 10M size taking into consideration precision, recall, RMSE, NDCG and F1 score.

5. F1 score: It is the harmonic mean of the precision and recall.

Data Size	Model	RMSE	Precision	Recall	F1	NDCG
1M	PREVIOUS MODEL (P.Q. Model)	0.698	0.583	0.636	0.610	0.313
	PROPOSED MODEL (P.Q. + U.R)	0.488	0.750	0.818	0.784	0.562
10M	PREVIOUS MODEL (P.Q. Model)	0.866	0.486	0.439	0.63	0.213
	PROPOSED MODEL (P.Q. + U.R Model)	0.698	0.621	0.639	0.706	0.426

$$F1 = \frac{2(precision*recall)}{Precision + recall}$$

- Personalized questionnaire with uncertainty reduction implemented using Movielens dataset of 1M size taking into consideration precision, recall, RMSE, NDCG and F1 score.

- Personalized questionnaire with uncertainty reduction implemented using Movielens dataset of 10M size taking into consideration precision, recall, RMSE, NDCG and F1 score. 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

C. 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 taking into consideration precision, recall, RMSE, NDCG and F1 score.

- Pozo[1] using Movielens dataset. We have to perform evaluation of the system in order to evaluate the recommendation by measuring the RMSE, recall, Precision, NDCG and F1 score.

VII. Result

Below is Table III, which contains RMSE, Precision and Recall result of both the previous model (P.Q.) and Proposed model(P.Q.+U.R).

Table III. Result

The Graph in figure 6 and figure 7 below presents RMSE result of 1M Movielens data size and 10M Movielens data size respectively. Graph in figure 8 and figure 9 presents Precision result of 1M Movielens data size and 10M movielens data size respectively. Graph in figure 10 and figure 11 presents Recall result of 1M Movielens data size and 10M movielens data size respectively. Graph in figure 12 and 13 presents F1 measure result of 1M Movielens data size and 10M movielens data size respectively, while Graph in figure 14 and 15 presents NDCG result of 1M Movielens data size and 10M movielens data size respectively.
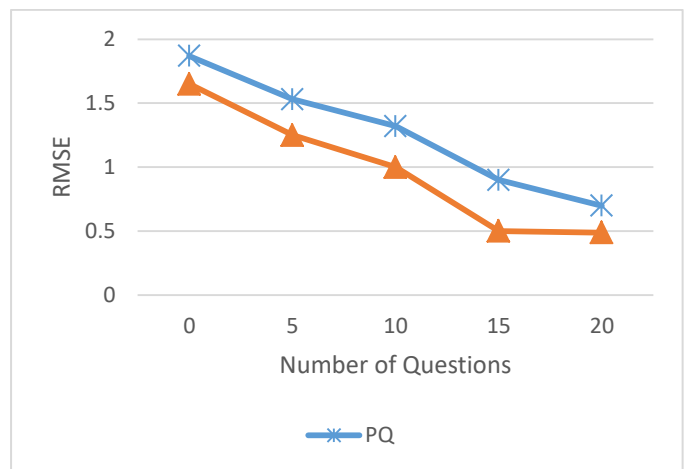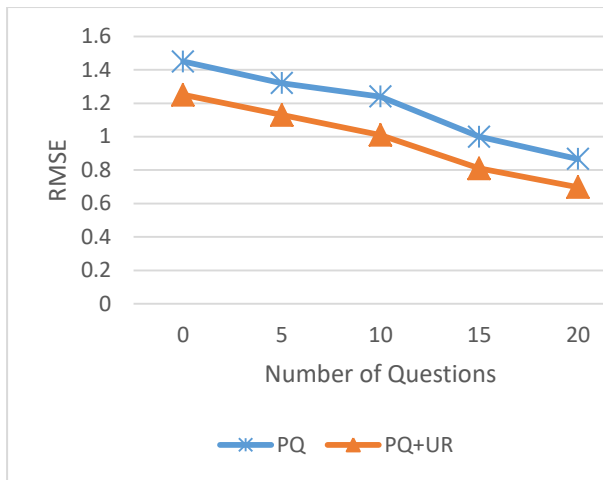

Figure 6. RMSE 1M dataset size

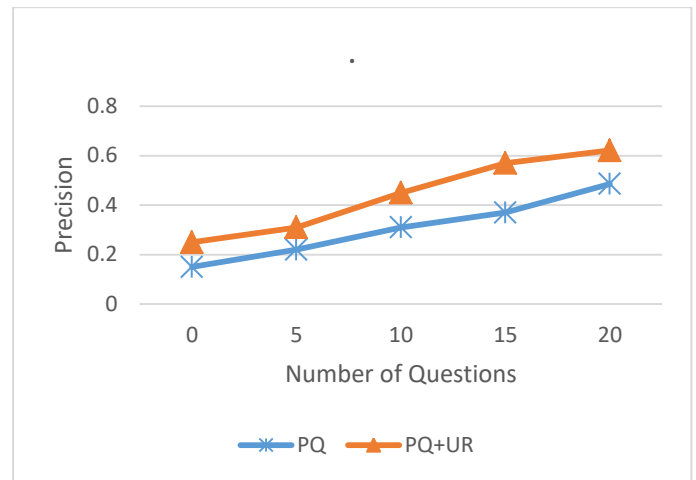Figure 7. RMSE 10M dataset size

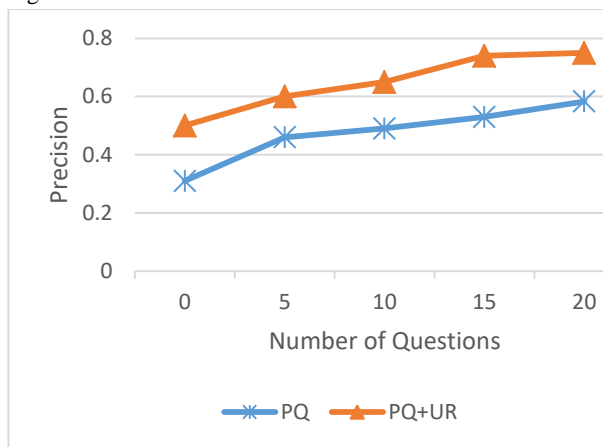Figure 9. Precision 10M dataset size
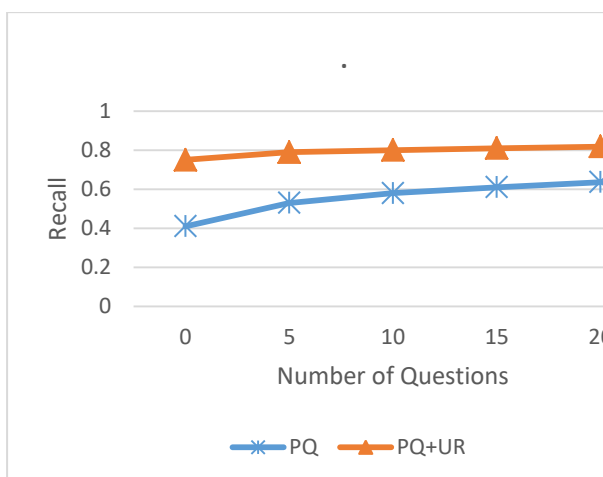

Figure 8. Precision 1M dataset size
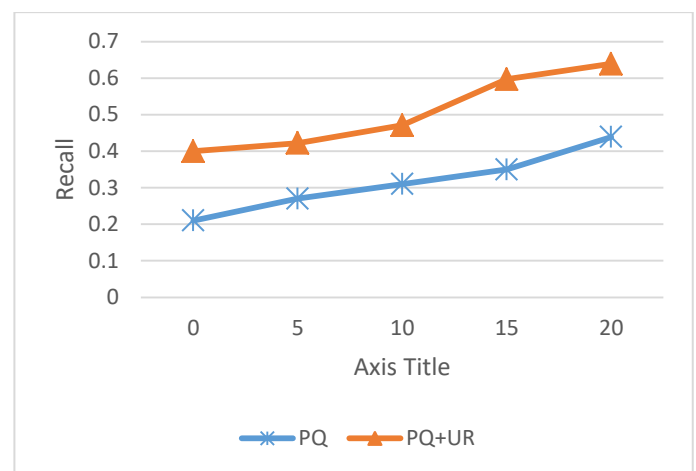

Figure 10. Recall 1M dataset size
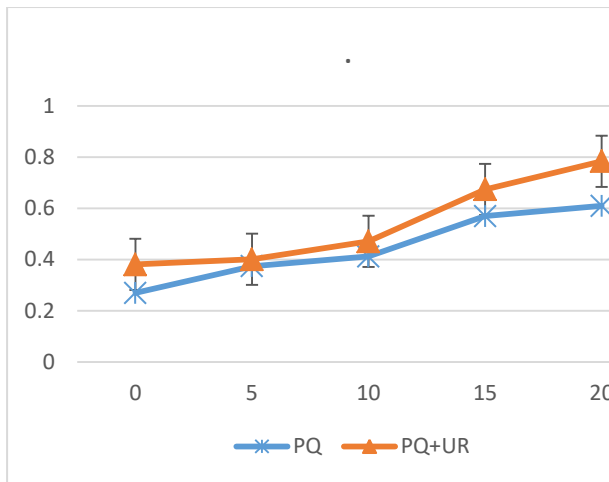

Figure 11. Recall 10M dataset size
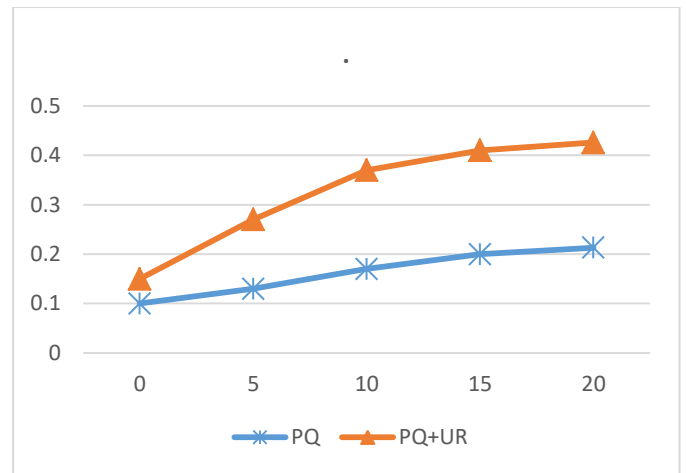
Figure 12. F1 1M dataset size
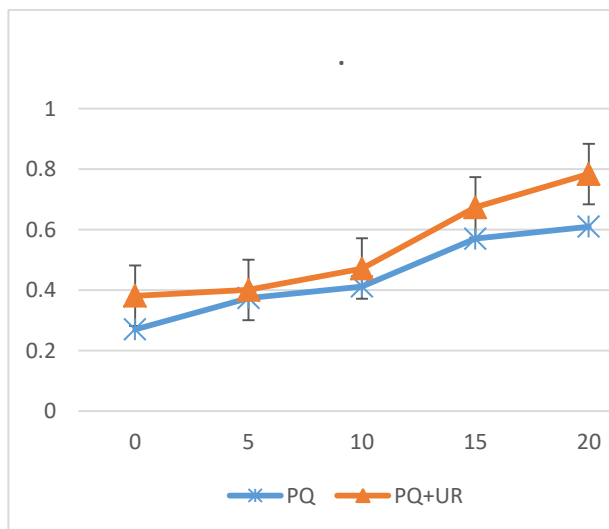

Figure 13. F1 measure 10M dataset size

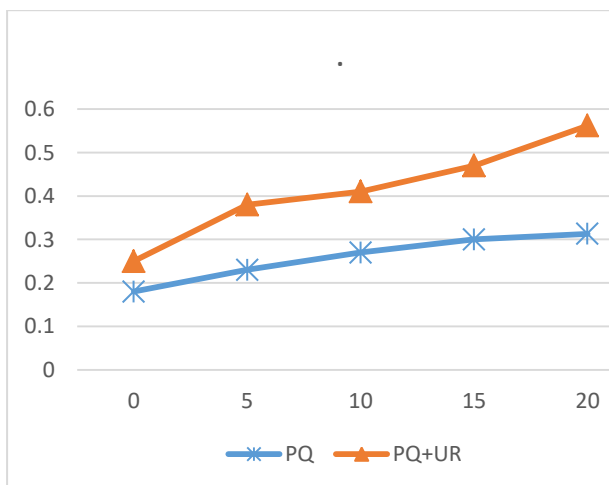Figure 14. NDCG measure 1M dataset size

Figure 15. NDCG measure 10M dataset size

VIII Conclusion and Future Work

In this worked we improved personalized questionnaire approach of solving cold user problem using uncertainty reduction strategy. Experimental results on Movielens dataset signify an improvement RMSE, Precision, Recall, F1 and NDCG of 0.200, 0.227, 0.261, 0.174 and 0.249 respectively in 1M dataset and 0.168, 0.159, 0.205, 0.076 and 0.213 in 10M dataset respectively over personalized questionnaire work of Pozo(2018).

References

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.

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.

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.

Merialdo, A. (2001). Improving collaborative filtering for new-users by smart object selection. In *ICMF*.

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

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

Zarzour, H., Maazouzi, F., Soltani, M., & Chemam, C. (2018, May). An improved collaborative filtering recommendation algorithm for big data. In *IFIP*

International Conference on Computational Intelligence and Its Applications (pp. 660-668). Springer, Cham.

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).

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).

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).

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*.

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

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

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.

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

Gunaruwan, T., & Neluka G. (2007). A Modular Framework for Extensible and Adaptable Recommendation Algorithms. NSBM Journal of Management 2.1.