[Volume 1 issue 3 April 2013]
Page No.123-128 ISSN :2320-7167

Possible Solutions of New User or Item Cold-Start Problem

Angira Bodhit[1], Kiran Amin[2]

1) Pursuing M.Tech. – IT , Computer & Information Technology Department,
U.V.Patel College of Engineering,Ganpat University, Gujarat, India.
angu712@gmail.com

2) Head of the Computer Department, U.V.Patel College of Engineering,
Ganpat University, Gujarat, India.
kiran.amin@ganpatuniversity.ac.in

Abstract

The new user cold start issue represents a serious problem in recommender systems as it can lead to the loss of new users who decide to stop using the system due to the lack of accuracy in the recommendations received in that first stage in which they have not yet cast a significant number of votes with which to feed the recommender system's collaborative filtering core. For this reason it is particularly important to design new similarity metrics which provide greater precision in the results offered to users who have cast few votes. The possible solutions, now in days, are profile expansion and developing a new method of finding similarity between existing users. In this paper, we focus on both of these methods. The metric has been tested on the Netflix and Movielens databases, obtaining important improvements in the measures of accuracy, precision and recall when applied to new user cold start situations. The profile expansion technique is proposed and evaluated by three kinds of techniques: item-global, item-local and user-local. These techniques are also tested on the Netflix and Movielens database. The results obtained by both the techniques are good and we can use that for our future work.

1. Introduction

With the advancement of electronic commerce, automated product recommendation has been perceived as a critical tool for boosting sales in online stores. By providing personalized recommendation of products to users, online stores have been able to increase revenue through up-selling and cross-selling[1,3]. There have been numerous ways of product recommendation methods that utilize various types of data and analysis tools. One of the most successful methods is collaborative filtering (CF) that recommends products based on the similarity of users in online stores that is calculated using users' ratings on items. CF has been proved to be successful by numerous studies and has been implemented by many real-world businesses. The most critical component of the CF mechanism is finding similarities between users effectively[2,4].

Recommender systems are commonly employed nowadays in any domains, being especially successful in e-commerce sites. Many online businesses have found in recommender systems the perfect tool to increase sales and customer satisfaction at the same time. The main reason of recommender systems success is that they can provide personalized recommendations, thus helping users to find valuable products for themselves among a wide amount of choices. In a society that generates an impressive amount of information every single day, recommender systems are an essential tool to guide users towards their needs[3,5,6].

The technique of Collaborative Filtering (CF) that recommends items based on the opinions of other users is very popular given its good results in many domains. In this technique, each user is represented by a profile consisting of her opinions about the items she has rated[2,3].

The system will look for users with a similar profile, and it will recommend items rated high by these similar users. Obviously, the user profile is incomplete, that is, it only contains opinions about a reduced set of items: those the user knows about and has decided to rate. The higher the profile, the more information the system has about the user. This usually results in better recommendations. On the other hand, when there is little available information, the recommender system performs poorly. This is the well-known cold-start problem[4].

In particular, Collaborative Filtering techniques suffer the so-called new user problem, that is, how to recommend items to users that have recently joined the system and thus have an empty or very small profile. Actually, this problem does not affect only new users, but also users that rate few items, despite having been using the system for a long time. For example, users that do not use the system very often, or that are reticent to rate items. In many domains where Collaborative Filtering techniques are used this is a rather common situation. Forexample, in movie recommendation systems a high number of users adopt this behaviour[2].

We will see the details of the main two possible solutions, named new similarity measure between two existing users and profile expansion technique, in the next section.

The results obtained by the above two methods are covered in the section 3. And the comparison of results and the conclusions are discussed in section 4.

2. Description

2.1 A new similarity measure technique[1]

There are many similarity measurement techniques which were used by the scientists, which are Pearson's correlation, cosine correlation, adjusted cosine for similarity measure between items, Constrained Pearson's correlation, Spearman's Rank Correlation etc. All of the above have some limitations.

The first one, Pearson's correlation (COR), measures the linear correlation between two vectors of ratings. The cosinemeasure (COS) looks at the angle between two vectors of ratings where a smaller angle is regarded as implying greater similarity. The third one, adjusted cosine, is used in some IBCF methods for similarity among items where the difference in each user's use of the rating scale is taken into account. The fourth one, constrained Pearson's correlation, is a slightly modified version of Pearson's correlation that allows only the pairs of ratings on the same side, e.g. both being positive or both being negative, to contribute to the increase in the correlation. The fifth one is called Spearman's Rank Correlation where similarity between two vectors is calculated based on the similarity of ranks of values in the vectors.

The two most-widely used similarity measures for CF are COR and COS.Besides of their popularity, they are limited to be used in new user cold-start situations where only a small number of ratings are available for similarity calculation. The major limitations can be briefly summarized as follows:

(1) Very limited number of co-ratings under data sparsity.
(2) If the number of co-rated items is 1, COR cannot be calculated and COS results in 1 regardless of differences in individual ratings.
(3) If all the available ratings of a given user are flat, e.g. <1,1,1> , <3,3,3> or <4,4,4>, COR cannot be calculated between the user (and, hence, often regarded as 0) and another since the denominator part of the correlation formula becomes zero.
(4) If two vectors are on the same line, e.g. vectors <2,2> and <3,3>, COS results in 1 regardless of the difference between the two.
(5) Both COR and COS can be sometimes misleading, where very different users may appear to be very similar to each other by the similarity measures, and vice versa.

Part of the above problems can be clearly observed through simple experiments using a sample dataset as seen in Fig. 1. Users 1 and 3 are showing very similar ratings for the two items in fig 1, but the cosine value between them are smaller than that of users 1 and 2 whose rating vectors are on the same line.

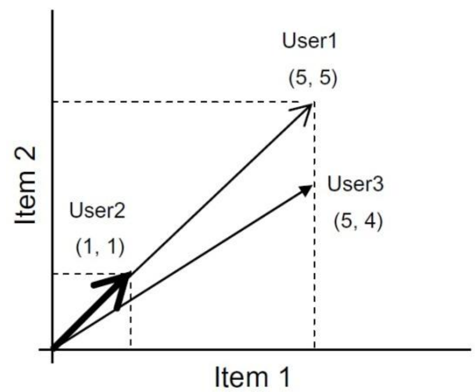

Fig. 1 Ratings given by 3 users on 2 items

To overcome these problems; the new similarity measure, named PIP (Proximity–Impact–Popularity) measure, is being developed. The main features of this measure are listed below:

(1) The measure should utilize domain specific meanings of data, rather than just employing traditional similarity or distance measures, in order to be more effective in cold-start recommendation conditions.
(2) In order to be more practical, the measure should allow easy plug-in to existing CF systems by replacing only the similarity measures of the systems, not requiring huge re-implementation or additional data collection.
(3) The measure should not only show better results in new user cold start conditions but also comparable results to other popular measures in non-cold-start conditions.

The measure is composed of three factors of similarity, Proximity, Impact, and Popularity, and hence, was named PIP. With the PIP measure, the similarity between two users, u_i and u_j, is calculated as:

$$SIM\ (u_i, u_j) = \sum_{k \in C_{i,j}} PIP(r_{ik}, r_{jk})$$

Where,r_{ik} and r_{jk} are the ratings of item k by user i and j, respectively, $C_{i,j}$ is the set of co-rated items by user $u_i$and u_j, and PIP(r_{ik}, r_{jk}) is the PIP score for the two ratings r_{ik} and r_{jk}. For any two ratings r_1 and r_2,PIP(r_1, r_2) = Proximity(r_1, r_2)* Impact(r_1, r_2)* Popularity(r_1, r_2).

First, the Proximity factor is based on the simplearithmetic difference between two ratings, but it further considers whether the two ratings are in agreement ornot, giving penalty to ratings in disagreement. That is, if two ratings are on the same side of a given ratingscale which is divided by its median, they are regarded to be in agreement.

Second, the Impact factor considers how strongly an item is preferred or disliked by buyers. When it is strongly preferred or disliked, we can regard that a clearer preference has been expressed for the item, and hence, bigger credibility can be given to the similarity.

Third, the Popularity factor gives bigger value to a similarity for ratings that are further from the average rating of a co-rated item.

Agreement:

For any two ratings r_1 and r_2, let R_{max} be the maximum rating and R_{min} the minimum in the rating scale, and let

$$R_{med} = \frac{R_{max} + R_{min}}{2}$$

A Boolean function Agreement(r_1,r_2) is defined as follows:
Agreement(r_1,r_2) = false if (r1 >Rmed and r2 <Rmed) or(r1 <Rmed and r2 >Rmed), and
Agreement(r_1,r_2) = true otherwise

Proximity:
A simple absolute distance between the two ratings is defined as:
D(r_1,r_2) = |r_1 - r_2| if Agreement(r_1,r_2) is true, and
D(r_1,r_2)= 2 * |r_1 - r_2| if Agreement(r_1,r_2) is false
Then the Proximity(r_1,r_2) is defined as:
Proximity(r_1,r_2) = $\{\{2*(R_{max} - R_{min}) + 1\} - D(r_1,r_2)\}^2$

Impact:
Impact(r_1,r_2) is defined as:
Impact(r_1,r_2) = $(|r_1-R_{med}| + 1)(|r_2-R_{med}| + 1)$ if Agreement(r1, r2) is true, and
Impact(r_1,r_2) = $\frac{1}{(|r1 - \text{Rmed}\,| + 1)(|r2 - \text{Rmed}\,| + 1)}$if Agreement(r1, r2) is false

Popularity:
Let μ_k be the average rating of item k by all users.
Then Popularity(r_1,r_2) is defined as:
Popularity(r_1,r_2) = $1 + (\frac{r_1 + r_2}{2} - \mu_k)^2$
if (r_1>μ_k and r_2>μ_k) or
(r_1<μ_k and r_2<μ_k), and
Popularity(r_1,r_2) = 1 otherwise.

2.2 Profile expansion technique[2]

Profile of any user is being created by the given set of data by user and by the set of ratings given by the user to the different items. The user profile is defined as a list of item-rating pairs, where each pair is made of an item $i \in I$ the user has rated, and the corresponding rating v_{ui},.$P_u$is defined as the user profile of a user $u \in U$. The user profile is comprised of *s*item-rating pairs, being *s* the number of items rated by the user.

P_u= <(i_1,v_{ui1}), (i_2,v_{ui2}), …. , (i_s,v_{uis})>

The kNN algorithms use the profile to compute the similarity between the user and each one of the other users in the system. Then, the most similar users are selected as neighbours. Finally, the neighbours are used to compute the recommendation.

It is important to note that this is a sequential process that begins with the user profile. If the user profile is very small, thesimilarity computation may not work properly, as there is not enough information about the user interests. As a paradigmaticcase, the Pearson correlation, a similarity measure commonly used in Collaborative Filtering, always reports a perfectsimilarity if two users have rated only a single item in common. Thus, with a small user profile, such similarity measureshould not really be used.

To alleviate this problem, Profile Expansion (PE) techniques are being used. The idea is to increase the size of the original profile with additional items, in order to improve the next steps. The authors thus define the expanded profile, P'_u, as the union of the original profile with up to *l* additional items. That can be seen in the following equation:

$P'_u = P_u$U<(i_{s+1},v_{uis+1}), (i_{s+2},v_{uis+2}), …. , (i_{s+l},v_{uis+l})>

Different techniques can be used to choose the additional items that will be part of the expanded profile. PE techniques are classified according to two criteria:

1. *Local versus global*. Similarly to query expansion techniques discussed before, PE approaches can be either local or global. Local techniques are based on the current recommendation results, that is, on information obtained from the current user profile. On the other hand, global techniques use information available independently of a particular user profile.
2. *Item-based versus user-based*. Item-based techniques choose the items to expand the profile among a given set of items. On the other hand, user-based techniques choose the items among those rated by a given set of users.

2.2.1 Item-global profile expansion

Item-global techniques attempt to find a set of items similar to the items already present in the user profile. As shown in Fig. 2, the expansion takes place before any further computation. Thus, to find similar items, these techniques need to use information globally available in the system.

To speed up the profile expansion step, an implementation may choose to pre-compute a list with the most similar items, for each one of the items.

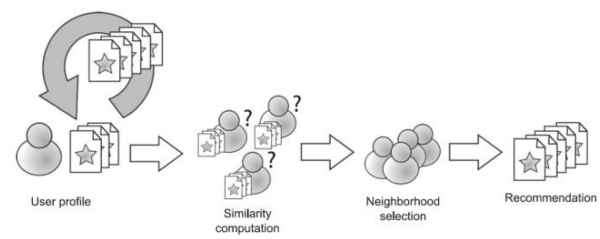

Fig 2.Item-global profile expansion

2.2.2 Item-local profile expansion

Item-local techniques choose the items to expand the profile based on the items recommended to the user, as shown in Fig. 3. They require an initial recommendation, which is only used for profile expansion. Then, a second recommendation, using the expanded profile, is computed and shown to the user.

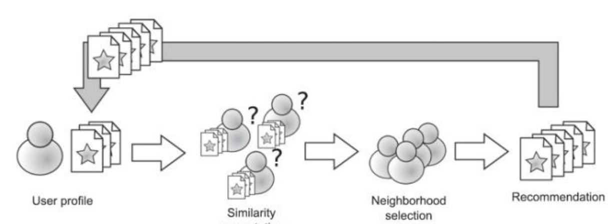

Fig 3.Item-local profile expansion

2.2.3 User-global profile expansion
Similarly to item-global techniques, user-global approaches use information generally available about users to expand the user profile. The idea is to add to the profile items rated by similar users. When collaborative information is used for computing neighbour similarity, this technique is very similar to user-local approaches presented next.

2.2.4 User-local profile expansion
As shown in Fig. 4, user-local techniques are based on the current user neighbours. The items chosen to expand the profile are selected among the items rated by those neighbours.

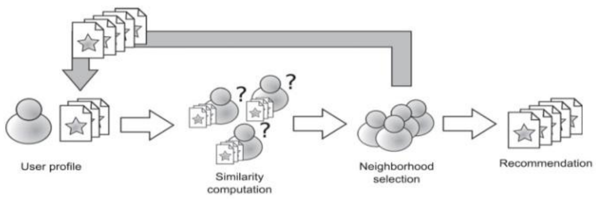

Fig 4.User-local profile expansion

3. Experiments

The experiments are done on the database of the Netflix[8] and Movielens[7] for both the techniques, PIP and PE. In PIP technique's experiments, the Netflix contains over 113,885 ratings from 885 users to 1000 movies and the Movielens contains over 100,000 ratings for 1682 movies, given by 943 users. In PE technique's experiments, the Netflix contains over 100 million ratings from 480,189 users to 17,770 movies and the Movielens contains over 10 million ratings for 10,681 movies, given by 71,567 users of the online movie recommender service. These two datasets are commonly used for Collaborative Filtering evaluation.

3.1 PIP experiments[1]

The first experiment is done with full ratings. It simply compared the recommendation performance of the CF method using the two datasets for five similarity measures: COR, COS, PIP, CPC and SRC. This experiment also included two baseline recommendations, Base1 and Base2, which perform recommendations basedon just the average item ratings and the average user ratings respectively.

The result in Fig. 5 shows that there is not much difference among the five measures when applied to full ratings. COR is showing the best performance for the Movielens and Netflix datasets.

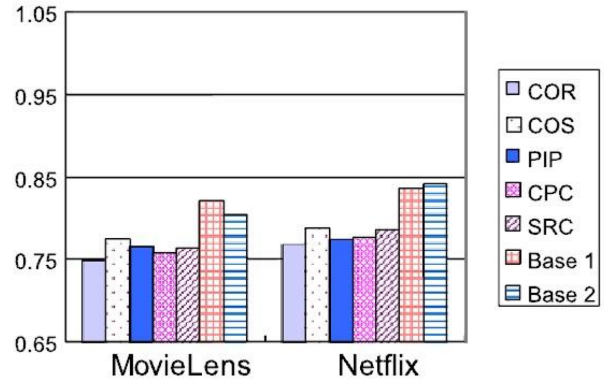

Fig 5.Comparison of the different techniques using full ratings.

The next experiment is done with *artificial cold-starting*. Since the PIP was designed to improve recommendation performance in cold-starting conditions, this experiment tested artificial new user cold-starting conditions by allowing the similarity computation to use only a small number of ratings per each user.

The results, in fig. 6, are very positive for the PIP measure. First, for two datasets, PIP is showing the overall best performance with lowest mean absolute error (MAE) over the range of number of ratings used.

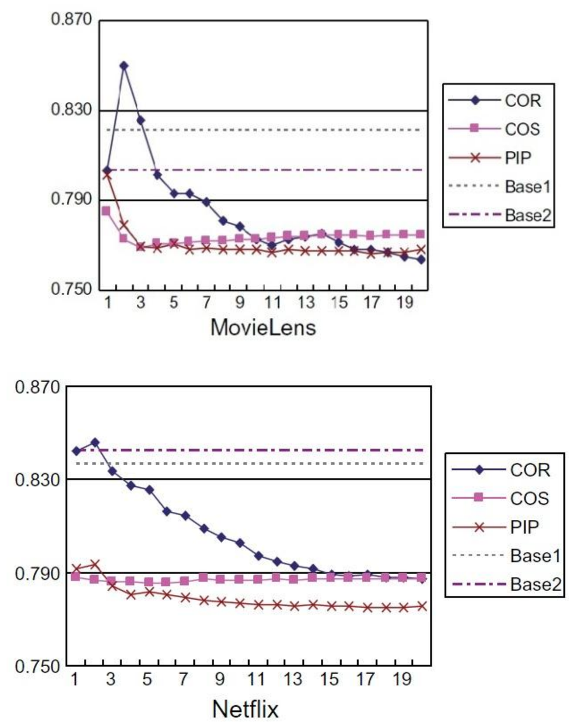

Fig 6.Artificial new user cold-start experiments – X-axis represents the number of ratings used and Y the prediction accuracy measured in MAE.

Based on the previous observation that the COR measure provides better results for the Movielens and Netflix datasets when all ratings are used, and that the COR measure begins to outperform PIP as the number of ratings increases, a hybrid approach combining the two measures was tested. Simply put, the hybrid approach uses PIP when the number of ratings for similarity calculation is smaller than or equal to a given threshold, and applies COR when it is greater than the threshold. Threshold values of 5, 10, 15, 20, and 25 were used in this experiment. The result is shown in Fig. 7.

The result shows superior performance of the hybrid approaches where they are showing better results than PIP and COR in most cases. However, when the percentage of cold-start users is 0, COR is showing as good performance as hybrid approaches of 10 or 15. Conversely, when the percentage is 100, PIP is showing equivalent results as hybrid approaches. Hence, it can be concluded that the hybrid approaches in the example dominatedthe non-hybrid ones except for only non-realistic extreme cases.

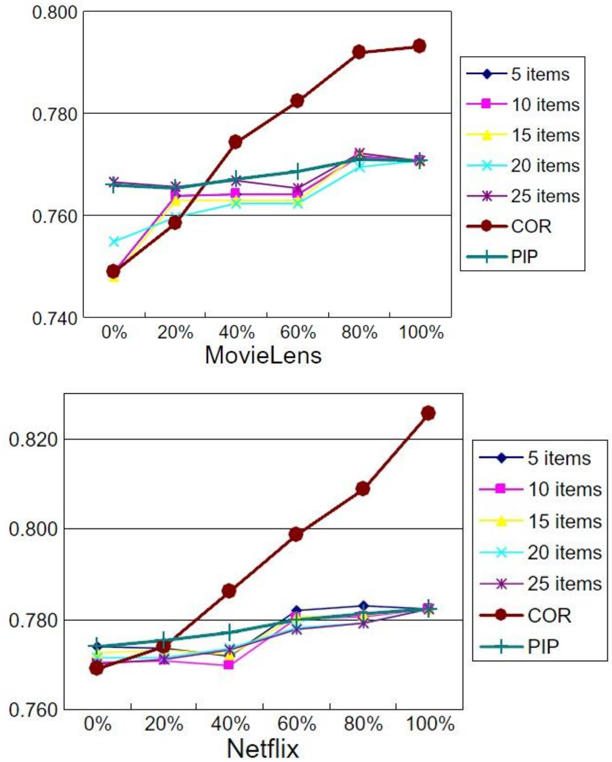

Fig 7.Hybrid recommendation experiments switching from PIP to COR measure with different percentage of cold-start users. X-axis represents the percentage of cold-start users. Y-axis represents the prediction accuracy in MAE.

3.2 PE experiments[2]

The performance of *item-global techniques* is firstly evaluated in particular the item-global method based on cosinesimilarity among item ratings.They have studied the overall performance of the technique, as well as the impact of the parameter l, that determines themaximum number of ratings to add to the user profile. In general, it can be seen that item-global techniques significantly improve the results, especially the precision in the first places of the recommendation list.The authors have studied the impact of item-global techniques on the Mean Average Precision (MAP) that takes intoaccount all recommended item, and not just the top 5 as in the above discussion. Results are shown in Fig. 8. It can be seenthat item-global techniques also improve the overall precision.
The usage of item-global techniquesoffers a huge improvement, offering a quite good precision with just a single rating.

Unlike item-global techniques, the item-local approach has obtained very bad results. In most cases, item-local techniques not only do not improve the results, but also worsen them in many situations.

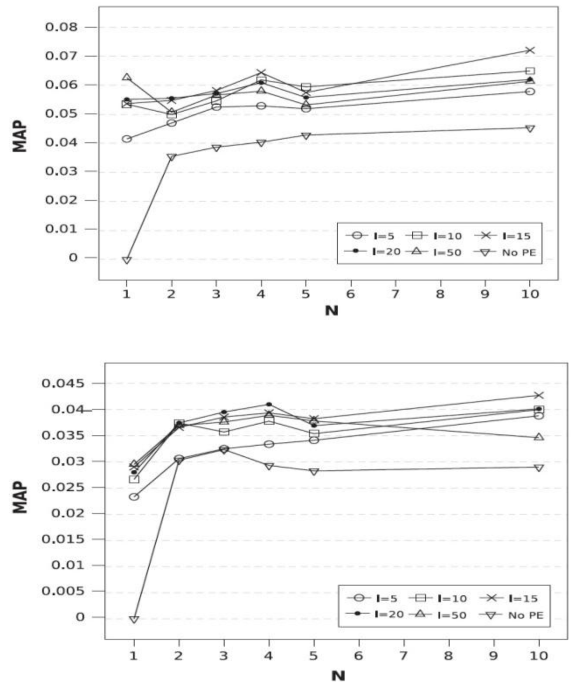

Fig 8.MAP evolution according to N, using item-global techniques. Results computed on Movielens (top) and Netflix (bottom)datasets.

The four user-local methods are evaluated, which are top-rated, most-rated, local item neighbours, anduser-local clustering. The performance of each method is evaluated firstly.User local approaches improve the results, evenconsidering that 10 ratings per user is enough information to compute precise recommendations without profile expansion.The improvement is much better than the one obtained with item-global techniques, in that situation. Even with l = 10, userlocal techniques obtain good results in both Movielens and Netflix. Both user-local clustering and most-rated methods obtainvery good results on the top elements of the recommendation list. That is confirmed if we study the evolution of the Mean Average Precision according to the number of ratings per evaluationuser, N, as can be seen in Fig. 9. The local item neighbours is the best method if average precision is desired.

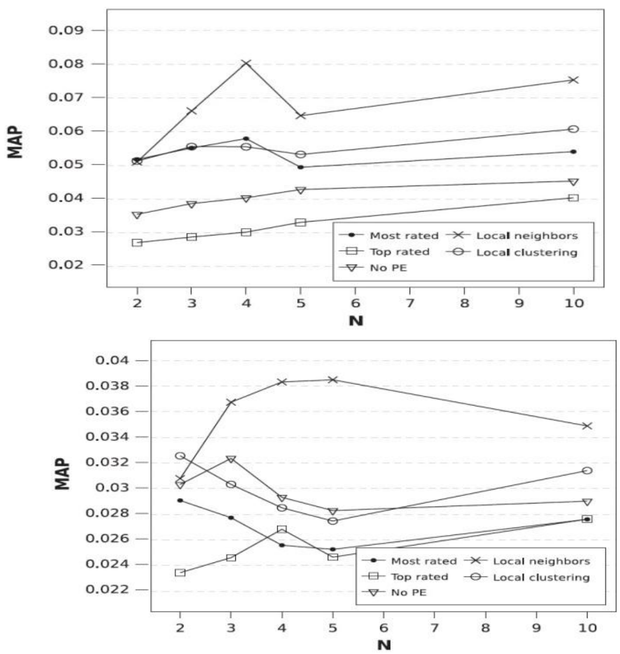

Fig 9.MAP evolution according to N, using different user-local techniques with l = 10. Results computed on Movielens (top) and Netflix (bottom) datasets.

4. Conclusion

From the above two proposed methods, the different experiments are done. The results are reported in the form of graph. From that we can conclude that a hybrid CF approach was also suggestedthat can combine the strengths of PIP and other similarity measures, showing very successful results. In the PE, several novel techniques are proposed, that can be classified in: item-global, item-local and user-local.The main advantage of these techniques is that they are just based on user ratings, so they do not require additional informationsuch as demographic or content-based data. This is an important advantage over many existing approaches. In general, user-local are the best approaches, specially the most rated and user-local clustering methods.

5. References

[1] H. J. Ahn, “A new similarity measure for collaborative filtering toalleviate the new user cold-starting problem”, *Information Sciences*, Elsevier, New Zealand, 178 (2008), pp. 37-51.

[2] V. Formoso, D. Fernandez, et. al, “Using profile expansion techniques to alleviate the new user problem”,*Information Processing and Management*, Elseveir, Spain, 2012.

[3] J. Bobadilla, F. Ortega, et. Al, “A collaborative filtering approach to mitigate the new user cold start problem”, Knowledge-Based Systems, Elseveir, Spain, 26 (2012), pp. 225-238

[4] Heung-Nam Kim, Abdulmotaleb El-Saddik , Geun-Sik Jo, “Collaborative error-reflected models for cold-start recommender systems”, Decision Support Systems, vol. 51 (2011) page no. 519–531, ELSEVIRE, March 2011, Canada.

[5]Mohd Abdul Hameed, RamachandramS.,Omar Al J.,IGCRGA: A Novel Heuristic Approach for Personalization of Cold Start Problem**,** 2011 Fifth Asia Modeling Symposium, IEEE, India , 2011.

[6] G. Adomavicius, A. Tuzhilin, Toward the next generation of recommender systems: a survey of the state-of-the-art and possibleextensions, IEEE Transactions on Knowledge & Data Engineering 17 (2005) 734–749.

[7] MovieLens, MovieLens dataset, http://www.grouplens.org/

[8] Netflix, Netflix movie dataset, http://www.netflixprize.com/