Available online at www.elixirpublishers.com (Elixir International Journal)

Computer Engineering

Elixir Comp. Engg. 86 (2015) 34907-34912

ISSN: 2229-712X

Provide a Method to Determining the Degree of Similarity between Users in Recommender Systems Based on Collaborative Filtering

Mohammad Nazari Farokhi, Sirous Moradi and Mojtaba Salehi

Department of Computer, Khorramabad Branch, Islamic Azad University, Khorramabad, Iran.

ARTICLE INFO

Article history:
Received: 21 July 2015;
Received in revised form: 28 August 2015;
Accepted: 4 September 2015;

Keywords
Recommender systems,
Similarity Measure,
Collaborative Filtering,
Recommendation Accuracy,
Select Features.

ABSTRACT

Collaborative filtering is one of the most widely used techniques in the recommender systems. Determining the similarity between users is the base of collaborative filtering based recommender systems. Obviously, choosing a proper similarity function improves the accuracy of recommendation in recommender systems. In this paper we provided a new similarity measure using features and have compared results with improved similarity measure and pearson traditional similarity measure. the comparison results show that our proposed method not only increases the accuracy of recommender, but also increases the prediction quality.

Introduction

In recent years, the Internet has become more important in the daily life of people. People use the Internet to communicating with others, buying and selling electronic products, searching information and doing other things as well. Nevertheless it is not surprising that the mass of new data created every day and availability of them is beyond the limited capacity of our processing. However, all of us in everyday life, are usually forced to select among the mass of items and objects without adequate experience about the options that we have. Up to day, several methods to solve the problem of information have been proposed. One of these methods is the use of search engines, but yet the engines have not been very successful in personalized search results and often retrieve the same results for all users while may be two users have totally different profiles and consider different aspects of search results. Another practical method is the use of recommender systems. Recommender systems are designed to help of this natural and social process and cope with the information overload [1,2]. These systems are considered as an innovative system that offers useful information, and can be used in various domains [3]. Recommender systems are generally divided into two categories: traditional and modern recommender systems. Traditional recommender systems have been used in the early researches for production of recommender systems and yet have been used widely. Modern recommender systems are often more complex than traditional methods, and their use commercially is still not universal. Traditional recommender systems are generally designed based on three techniques: content-based, collaborative filtering, and hybrid recommender. In the content based recommender systems, first items that the users give them an acceptable rating has to come. Then in the list of entire items, looking for samples that are similar to those rated by users, and recommends the most similar items among them to users [18]. This method requires two types of information; the information about user profiles and information about the content item. Collaborative filtering system is an automated predictive method about the interests of a user that is done by collecting data from many users namely in the form of cooperative [19]. The basic hypothesis of collaborative filtering systems is that those who was agreed on an issue in the past, in the future will be agreed. In the overall classification, collaborative filtering systems are divided in two groups of model based and memory based [13]. Model-based techniques try to create a model of data. Then, calculations are performed only on those models. Models are made using data analysis algorithms and machine learning to base on data taken learn, find good models. Memory based collaborative filtering system is an old method that is implemented easily and is used in many commercial systems. The reason of its name that called "memory-based" is that their algorithms calculate their suggestions directly on user-item matrix which are stored in memory. The advantage of this method is justified results. However, this method has some disadvantages. The main disadvantage is that the rating of Individuals is dependent. Another disadvantage of this approach is that its performance at low data density decreases sharply that this problem can be seen in abundance on the items on the web. Some of the recommender systems in order to reduce the limitations of previous methods use other method that is a combination of content based method and collaborative filtering. Hybrid recommender systems not only increase the predictions efficiency, but also overcome to the problems such as low density or the loss of information. However, the complexity of them is high and implementation is costly. One of the effective factors in the efficiency of a recommender system is similarity measure. Similarity measure by getting data from users, measures the similarity between users. If the accuracy of similarity measures used is low, resulting improper data extraction and thus provide incorrect recommendations to the user. The main foundation of recommender systems is to determine one group of similar users to the target user. To calculate the amount of users' similarities in recommender

Tele:
E-mail addresses: M_kasitman@yahoo.com

systems, we need measures that can be calculated these similarities correctly.

The structure of this paper is presented in the following. Section 2 is the research background. Section 3 presents examine new similarity measure. Results and evaluations are presented in section 4 and section 5 concludes the paper.

Research Background

We can say that the appearance of the term of recommender system is almost in the mid-1990s. At the time, investigators were focused on rating structures. Making intelligent system to assist users in searching, sorting, classifying, sorting and sharing of information has been the first ideas for creation and development of recommender systems. An important aspect of these systems is the achievement of collaborative filtering. At first, the name of recommender systems has been accompanied by collaborative filtering. Since 1990, the first articles about recommender systems were presented by using collaborative filtering [4]. In 1996, interest in collaborative filtering systems led that a workshop on the same subject at the University of Berkeley California started to work, and its result in the 1997 was a special issue named recommender systems [1,5]. The basis of collaborative filtering systems is to determine the similarity between users.

So far, many methods, including Pearson correlation coefficient, cosine similarity measure and adjusted cosine similarity to determine the similarity of users have been used. Billsus et al. [6] have proposed a method for determining the similarity of users based on reducing dimension through singular value decomposition of a primary matrix of the user's ratings in k dimensions. Breese et al. [7] have used the reverse user frequency to calculate the item weight and claimed that have achieved to the desired results. Yu and Jin, later concluded that the reverse user frequency does not provide acceptable results than the Pearson correlation coefficient. Zhang et al. [8] have adopted a matrix conversion method for measuring similarity among users. Because the number of items categories was far less than the number of items, this method caused data sparsity and scalability. Papagelis et al. [9] have proposed a method for reducing dispersion and trust conclusion that have examined the indirect relationships among users in social networks, and valuable sources of data overload to deal with cold start problem and data sparsity.

Yu et al. [10] have adopted mutual data for weighting to the items and in the tests, have applied entropy as a way of weighting. Entropy method somewhat enhances accuracy of recommendation. While mutual data method is done this better than the entropy procedure. But the approach of mutual data in terms of implementation was very complicated and it led to little progress [11]. Ahn has proposed a new similarity measure to address the problem of cold start in memory based collaborative filtering. This method that is known as PIP, has reduced cold start problem of new user, but when the number of data is high, it cannot improve performance. Goldberg et al. have employed advanced collaborative filtering algorithms with the main component analysis [12].

As noted above, one of the similarity measures in the recommender systems is the Pearson correlation coefficient. In a base recommender system that only has the ability to use previous data rate on predictions of possible rate user to items, Pearson correlation coefficient is generally as the main component. For example, if two users A and B have close rates to items 1 and 2, it is likely have close interest rates to item 3, so the items will be the same rate. If we want to predict the possible rate of user A on item 3, should help the rate of user B on this item. To express this factor mathematically, the following formula is used:

$$SimPCC_{a,b} = \frac{\sum_{i=1}^{I_{a,b}}(r_{a,i}-\bar{r_a})\times(r_{b,i}-\bar{r_b})}{\sqrt{\sum_{i=1}^{I_{a,b}}(r_{a,i}-\bar{r_a})^2}\times\sqrt{\sum_{i=1}^{I_{a,b}}(r_{b,i}-\bar{r_b})^2}} \quad (1)$$

In this formula, $r_{a,i}$ represents user rate a on the item i , $\overline{r_a}$ is the average user rates a and independently from user b. $I_{a,b}$ is a set of items that is rating by two a and b users. In a base recommender system which on possible rate user to items predictions, only has the ability to use previous data rate in the system, cosine similarity measure is generally a key component. In fact at this ratio, whatever users assign closer rates to these two items, it means that these two items have high similarity for users and we can say that in near future also will allocate for roughly equal rates. To understand the process of calculating the cosine similarity coefficient, we consider in mind two items as two vectors that will calculate the angle between them. So, the components of these vectors typically will be compared mutually so that angle cosine between two vectors will be calculated using the following formula [14].

$$Sim(i,j) = Cos(\vec{i},\vec{j}) = \frac{\vec{i}.\vec{j}}{\|\vec{i}\|\times\|\vec{j}\|} = \frac{\sum_{k=1}^{N} w_{k,i}w_{k,j}}{\sqrt{\sum_{i=1}^{N} w_{k,i}^2}\sqrt{\sum_{i=1}^{N} w_{k,j}^2}} \quad (2)$$

In this formula $w_{k,j}$ means the k-th component of the item j vector.

Jaccard’s coefficient is the other similarity measure. Jaccard is used in recommender systems, which users rates at which is binary. If the systems have a multi level rates, like [1-5], should be written in a binary rate of (0,1). In binary rating, numbers 1and 0 are respectively user interest and lack of user interest in the item. On this scale, similarity of users is calculated of division share favorite items of both users on the collection of these items:

$$Sim(u,v) = \frac{|I_u| \cap |I_v|}{|I_u| \cup |I_v|} \quad (3)$$

In this regard, I_u and I_v are respectively interest item sets of u and v users [15]. The main problem of Jaccard s method is that, it’s not considers the absolute measures (independent) [7]. The last similarity measure that examine in this paper is improved similarity measure. In this measure, items of relative information usually have used vectors to explain. The formula of item relative similarity is as follows:

$$Sim\ c = (I_a\ ,I_b)\frac{N_C{}^{I_a\cap I_b}}{N_C - \overline{N\ _C{}^{I_a\cup I_b}}} \quad (4)$$

$N_C{}^{I_a\cap I_b}$ is some of the common characteristics among items. I_a , I_b, N_C are a number of items. $\overline{N\ _C{}^{I_a\cup I_b}}$ are the number of ratios that are not owned by I_a item nor by I_b item. At the same item, weighted factors have noted to the relative similarity item and ranking similarities that are defined as follows:

$$\lambda_c = \frac{Sim_c^{\,2}(I_a,\ I_b)}{sim_R^{\,2}(I_a,\ I_b) + sim_c^{\,2}(I_a,\ I_b)} \quad (5)$$

$$\lambda_R = \frac{sim_R^{\,2}(I_a,\ I_b)}{sim_R^{\,2}(I_a,\ I_b) + sim_c^{\,2}(I_a,\ I_b)} \quad (6)$$

$\lambda_{R+}\lambda_c = 1$ applies weighting factors to integrate similarity items in the form of the following equation:

$$Sim(I_a, I_b) = \lambda_c \times sim_c(I_a, I_b) + \lambda_R \times sim_R(I_a, I_b) \quad (7)$$

It is obvious that not for rating new item that its λ_R is zero, the maximum value of λ_c is 1. This method solve scold start problem of collaborative filtering at the same time, but only focuses to discuss new items and ignores new user on a similar subject. At the improved similarity measure, the accuracy of prediction of the new items is not enough valid and we need more research in this field [16].

Examine new similarity measure

Our review network is a two-part network consisted of users and items. At this two -part network, it seems that records contains the file of users characteristics, the file of items characteristics and the file of ratings given by users to items, are provided for the recommender system. According to the recorded data, each user has characteristics that have been received during registration at the website by own. According to the nature of these properties, they can be divided into two categories: numeric and nonnumeric characteristics. Features such as age, height, experience, number of children, etc. are a number of numeric characteristic.

$$S_q^l = \{id \mid f_q^{id} = Str_q^l\}$$
$$S_q^{l'} = \{id' \mid f_q^{id'} = Str_q^{l'}\}$$

These features are suitable for calculation of the similarity of people and different amounts of them can be easily compared. However, features such as occupation, place of residence, nationality, discipline and etc. contains a lot of information on tastes and personality of individuals, but unlike numeric characters, without analysis cannot be used to compare individuals. For example, it's not easy to say that, a lawyer is more alike to an engineers or a writer. So the main challenge in using nonnumeric features to compare them is how to calculate the difference between the different levels of each of them. So the characteristic of each individual is shown below by id in the form of features set of id:

$$F^{id} = F_p^{id} \cup F_q^{id}$$
$$F_p = \{f_p \mid 1 \le p \le M_1\}$$
$$F_q = \{f_q \mid 1 \le q \le M_2\} \quad (8)$$

Where F_p^{id} and F_q^{id}, are respectively numeric feature and nonnumeric feature sets. The number of numeric feature M_1 and non-numeric M_2 is considered. Any numeric feature according to its nature, can easily take areal number. For ease of calculation, we need to normalize numeric feature values before processing. This is done to write numeric feature values in [0,1]. To do this, each numeric feature is normalized as follow:

$$\hat{f}_p = \frac{f_p - f_p^{min}}{f_p^{max} - f_p^{min}} \quad (9)$$

The above relationship, shows the written method of the value of p-th feature in [0,1], at which f_p^{min} and f_p^{max} are respectivelythe minimum and maximum value that p-th feature can receive.

Euclidean space based on numerical characteristics of users

by considering the numerical characteristics of users, can consider Euclidean space for them. In this way, each user according to the recorded values for numerical feature scan be shown by a point in Euclidean space. By considering Euclidean space, Euclidean distance of two different user regard with the available information contained in their profile can be calculated as follows:

$$Eucl_{id}^{id'} = \sqrt{\sum_{p=1}^{M_1} \left(f_p^{id} - f_p^{id'}\right)^2} \quad (10)$$

The final amount of the Euclidean distance in the defined space is between zero and one, and what ever the space is being less for a pair user, the two users are more similar to each other.

The intra-feature distance for different values of each non-numeric feature

For each numerical characteristic, the difference between the different level scan be easily calculated. For example, the user's age difference simply can be calculated by subtracting the values of the features. However, in non-numeric features need to provide a way for different values.

Suppose, id-th user and id'-th user are respectively a teacher and a doctor. It means we have:

$$f_{teacher}^{id} = teacher,$$
$$f_{doctor}^{id'} = doctor. \quad (11)$$

To calculate the difference between the teachers and doctors, all users that their job is teaching and medicine have been searched, and are saved in $S_{teacher}^{id}$ and S_{doctor}^{id} sets respectively. In general, to find the difference between users by f_q^l perform as follow:

For example, if discussed feature is job=q, and $f_{job} \in$ {teacher, carpenter, engineer, doctor} is, $Str_{job}^1 = teacher$, $Str_{job}^2 = carpenter$, $Str_{job}^3 = engineer$ and $Str_{job}^4 = d$octor; in this case:

$$S_q^1 = S_{job}^{teacher} = \{id\ users\ all \mid f_{job}^{id} = Str_{job}^1 = teacher\}. \quad (12)$$

At last, the intra-feature distance for l and l' from f_q feature is calculated by the Mean Euclidean Distance for all users in S_q^l and $S_q^{l'}$ sets:

$$IntraDist(f_q^l, f_q^{l'}) = \frac{\sum_{id \in S_q^l} \sum_{id' \in S_q^{l'}} \left(Eucl(id, id')\right)}{length(S_q^l) \times length(S_q^{l'})} \quad (13)$$

As shown, the above equation only considers anon-numeric feature. For considering all the non-numeric features to calculate the distance between the two users, the following formula is used:

$$Intra_{id}^{id'} = \sqrt{\sum_{q=1}^{M_2} \left(IntraDist(f_q^l, f_q^{l'})\right)^2} \quad (14)$$

By combining the above two equations for numerical features of the Euclidean distance and intra-feature distance for non-numeric features, can achieve equation that by it the difference of two users using their user profile can be identified:

$$Dist_{id}^{id'} = \sqrt{(Eucl_{id}^{id'})^2 + (Intra_{id}^{id'})^2} \quad (15)$$

Selected shared items by two users with id and id' can be shown as below:

$$C_{id}^{id'} = \{I^j | R_{id}^j > 0 \text{ AND} R_{id'}^j > 0\} \quad (16)$$

Also, the items set that are selected at least by one of the above users are displayed as follows:

$$A_{id}^{id'} = \{I^j | R_{id}^j > 0 \; OR R_{id'}^j > 0\} \quad (17)$$

By using two $C_{id}^{id'}$ and $A_{id}^{id'}$ values, selected share of two above users can be calculated:

$$J_{id}^{id'} = \frac{\text{length}(C_{id}^{id'})}{\text{length}(A_{id}^{id'})} \quad (18)$$

$\text{length}(C_{id}^{id'})$ is the number of items that have been rated by two users and $\text{length}(A_{id}^{id'})$ is the number of items that are placed in $A_{id}^{id'}$ set.In this amount that is a number between zero and one, whatever the calculated number is closer to one showing more similarity of two users and vice versa. As previously mentioned, the rank given by users to existing items has been showed by R_{id}^j which shows the rank of id-th user to j-th item. Now suppose the minimum and maximum ranks that can be allocated to the items by users, are one and R^{max} respectively. Thus, we conclude, the minimum and maximum possible difference in votes of two users to an item are zero and R^{max} - 1 respectively .So can expect, two users that have many similarities to each other, having similar votes to their shared items. Hence, the measure of rating error mean is proposed as the last similarity measure in the recommended approach. On this basis, for each analyzed user pair, all share items that has been calculated on $C_{id}^{id'}$ variable are usually examined as follows:

$$\text{Mean Vote Erro} = MVE_{id}^{id'} = \frac{\sum_{j\in C_{id}^{id'}} |R_{id}^j - R_{id'}^j|}{\text{length}(C_{id}^{id'})} \quad (19)$$

For example, suppose that two users jointly give below ranks to 3:

Id= {the first item=2, the second item=5, the third item=1},

Id'= {the first item=3, the second item=2, the third item=4},

In this way, the rating error mean is $\text{length}(C_{id}^{id'}) = 3$:

$$MVE_{id}^{id'} = \frac{|1-4| + |5-2| + |3-2|}{3} = \frac{7}{3}$$

This number is always between zero and $R^{max} - 1$ and whatever is closer to zero showing more similarity of two users; it means that the degree of similarity between two users has an inverse relationship with the $MVE_{id}^{id'}$ rating error mean.

New similarity measure

Proper selecting of a similarity function to determine similarities among users is a critical factor in collaborative filtering systems; because it severely affects the accuracy of the recommendations. Considering all the components introduced in the previous sections, we can achieve similarity measure of users in recommender systems by combining these components:

$$\text{Similarity}(id, id') = \frac{\left(\text{length}(C_{id}^{id'})\right)^2}{\sqrt{(Eucl_{id}^{id'})^2 + (Intra_{id}^{id'})^2} \times \sum_{j\in C_{id}^{id'}} |R_{id}^j - R_{id'}^j| \times \text{length}(A_{id}^{id'})} \quad (20)$$

Experiments

Data Set

Movielens is used for the evaluation tests of the recommendation system of data set. This data base has been a reference in researches of recommendation systems over the past few years. This data set contains over one million registered votes that are produced by 6040 different users. The number of existing movies in this data set is 3900, and user's data are placed only in data set that at least has registered 20 ranks in data sets. This data base is including 4 texts that one file is related to the description and the other three files contain records of data set.

Evaluation Metrics

Recommended similarity measure is being programmed in the MATLAB environment and is implemented on Movielens data collection. For the best comparison in results, Pearson similarity measure and also improved similarity measure are used as competing methods and their records for ease of use in matrices are stored. The most important generated matrices are included as below: The two-part user-item matrix that including 6040 rows and 3900 columns. The users' information matrix is including 6040 rows in the number of network users and has allocated three age, occupation and gender fields to each user Intra -feature distance matrix is used for non-numeric feature of data collection, it means that job feature users is studied. Users' similarity matrix, that is empty at the beginning of the simulation and is completed by implementation of the recommendation measure as well as Pearson similarity measure and improved similarity measure.

Experimental procedures and results

Considering U user with the greatest similarity to a specified user, the average numbers of items that are the same with these users are measured. Moreover, to calculate the number of shared items, first for a U user, select a number of the most similar users with the recommendation measure and also the improved similarity measure and Pearson measure and form two separate set of users. For each of the collection, and for all the same users, all shared items with user would be counted and using them forms a histogram. One of the advantages of users' similarity is that they tend to select similar items, even if the assigned rating to them after experience the item is being decreased. So by counting the number of shared items of user 1 and other similar users and averaging, can compare the accuracy of recommended measure, Pearson measure and improved similarity measure. In the average measure of rating shared items, whatever the average of allocated ranks to shared items is being high, showing that two users have more similarity, and the measure that has detected two users similar has a good performance. Three following Figures represent the histogram of the number of shared items for 50, 100 and 150 similar user.

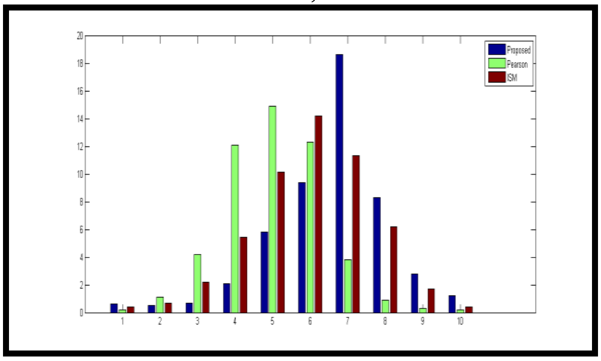

Figure 1. Average number of shared items for the 50 similar user.

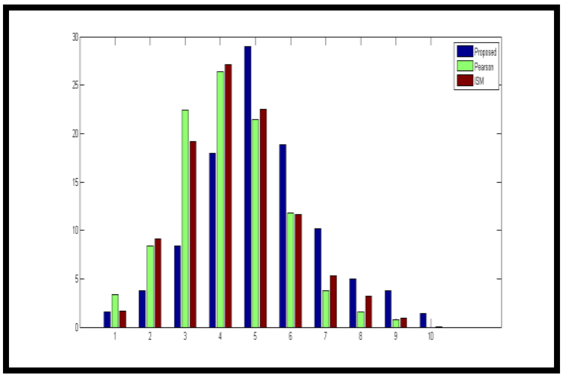

Figure 2. Average number of shared items for the100 similar user.

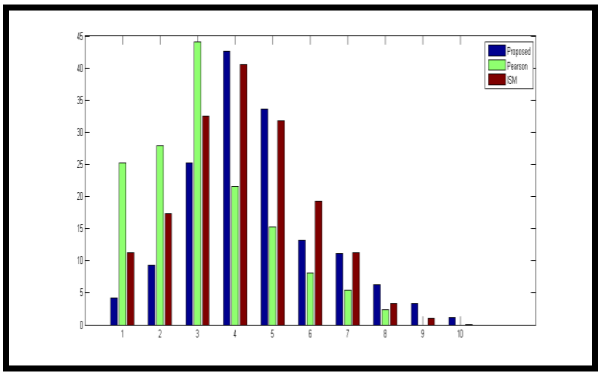

Figure 3. Average number of shared items for the150 similar user.

As you can see, in all three conditions the performance of is better than Pearson and improved similarity measure, and the average number of shared items for recommended measure is higher than Pearson measure and improved similarity measures. The follow figure shows the average number of shared items for similar users. The horizontal axis of this graph shows the number of similar users and the vertical axis is the average number of shared items between user and similar users for all network users.

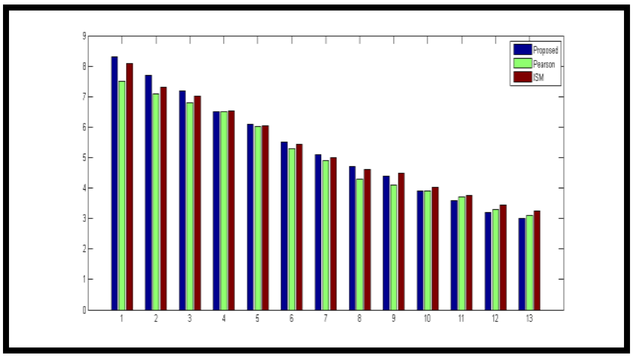

Figure 4. Average number of shared items based on the number of similar users.

As you can see, the recommended measure makes users have, on average, more shared items by other similar users. In other words, the recommended measure selects users as the similar users that are more common with the current user. Finally, at the chart below, average error rating of the shared items by the current user and other similar users is shown. To this, the horizontal axis shows the number of shared users. For each value of the horizontal axis, all the shared items of the current user and the same users as well as the allocated rating to each of these items are considered. Whatever the average error rating is lower, means that similar selected users are more similar to the current user, and this similarity is not as item-selection, but it is in terms of satisfaction of shared items. In other words, low values of the average error rating means that each item that is desired for current user is desired for other similar users and vice versa.

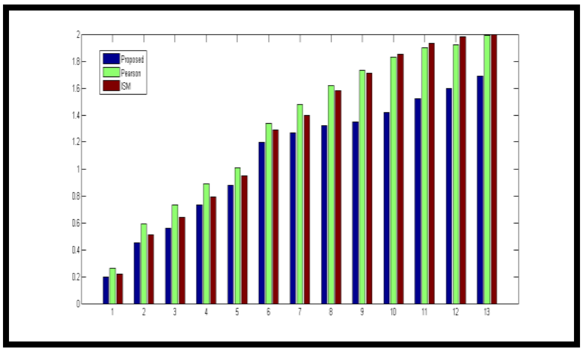

Figure 5. Average error of rating to the number of similar users.

As you can see, the average error rating of recommended similarity measure to Pearson measure and improved similarity measure in all cases is less. This shows that recommended similarity measure selects users as the similar user that has similar preferences to the current user.

Conclusion

In this paper a new similarity measure has been propose and has been implemented on a known data set. For the best comparison of results, Pearson similarity measure and improved similarity measure are used as competing methods. The results showed that the performance of recommended measure from Pearson similarity measure and improved similarity measure is better and our proposed method not only increases the accuracy of recommender, but also increases the prediction quality

References

[1] P. Resnick and H. R. Varian, "Recommender systems," Commun. ACM, vol. 40, pp. 56-58, 1997.

[2] F. Walter, S. Battiston, and F. Schweitzer, "A model of a trust-based recommendation system on a social network," Autonomous Agents and Multi-Agent Systems, vol. 16, pp. 57-74, 2008/02/01 2008.

[3] M.-H. Hsu, "Proposing a charting recommender system for second-language nurses," Expert Syst. Appl., vol. 38, pp. 9281-9286, 2011.
[4] D. H. Park, H. K. Kim, I. Y. Choi, and J. K. Kim, "A literature review and classification of recommender systems research," Expert Systems with Applications, vol. 39, pp. 10059-10072, 9/1/ 2012.
[5] S. Perugini, M. Gonçalves, and E. Fox, "Recommender Systems Research: A Connection-Centric Survey," Journal of Intelligent Information Systems, vol. 23, pp. 107-143, 2004/09/01 2004.
[6] T. Tanimoto, "An elementary mathematical theory of classification and prediction. 1958," International Business Machines Corporation.
[7] H. Liu, Z. Hu, A. Mian, H. Tian, and X. Zhu, "A new user similarity model to improve the accuracy of collaborative filtering," Knowledge-Based Systems, vol. 56, pp. 156-166, 1// 2014.
[8] S.-H. Cha, "Comprehensive survey on distance/similarity measures between probability density functions," City, vol. 1, p. 1, 2007.
[9] E. Vozalis and K. G. Margaritis, "Analysis of recommender systems algorithms," in Proceedings of the 6th Hellenic European Conference on Computer Mathematics and its Applications (HERCMA-2003), Athens, Greece, 2003.
[10] K. Yu, X. Xu, M. Ester, and H.-P. Kriegel, "Feature Weighting and Instance Selection for Collaborative Filtering: An Information-Theoretic Approach," Knowl. Inf. Syst., vol. 5, pp. 201-224, 2003.
[11] R. Jin, J. Y. Chai, and L. Si, "An automatic weighting scheme for collaborative filtering," presented at the Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, Sheffield, United Kingdom, 2004.
[12] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins, "Eigentaste: A Constant Time Collaborative Filtering Algorithm," Inf. Retr., vol. 4, pp. 133-151, 2001.
[13] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," presented at the Proceedings of the 10th international conference on World Wide Web, Hong Kong, Hong Kong, 2001.
[14] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Analysis of recommendation algorithms for e-commerce," in Proceedings of the 2nd ACM conference on Electronic commerce, 2000, pp. 158-167.
[15] J. Bobadilla, F. Ortega, A. Hernando, and J. Bernal, "A collaborative filtering approach to mitigate the new user cold start problem," Knowledge-Based Systems, vol. 26, pp. 225-238, 2012.
[16] W. Yueping and Z. JianGuo, "A collaborative filtering recommendation algorithm based on improved similarity measure method," in Progress in Informatics and Computing (PIC), 2010 IEEE International Conference on, 2010, pp. 246-249
[18] D. Asanov, "Algorithms and methods in recommender systems," Berlin Institute of Technology, Berlin, Germany, 2011.
[19] J. Kleinberg and M. Sandler, "Convergent algorithms for collaborative filtering," in Proceedings of the 4th ACM conference on Electronic commerce, 2003, pp. 1-10.