A Review on Similarity Measurement Methods in Trust-based Recommender Systems

Morteza Ghorbani Moghaddam  
University Putra Malaysia, Malaysia  
morteza.gh.m@gmail.com

Norwati Mustapha  
Aida Mustapha, Nurfadhлина Mohd Sharef  
University Putra Malaysia, Malaysia  
{norwati, aida_m, nurfadhlina}@ upm.edu.my

Anousheh Elahian  
Virtual University of Shiraz, Iran  
anousheh.elahian@gmail.com

Abstract

These days, due to growing the e-commerce sites, access to information about items is easier than past. But because of huge amount of information, we need new filtering techniques to find interested information faster and more accurate. Therefore Recommender Systems (RS) introduced for solving this problem. Although several recommender approaches have proposed, Collaborative Filtering (CF) approaches are the most successful ones. These approaches use historical behaviors of users for making recommendation. Next generation of CF, called Trust-based CF, use social relations and activities for measuring trust between users. One important step in these approaches is measuring the similarity between users, which affect recommendation results. Therefore variety methods for this reason have been proposed. In this paper, we will review and categorize the measurement methods. We will also analyze the methods to identify their characteristics, benefits and drawbacks.

Keywords: measurement methods, Trust-based approaches, recommender systems, Collaborative Filtering, E-commerce.

Introduction

Daily huge amount of information publish through internet. It means that internet is a valuable useful data source. Because of fast growing, following the information flow is not possible physically. Therefore one of the major problems is information overload and for cope with it, filtering of information is essential (Punyavathi & Jyothi, 2013) . One of the most approaches for this reason, is using of Recommender Systems (RS). RS guide users to find interested items by personalized approaches in huge data sources (Bobadilla, Ortega, Hernando, & Gutiérrez, 2013). One of the most important between RS and Information Retrieval (IR) systems are personalization. The results of RS are calculated based on previous interest of users. Therefore they can predict future interest of user even before
themselves.

Based on previous literatures, RS have been categorized to four categories: Collaborative filtering (CF), Content-based (CB), Demographic-based filtering (DF) and hybrid approaches. Differences among the approaches refer to different methods of inferring interest. CF approaches focus on previous behaviors of users. In this approach the most similar users to active user, who the system is predicting his/her interest, are identified and place in a set, called neighbors. Aggregations of their previous interests are used for making prediction or recommendation. CB approaches try to extract the most important features from about interested items. For this reason they need to access to content of items and extract the features from contents. Due to improved text mining approaches, CB approaches have good results about text items but access to content of images and video is not easy. Therefore these approaches have some limitations. The third method is DF. This approach, similar to CF, find similar user to the active user and serve them for making prediction. But the data which are used for measuring similarity and finding similar users are different. In DF, demographic information of users such as age, sex, education … is used for making similarity. However in CF behavior of user are used. In practice, combination of three discussed methods, called hybrid methods, are used. The objective of hybrid methods improving the result of recommendation by using of individual advantages of the previous approaches (Burke, 2002).

Next generation of CF, called social CF, uses web 2 which allows users to interact and share information to each other. One of the most important data which used in social CF is trust. In this paper, trust is defined as one’s belief toward others in providing accurate ratings relative to the preferences of the active user. Previous researches showed that results of trust-based approaches are more reliable than traditional CF approaches (Guo, Zhang, & Thalmann, 2013; Golbeck, 2006).

As discussed before, CF techniques are based on similarity between users by using theirs previous interests. Whatever the measured similarity be more accurate, the result of recommendation is more useful and reliable. For this reason, researchers have tried to introduce variety measurement methods. The similarity methods need different requirements and choosing the suitable method can advance results of the system.

In this paper, we discuss about different proposed similarity methods and will analyze them to highlight advantage and disadvantages of each approach. The rest of this paper is organized as follows: in next section, we discuss about traditional CF and trust-based approaches. Section 3 includes introducing and discussion about different similarity measurement methods. Finally, the last section summarizes the paper.

Recommender Systems Categories

Figure 1 shows categories of CF approaches. These approaches focus on previous behavior of users and use their previous rates about items to predict their interest to not-
rated items. Rate of users, may be explicit (Golbeck, 2006; Massa & Avesani, 2007) or implicit (Jin & Chen, 2010; Moghaddam, Mustapha, Mustapha, Mohd Sharef, & Elahian, 2014). The former refers to the case that the user explicitly defines his/her interest to each item. For example number of starts that user define to each movie can be used as his explicit rate. If the explicit rate is not exists, several researches proposed heuristic approaches for inferring interests. For example Jin and Chen (2010) used number of times that a user listen to music as a parameter to measure his/her interest to the music. Also proposed approach in Moghaddam et al. (2014) used number of assigned tags to each item as a parameter for inferring the rate.

Based on methodology, CF approaches are categorized to model-based and memory-based approaches. Model-based approaches, try to model behavior of users based on their previous activities. It means that previous interests of users to items are used to constructing a predictive model for each user. Base on the model the system can predict interest of users to new items and make recommendation to them. Different researches have used variety model for this reason, which Genetic algorithm, Neural networks, Fuzzy models, Bayesian models and clustering approaches are more used (Bobadilla, et al., 2013; Burke, 2002).

Memory based-approaches, focus on users-items matrix, which is a two dimensional matrix that users are rows and items are columns and content of each cell, is interest of the user to the item. Memory-based approaches include three steps. At the first step, the system measure similarity between users. The second step includes identifying the most similar users to each user based on measured similarities. They will place in a new set, called neighbors, and aggregation and summarization of their interest use for making prediction is last step. Memory-based approach may be user-based or item-based. The former find similarity between users and these similar users suggest the new items to the active user. The latter uses similarity between rated items by the active users and other items to make prediction.

The result of recommender systems may be prediction or recommendation. The first one, need two parameters as input. The user and the item. At this case the system predict interest of defined user to defined item. If the result is recommendation, the system has just one parameter that is the user. At this case it will recommend a list of new items which are more potential to be interested.

![Collaborative Filtering](image)

*Figure 1. Collaborative filtering.*
Method

In this section, we review the most important measurement methods which have proposed. A list of discussed measurement methods have been listed in Table 1. In this section our focus is on user-based approaches. However many of discussed methods may also be usable for item-based approaches by doing some minor changes.

Suppose that $U = \{u_1, u_2, ..., u_N\}$ and $P = \{p_1, p_2, ..., p_M\}$ are set of users and items. Therefore the users-items matrix is:

$$R = (r_{i,j})_{N \times M}, \quad i = 1,2,...,N; j = 1,2,...,M$$

Pearson Correlation Coefficient (PCC) is one of the most important and common measurement methods that is defined as below:

$$Sim(u, v)^{PCC} = \frac{\sum_{p \in I}(r_{u,p} - \bar{r}_u)(r_{v,p} - \bar{r}_v)}{\sqrt{\sum_{p \in I}(r_{u,p} - \bar{r}_u)^2 \sum_{p \in I}(r_{v,p} - \bar{r}_v)^2}}$$

(1)

Where $u$ and $v$ are the users who want measure their similarity. $r_{u,p}$ and $r_{v,p}$ are rate of the users about item $p$. $I$ is set of co-rated items that are rated by both user $u$ and $v$. $\bar{r}_u$ and $\bar{r}_v$ are average rate of users about items in set $I$. PCC gives a value between +1 and −1 inclusive, where 1 is total positive correlation, 0 is no correlation, and −1 is total negative correlation.

PCC have used in several trust-based approaches. For example the proposed approach in Massa & Avesani (2007), called MoleTrust, used breadth-first search method for finding trusted users. They served PCC method for measuring similarity between users’ rates and used average weighted of previous rates for making prediction. TidalTrust (Golbeck, 2006), used PCC for same reason. TidalTrust used a depth-first search for finding more trusted users. Merge (Guo et al. 2013), used PCC measurement method for creating extended items based on trust relations between users. Merge used extended rated items instead of direct rated items for prediction. In this approach, for each user, extended items are the items which are rated by the user explicitly or at least one of his/her direct friends. Merge also calculated predicted rates for items in the extended set. Suppose that item $i$, is member of the extended set for active user $u$. If the user $u$, rated directly to the item $i$ therefore his/her previous rate will save. Otherwise, the average rates of his/her direct friend(s) weighted by importance will save as predicted rate value. For the importance weight, combination of trust, rating similarity and social similarity has been used. For rating similarity, they used PCC similarity method based on directed rates. Also for measuring social similarity, they used Jaccard similarity method, that is discussed in continue. Ray and Mahanti (2010) proposed a trust-based approach that focused on accuracy of prediction. For this reason they reconstructed the trust networks. As weight of trust in trustworthy graph, they used combination of trust and PCC similarity between users.
Table 1

<table>
<thead>
<tr>
<th>Similarity Measurement Methods</th>
<th>Input parameter</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cosine</td>
<td>Value of common rates</td>
<td>[0,1]</td>
</tr>
<tr>
<td>Adjusted cosine</td>
<td>Value of rates</td>
<td>[-1,1]</td>
</tr>
<tr>
<td>Pearson Correlation Coefficient</td>
<td>Values of common rates</td>
<td>[-1,1]</td>
</tr>
<tr>
<td>Constrained Pearson Correlation Coefficient</td>
<td>Values of common rates</td>
<td>[0,1]</td>
</tr>
<tr>
<td>Weighted Pearson Correlation Coefficient</td>
<td>Values of common rates</td>
<td>[-1,1]</td>
</tr>
<tr>
<td>sigmoid function based Pearson Correlation Coefficient</td>
<td>Values of common rates</td>
<td>[-0.6,0.6]</td>
</tr>
<tr>
<td>Confidence-aware Pearson Correlation Coefficient</td>
<td>Values of common rates</td>
<td>[0,1]</td>
</tr>
<tr>
<td>Jaccard</td>
<td>Values of common rates</td>
<td>[0,1]</td>
</tr>
<tr>
<td>Jaccard'</td>
<td>Values of common rates</td>
<td>[0,1]</td>
</tr>
</tbody>
</table>

Although PCC has been used in many trust-based CF approaches, it has some weaknesses that caused other methods be proposed. Constrained Pearson Correlation Coefficient (CPCC) has been proposed in Shardanand (1994) and formulated as below:

\[
Sim(u, v)^{CPCC} = \frac{\sum_{p \in I}(r_{u,p} - r_{med})(r_{v,p} - r_{med})}{\sqrt{\sum_{p \in I}(r_{u,p} - r_{med})^2 \sum_{p \in I}(r_{v,p} - r_{med})^2}}
\]  

(2)

Where \(r_{med}\) is medium rate in scale of ratings. This method measures effect of positive and negative rates. For this reason CPCC supposed that all rates that are bigger than \(r_{med}\) are positive and others are negative rates.

PCC just uses the value rates, but it seems that number of common rates between two users is also important for measuring their similarity. It means that if two users, have rated many common items, therefore the measured similarity will be more reliable. Therefore, for solving the problem, Weighted Pearson Correlation Coefficient (WPCC) has been proposed (Herlocker, Konstan, Borchers, & Riedl, 1999):

\[
sim(u, v)^{WPCC} = \begin{cases} 
\frac{\sum_{p \in I}(r_{u,p} - r_{med})(r_{v,p} - r_{med})}{\sqrt{\sum_{p \in I}(r_{u,p} - r_{med})^2 \sum_{p \in I}(r_{v,p} - r_{med})^2}} & |I| \leq H \\
\frac{\sum_{p \in I}(r_{u,p} - r_{med})(r_{v,p} - r_{med})}{\sqrt{\sum_{p \in I}(r_{u,p} - r_{med})^2 \sum_{p \in I}(r_{v,p} - r_{med})^2}} & otherwise
\end{cases}
\]  

(3)

Where H is penalty parameter that define a threshold value to detect small common sets. If number of common rates between two users is less than H, the measured similarity based on PCC decrease. Otherwise the PCC value is used.

TrustWalker (Jamali & Ester, 2009) is a random walk model which has combined trust-based and item-based collaborative filtering approaches to improve accuracy of predictions and solve cold-start and sparsity problems. In this approach, for predicting the rate of user u
about item i, the system use his direct friends. If they have rated before to the item i, will return their previous rate. Otherwise based on a probabilistic method, they will return one of previous rated items or will ask their direct friend. TrustWalker used a sigmoid function based PCC (SPCC) to avoid favoring the size of I too much:

$$\text{sim}(u, v)^{\text{SPCC}} = \text{sim}(u, v)^{\text{PCC}} \frac{1}{1+\exp(-\frac{|I|}{2})}$$  \hspace{1cm} (4)

If the size of the set of common users is big enough, then the second part of equation 4 would converge to 1, but for small sets of common users, the factor would be 0.6. The number 2 in the denominator of the exponent is because they wanted to have a factor of greater than .9 if the size is greater than 5.

As discussed before Merge (Guo, et al., 2013) is a trust-based approach that uses extended items for measuring the similarity between users. In this approach each item in set of extended items has a rate value and a confidence value. They believed that confidence of rates is important for measuring similarity. Therefore in a new similarity method, based on PCC have been proposed (ibid). This new method, called Confidence-aware Pearson Correlation Coefficient (CoPCC), has been formulated as below:

$$\text{Sim}(u, v)^{\text{CoPCC}} = \frac{\sum_{p \in I} a_{u,p}(r_{u,p} - \bar{r}_u)(r_{v,p} - \bar{r}_v)}{\sqrt{\sum_{p \in I} a_{u,p}(r_{u,p} - \bar{r}_u)^2} \sqrt{\sum_{p \in I} a_{v,p}(r_{v,p} - \bar{r}_v)^2}}$$  \hspace{1cm} (5)

The difference between PCC and CoPCC is confidence factor that is $c_{u,p}$ related to rates of user u about item p. if $c_{u,p}$ is 1, therefore CoPCC result is same as PCC result.

In Cosine similarity method, interests of user u and user v supposed as two vectors in m-dimensional space. At this case m is number of items in I (m=|I|). Therefore similarity between users is computable through angle between two vectors:

$$\text{sim}(x, y)^{\text{COS}} = \frac{\vec{r}_u \cdot \vec{r}_v}{\| \vec{r}_u \| \| \vec{r}_v \|}$$  \hspace{1cm} (6)

Where the magnitude of vector is represented as $\| . \|$. Therefore, based on users’ rates, Cosine similarity method can be formulated as below:

$$\text{sim}(x, y)^{\text{COS}} = \frac{\sum_{p \in I} (r_{u,p}r_{v,p})}{\sqrt{\sum_{p \in I} (r_{u,p}^2) \sum_{p \in I} (r_{v,p}^2)}}$$  \hspace{1cm} (7)

The Cosine method ignores average rate of users and usually is used for the approaches that measure interest of users based on third party data. For example proposed approach (Jin & Chen, 2012), called MWalker, used annotated tags as third party, for profiling users and items. For measuring similarity between users they used Cosine similarity method and
measured their similarity based on users-tags matrix. MWalker used an improved random walk model for making recommendation. They used a heuristic method for measuring interest of users to music. For this reason they used number of times that a user listened to music as his/her rate. Also AgeTrust (Moghaddam et al., 2014) used Cosine to measure similarity between users based on their interests to items. It used number of used tags as interest of the user to the item. In their approach, they discussed about time of friendship and showed that time of friendship can be used for weighting trust between users.

Different users, have different preferences about rating. Some people used to rate items with high values (even for not many interested items), some others prefer to use low rates for rating. The second group doesn’t use high rates even for the most rated items. Cosine similarity method, ignore these preferences and just focus on values of rates. For solving this problem Adjusted Cosine (ACOS) (Ahn, 2008) has proposed based as equation below:

\[
Sim(u, v)^{ACOS} = \frac{\sum_{p \in P} (r_{u,p} - \bar{r}_u)(r_{v,p} - \bar{r}_v)}{\sqrt{\sum_{p \in P} (r_{u,p} - \bar{r}_u)^2 \sum_{p \in P} (r_{v,p} - \bar{r}_v)^2}}
\]

(8)

Where P is set of all items and \(\bar{r}_u\) is average rates of user u about all rated item. In this method, if user u has not rated the item \(p \in P\), therefore \(r_{u,p}\) is zero. Although ACOS is similar to PCC, there is an important difference between them that is related to set of items. In PCC, set of common rated items are used, however in ACOS set of all items are used.

for measuring social similarity between users, Merge (Guo et al, 2013) used Jaccard similarity method (Koutrica & Bercovitz, 2009).

\[
sim(u, v)^{Jaccard} = \frac{|I|}{|I_u \cup I_v|}
\]

(9)

Where \(I_x\) is set of rated items by user x and I is set of common rated items by both users. Jaccard supposed that if two users have many common rating items, therefore their similarity is high. This method ignores rates values and just focuses on size of common rates. The main weakness of this approach is not using rates values. In Merge, as discussed before, for measuring the importance weight of predicted rates in extended items, combination of trust, rating similarity and social similarity has been used. For measuring social similarity, they used Jaccard similarity method. The intuition is that two users are socially close if they share a number of trusted neighbors. In Merge, a trusted neighbor who also shares some social friends is regarded as more important than the user who has no friends in common with the active user. For this reason, the social similarity is defined as the ratio of shared trusted neighbors over all the trusted neighbors, and computed by the Jaccard Index.

A new version of jaccard, called Jaccard’, has been proposed (Liu, Hu, Mian, Tian, & Zhu, 2014). Jaccard’ highlighted effect of the set of common rating items and formulated as
A Review on Similarity Measurement Methods in Trust-based Recommender Systems

below:
\[
sim(u, v)^{\text{jaccard'}} = \frac{|I|}{|I_u| \times |I_v|}
\] (10)

Because of using multiplication operator in denominator, \( \sim(u, v)^{\text{jaccard'}} \leq \sim(u, v)^{\text{jaccard}} \). Reference (ibid) used combination of Jaccard’ and a new heuristic similarity method, called PSS, to measure similarity between users more accurate.

There are several similarity methods that based on our best knowledge, have not used in any trust-based approach. We discuss about these method at continue.

Opposite of jaccard, there is Mean Square Distance (MSD) (Cacheda, Carneiro, & Fernández, 2011) that just focuses on value of rates and formulated as below:
\[
sim(u, v)^{\text{MSD}} = 1 - \frac{\sum_{p \in I} (r_u, p - r_v, p)^2}{|I|}
\] (11)

The main weakness of this approach is ignoring reliability. It means that if the number of common rating items is high, therefore the measure similarity is more reliable. However in this method, just difference between users’ rates is used and \(|I|\) is used just to make average. Also, this method ignores different preferences of users about rating that discussed before.

Spearman’s Rank Correlation (SRC) (Herlocker, Konstan, & Terveen, 2004), is a famous method in several research areas. In this method, instead of using the real value of rates, rank of rates are used. SRC suppose that rates of similar users have same ranks about items. Therefore this method tries to solve the problem of variety preferences. SRC formulated as equation below, that \(dp\) is difference between rank of the users rates about items \(p\).
\[
sim(u, v)^{\text{SRC}} = 1 - \frac{6 \sum_{p \in I} d_p^2}{|I|(|I|^2 - 1)}
\] (12)

All of discussed methods used local rating information of user about items for making similarity. But User Rating Preference (URP) (Liu et al, 2014) is based on global preference of users. This method measures similarity between users based on averages and standard variances of their previous rates. This method formulated as equation below:
\[
sim(u, v)^{\text{URP}} = \frac{1}{1 + \exp(-|\mu_u - \mu_v| |\sigma_u - \sigma_v|)}
\] (13)

It seems that heuristic methods are next generation of similarity methods. These methods are based on combination of effective factors on similarities among users. PIP (Ahn, 2008) used combination of Proximity, Impact and Popularity to propose a new and more accurate method.
\[
sim(u, v)^{\text{PIP}} = \sum_{p \in I} (\text{Proximity}(r_{u,p}, r_{v,p}) \cdot \text{Impact}(r_{u,p}, r_{v,p}) \cdot \text{Popularity}(r_{u,p}, r_{v,p}))
\] (14)
At the first step, this method calculates agreement of rates as Boolean value and also calculates difference between users’ rates by using that. At the next step, (ibid) proposed new formulas for measuring Proximity, Impact and Popularity and combined them as the new factor PIP.

Similar to PIP, PSS (Liu, et al., 2014) used combination of three factor for measuring the similarity. Combination factors in PSS are Proximity, Significance and Singularity. A comparison study has done and different measurement methods have compared to each other (ibid).

Conclusion

Although growing the internet made sharing of information easier than past, Due to the information overload problem, finding the interested information is not easy. To cope with information overload, we need to filter the information. For this reason, we should improve information filtering techniques. One application of information filtering techniques is the recommender systems, which help and guide users to find interested items from a large scale datasets. The literature has broadly categorized recommender systems into four different approaches: Collaborative filtering, Content-based, Demographic filtering and hybrid approaches. Collaborative Filtering is the most successful technology for recommender systems. The technology does not rely on actual content of the items, but instead requires users to indicate preferences, most commonly in the form of ratings. While CF is known for its traditional problems such as cold-start, sparsity and modest accuracy, a trust-based CF has been previously proposed to solve such issues by focusing on trust values among the users.

Trust-based CF techniques are based on similarity between users by using theirs previous interests. Whatever the measured similarity be more accurate, the result of recommendation is more useful and reliable. For this reason, researchers have tried to introduce variety measurement methods. The similarity methods need different requirements and choosing the suitable method can advance results of the system. In this paper, we discussed about different proposed similarity methods and analyzed them to highlight advantage and disadvantages of each approach.

References

A Review on Similarity Measurement Methods in Trust-based Recommender Systems


