NRCF: A Novel Collaborative Filtering Method for Service Recommendation

Huifeng Sun[1], Zibin Zheng[2]
[1]State Key Lab of Networking & Switching Technology
Beijing University of Posts and Telecommunications
Beijing, China
huifengsun.zj@gmail.com, chjl@bupt.edu.cn

Junliang Chen[1], Michael R. Lyu[2]
[2]Dept. of Computer Science & Engineering
The Chinese University of Hong Kong
Hong Kong, China
{zbzheng, lyu}@cse.cuhk.edu.hk

***Abstract*—Since there are many Web services on the Internet, personalized Web service selection and recommendation is very important. In this paper, we present a new similarity measure for Web service similarity computation and propose a normal recovery collaborative filtering (NRCF) method for personalized Web service recommendation.**

***Keywords*-Service recommendation; collaborative filtering; recommender system; QoS;**

I. Introduction

Web service is a software system designed to support interoperable machine-to-machine interaction over a network. Web service discovery that deals with functional properties has been extensively studied [1]. However, Web service discovery cannot differentiate services with identical or similar functionalities. Facing large amount of functionally equivalent services, it is difficult and expensive for users to identify the optimal one, since user-received performance of Web services are highly related to user locations. Designing effective methods for personalized web service selection and recommendation is becoming more and more important in the field of service computing [2].

Quality-of-Service (QoS) is usually defined as a set of non-functional properties, such as round-trip time (RTT), price, failure-rate, etc. QoS are usually considered when making service selection [3]. Values of some Web service QoS properties (e.g., RTT and failure-rate) are influenced by the communication links and usually different from user to user. It will be very useful if we can make personalized QoS value prediction for a user using a small amount of available Web service QoS values. Based on the predicted QoS values, personalized Web service recommendation can be conducted for users. In this paper, we propose a novel collaborative filtering (CF) method, called normal recovery collaborative filtering (NRCF), for personalized Web service recommendation that take advantage of the small amount of available QoS information. By our CF methods, QoS values of Web services for an active user can be automatically predicted employing QoS information from other similar users, who have similar historical QoS experience on the web services to the active user. Existing CF methods are usually designed for product recommendation, and they don't consider characteristics of Web service QoS, but our NRCF do.

II. Normal Recovery Collaborative Filtering

In this section, we present NRCF. Given a Web service recommender system contains M users and N items (Web services), we get an $M \times N$ user-item matrix, in which entry $r_{m,n}$ denotes the QoS value (e.g., RTT values) of Web service n observed by user m. If the entry $r_{m,n}$ is empty, then $r_{m,n} = \varnothing$, denoting that Web service n has never been invoked by user m before.

A. Normal Recovery Similarity Measure

In our method, to measure the similarity between users, we first normalize each row of the original user-item matrix P by the lowest and highest QoS-values of the same row, so that each row has a value range of [0,1]. As a result, we map the original user-item matrix P into row-normal user-item matrix P_{nu}. Assume user u, v co-invoke num items, and vector $\vec{u}$, $\vec{v}$ are the observed-value vector of user u, v in the matrix P_{nu} respectively. $dist(\vec{u},\vec{v})$ calculates the Euclidean distance between $\vec{u}$ and $\vec{v}$ in the num-dimensional vector space, while $dist_{max}$ calculates the maximal Euclidean distance in the num-dimensional vector space that each dimension ranges from 0 to 1; Let $nr_{u,i}$, $nr_{v,i}$ be the value of item i from user u, v in P_{nu} respectively. We propose the formula of NR to measure the similarity between two users u and v as follow:

$$Sim(u,v) = 1 - \frac{dist(\vec{u},\vec{v})}{dist_{max}} = 1 - \frac{\sqrt{\sum_{i \in I}(nr_{u,i} - nr_{v,i})^2}}{\sqrt{\sum_{k=1}^{|I|}(1-0)^2}}$$
$$= 1 - \frac{\sqrt{\sum_{i \in I}\left(\frac{r_{u,i} - r_{umin}}{r_{umax} - r_{umin}} - \frac{r_{v,i} - r_{vmin}}{r_{vmax} - r_{vmin}}\right)^2}}{\sqrt{\sum_{k=1}^{|I|} 1}}$$

978-0-7695-4463-2/11 $26.00 © 2011 IEEE
DOI 10.1109/ICWS.2011.86

i.e.,

$$Sim(u,v) = 1 - \frac{\sqrt{\sum_{i \in I} (\frac{r_{u,i} - r_{umin}}{r_{umax} - r_{umin}} - \frac{r_{v,i} - r_{vmin}}{r_{vmax} - r_{vmin}})^2}}{\sqrt{|I|}}, \tag{1}$$

where $I = I_u \cap I_v$ is the set of items co-invoked by user u and v; $|I|$ is the number of I; $r_{u,i}$ is the value of item i from user u in the original user-item matrix P; r_{umin} and r_{umax} denote the lowest and highest values from user u in P respectively; and r_{vmin} and r_{vmax} denote the lowest and highest value from user v in P respectively. In Eq. (1), $Sim(u, v) \in [0, 1]$, where $Sim(u, v) = 0$ represents that two users are dissimilar and $Sim(u, v) = 1$ indicates that two users are completely similar.

To calculate similarity between two Web services, similarly, based on normalizing the items' QoS values, we map the original user-item matrix P into the column-normal user-item matrix P_{ni}, where each column has a range of [0,1]. The formula of NR to measure the similarity between two items i and j is as follow:

$$Sim(i,j) = 1 - \frac{\sqrt{\sum_{u \in U} (\frac{r_{u,i} - r_{imin}}{r_{imax} - r_{imin}} - \frac{r_{u,j} - r_{jmin}}{r_{jmax} - r_{jmin}})^2}}{\sqrt{|U|}}, \tag{2}$$

where $U = U_i \cap U_j$ is the set of users who invoked both items i and j; $|U|$ is the number of U; $r_{u,i}$ is the value of item i from user u in the original matrix P; and r_{imin}, r_{imax}, r_{jmin}, r_{jmax} denote the lowest value of item i, the highest value of item i, the lowest value of item j, and the highest value of item j in the original matrix P, respectively. In Eq. (2), $Sim(i, j) \in [0, 1]$, where $Sim(i, j) = 0$ represents that two items are completely dissimilar, and $Sim(i, j) = 1$ represents that two items are completely similar.

Specially, the situation when user u (or v) has the same value for all items make Eq. (1) cannot work, and the situation when item i (or j) has the same value for all users make Eq. (2) cannot work. But they exists in theory and don't exist in fact, because the user's QoS values (e.g., RTT) of worldwide Web services and the item's QoS values (e.g., RTT) from worldwide users can't be all the same, respectively. Therefore, we ignore these special situations.

B. Normal Recovery Collaborative Filtering

Based on NR, we propose NRCF. When predicting the unknown QoS value $\hat{ru}_{u,i}$ of user u on item i, NRCF recovers the original scale of user u or item i. In user-based QoS value prediction, we define user-based NRCF as:

$$\hat{ru}_{u,i} = r_{umin} + (r_{umax} - r_{umin}) \frac{\sum_{u' \in U} Sim(u,u') \times nr_{u',i}}{\sum_{u' \in U} Sim(u,u')}, \tag{3}$$

where U denotes the set of similar users to user u who have invoked item i; $nr_{u',i}$ is the value of item i from user u' in the row-normal matrix P_{nu}; r_{umin} and r_{umax} are the lowest and highest values from user u in the original matrix P, respectively; and $Sim(u, u')$ is computed by Eq. (1).

In item-based CF, to predict the unknown QoS value $\hat{ri}_{u,i}$ of user u on item i, we create item-based NRCF as follow:

$$\hat{ri}_{u,i} = r_{imin} + (r_{imax} - r_{imin}) \frac{\sum_{i' \in I} Sim(i,i') \times nr_{u,i'}}{\sum_{i' \in I} Sim(i,i')}, \tag{4}$$

where I denotes the set of similar items to item i which have been invoked by user u; $nr_{u,i'}$ is the value of item i' from user u in the column-normal matrix P_{ni}; r_{imin} and r_{imax} are the lowest and highest values of item i in the original matrix P, respectively; and $Sim(i, i')$ is computed by Eq. (2).

Different datasets have different value characteristics (e.g., different value distributions). To make use of the information from both similar users and similar items, a parameter $\lambda(0 \le \lambda \le 1)$ is employed to determine how much does the prediction rely on user-based NRCF or item-based NRCF. NRCF method makes prediction by the following equation:

$$\hat{r}_{u,i} = \lambda \times \hat{ru}_{u,i} + (1 - \lambda) \times \hat{ri}_{u,i}. \tag{5}$$

When $\lambda = 1$, our method only uses information of similar users to make prediction. When $\lambda = 0$, our method uses information of similar Web services for making missing value prediction. When $0 < \lambda < 1$, our method systematically combines the user-based NRCF method and the item-based NRCF method to fully utilize the information of both similar users and similar Web services.

III. Conclusion

In this paper, we propose NRCF to address the problem of personalized Web service recommendation. NRCF method investigates the characteristics of Web service QoS values and systematically fusing the information of similar users and similar Web services. We will conduct experiments comparing NRCF with other notable competing methods.

Acknowledgment

This work is supported by the National Grand Fundamental Research 973 Program of China (No. 2011CB302506), the National Natural Science Foundation of China (No. 61003067), and the Fundamental Research Funds for the Central Universities (No.2009RC0507).

References

[1] R. M. Sreenath and M. P. Singh, "Agent-based service selection," *Journal on Web Semantics*, vol. 1, no. 3, pp. 261–279, 2003.

[2] L.-J. Zhang, J. Zhang, and H. Cai, "Services computing," in *Springer and Tsinghua University Press*, 2007.

[3] O. Moser, F. Rosenberg, and S. Dustdar, "Non-intrusive monitoring and service adaptation for ws-bpel," in *Proc. 17th Int'l Conf. on World Wide Web (WWW'08)*, 2008, pp. 815–824.