International Journal of Innovative Computing, Information and Control
Volume 11, Number 3, June 2015

 ISSN 1349-4198
pp. 903–919

PERSONALIZED RECOMMENDATION OF E-HEALTH SERVICES BASED ON MUTUAL INFORMATION

Wenyu Zhang, Yuan Zhu*, Yan Wang, Dejian Yu and Shuai Zhang

School of Information
Zhejiang University of Finance and Economics
No. 18, Xueyuan Street, Xiasha, Hangzhou 310018, P. R. China
wyzhang@e.ntu.edu.sg; *Corresponding author: tzzy246@gmail.com
wangyan@zufe.edu.cn; yudejian62@126.com; zs760914@sina.com

Received September 2014; revised January 2015

Abstract. *Information asymmetry has led to selection difficulty for citizens who need to choose appropriate e-government services, especially e-health services in Government-to-Citizen (G2C) areas. This is because of the probability that certain malicious actors (e.g., citizens, health professionals, or institutions) might publish false information to misguide service seekers and make them feel uncertain when making decisions. This paper proposes a novel trust-based recommendation method to address this issue in e-health service recommendation. Service categories are first clustered to make recommendation more specific and accurate. Next, we adopt mutual information in information theory with time difference involved to measure uncertainty, and combine a new concept, hesitation degree, to define and calculate category-based citizen-citizen trust. The EigenTrust method is adopted to propagate local citizen-citizen trust and make it global and systematic. Furthermore, the trustworthiness of a service provider, i.e., citizen-provider trust, is considered. Thus, our final recommended results can be acquired through a two-step filtering process, i.e., collaborative filtering (CF) with citizen-citizen trust and filtering with provider trustworthiness. Experiments are conducted to prove the better performance and practicability of our method.*
Keywords: Government-to-citizen, E-health service, Service recommendation, Collaborative filtering, Mutual information, EigenTrust

1. **Introduction.** With the rapid development of information technology, the concept of electronic government (e-government) services has gradually replaced traditional manual governmental affairs. Government-to-Government (G2G), Government-to-Business (G2B), and Government-to-Citizen (G2C) services are the three main government transaction domains of services comprehensively used by civil servants, business people, citizens, etc. However, for regular citizens, the G2C e-government services are more popular and used. Among such services, electronic employment (e.g., online job seeking), information, education and training, e-health, etc., are common G2C services widely adopted by citizens in their daily life. Nevertheless, as ever-increasing G2C services are published by different government-related and social sectors, service overload becomes a growing prominent problem, especially for e-health services, where improper choices can cause great harm to an individual's health.

E-health is an emerging field at the intersection of medical informatics, public health, and business, and refers to those health services and information delivered or enhanced through the Internet and related technologies that increase efficiency, enhance quality, and encourage a new relationship between patients and health professionals [1]. Meanwhile, e-health services are G2C services related to e-health and published by health-related institutions or health professionals. Citizens or potential patients search for proper e-health services provided by satisfactory and trustable health institutions or professionals with corresponding websites. Frequent concerns in the e-health area include how to find appropriate treatment or therapeutic schedules, whether health professionals or institutions are sufficiently experienced, whether patients should trust professionals for their proposed treatments, and whether treatments are appropriately tailored for the patients. As Minich and Bland [2] emphasized the necessity of personalized lifestyle medicine in public health recommendation, we attempt to focus on health service recommendations in this study.

However, because of information asymmetry and plain introductions attached, citizens seldom understand all aspects of e-health services or their providers. Even worse, increasingly false and irregular health professionals and institutions benefit from such citizen uncertainty, providing them with improper or even harmful e-health services that can cause tremendous damage to the human body. In addition, certain citizens can also broadcast and introduce false information regarding e-health services, which further augments uncertainty. Thus, trust between citizens and between citizens and service providers is an indispensable element when choosing e-health services.

Furthermore, the push technology of current health service platforms, such as WebMD (http://www.webmd.com/) and haodaifu (http://www.haodf.com/), is solely based on user requirement keywords or clicking rates. Thus, the recommending results are not personalized and do not consider different individual characteristics or the trust between citizen users.

Recommendation systems are achieving widespread success and have attracted the attention of researchers, mostly in the field of e-commerce applications [3]; however, they are seldom used in the context of e-government, especially in the area of e-health services that belong to G2C services. These systems can provide personalized and suitable recommended selections for consumers. Content-based and collaborative filtering (CF)-based recommendations are the most popular recommendation methods for recommendation systems.

In order to manage the aforementioned problems that appear in the e-health field, we propose a novel trust-based personalized recommendation method applicable in the context of e-health service recommendation. Mutual information combined with the element of time difference is used in this paper to measure the uncertainty or the passing information between different users. A new introduced factor named hesitation degree is combined with this uncertainty to define our new trust between citizens. In addition, the EigenTrust method [4] is adopted to propagate said local citizen-citizen trust and make it global and systematic, which is helpful for effective CF. Furthermore, the trustworthiness of the health service providers is another considerable factor for the service seekers in this paper for our final recommendation.

The rest of the paper is divided into four sections. Section 2 is the related work of trust, CF, and e-health area. Our proposed trust-based recommendation method for e-health services is presented and elaborated in Section 3. Experiments are explained in Section 4, and Section 5 is the conclusion and some limitations of our work.

2. **Related Work.** CF has been extensively used in recommender systems usually combined with the k-nearest neighbor (KNN) approach. Items are recommended through predicted ratings calculated from the top k similar users/items. Pearson correlation coefficient and cosine measure [5] are the traditional methods frequently used for similarity computation. However, these similarity metric algorithms only focus on the rating similarity between users/items, which are too simple to completely and comprehensively

interpret different aspects of the “similarity” meaning. In addition, actual numbers of the calculated similarity results are much fewer than theoretically forecasted numbers because of data sparsity. Furthermore, merely rating similar users/items cannot represent all collaborative references as the nearest neighbors for the final recommendation prediction, in particular, for the recommendation in different industrial and governmental application domains. Recently, researchers [6-13] have introduced other information and approaches, e.g., demographic information, semantic similarity, and clustering methods, to combine with CF in order to extend the selection range of neighbors/references; meanwhile, to improve efficiency, the recommended results are personalized and made more precise with regard to their respective application fields. In the area of public administration and e-government, literature [3,8,10,12] therein extended the traditional CF into a hybrid CF recommendation method to manage the personalized problems in the field of government transactions combined with citizen traits, mostly in the G2B area. To the best of our knowledge, insufficient attention has been given to the area of G2C, particularly to the personalized recommendation issue of e-health services. Our previous work [14,15] applied CF combined with social network and fuzzy number, respectively, to the fields of manufacturing service recommendation and e-government procurement. In this paper, we attempt to extend our application research into the area of e-health recommendation.

Currently, trust is a popular topic that can act as a decisive factor during the selection of neighbor/reference users in CF for certain application fields. Trust values can be acquired through explicit trust information, such as [16,17], or implicit user relationships in social networks. Nevertheless, few works have provided explicit trust information and there is a probability for certain malicious users to mislead judgment. However, implicit trust can be calculated through existing profiles and finished transactions, which are more real and applicable than explicit trust. In general, there are two classes of implicit trust: 1) subjective trust, whose values are different between different users, with basic characteristics such as asymmetry, context-dependence, and dynamic [18]; 2) objective trust, i.e., trustworthiness, whose value is the same for each user. Cho et al. [19] perceived users who provided information that reflects their actual feelings or opinions as being trustworthy, which is an objective trust. However, such trust is not so adaptable as subjective trust in the real world. Thus, in this paper, we choose subjective trust calculated from extracted implicit trust information to be a reference standard of neighbor selection in our novel CF phase. Different definitions of trust might lead to different trust metric algorithms in various research areas. Researchers [20,21] regarded trust as the capability of recommendation, i.e., the accuracy of prediction of a certain user. Hyoung [22] employed a type of entropy-based consistency between two users as a trust metric in CF methods. However, these metrics did not consider the multiple facets of trust, i.e., people have different degrees of trust toward someone according to different categories of items, which can be regarded as category-based trust. In addition, the rating time difference of a finished transaction between two users can also influence the trust between them toward such type of category, which has always been neglected.

Service recommendation methods in current e-health systems do not consider user-user/doctor trust as the recommending reference. Meanwhile, to the best of our knowledge, there is no research work regarding user-user trust computation models in e-health. In addition, the recommended results are not formed through a personalized filtering process.

The definition of trust in the e-health area is different from the aforementioned trust-related works. A study on trust in e-government reveals that citizen trust in government organizations reduces their perception of risk using e-government services [23], which illustrated the fact that citizen trust has a strong correlation with risk. To some degree, risk can be translated into the uncertainty citizens experience toward certain people or items. In e-health systems, citizen purpose is to find the correct health services and reduce the risk/uncertainty of choosing incorrect services. If citizens have more information on given items, or feel less uncertain/risky about such items, trust is built and enhanced. Sun et al. [24] defined trust as a measure of uncertainty through entropy computation in his information theoretic framework of trust modeling. Inspired by this idea, we attempt to use mutual information acquired from entropy as a measure to calculate uncertainty and further acquire trust between two citizens. However, unlike Sun et al.'s method, we introduce the time difference of trust into our new entropy model to make it more adaptable to the real world. In addition, considering citizen-citizen trust as the only selection standard is not sufficient because the trustworthiness of health institutions or health professionals is also an indispensable standard that should be considered. In the patient-doctor relationship, the trustworthiness of doctors is an important reference standard that should not only include the success rate of previous therapeutic schedules, but also other influential elements, such as affinity. This idea can also be introduced into our e-health service recommendation method to calculate the citizen-provider trust as another important decisive element on final recommendation. Furthermore, given that citizens usually prefer weighting and considering suggestions from others, global trust should also be calculated through the EigenTrust method in this paper to make the recommendation more precise and realistic after the local citizen-citizen trust is acquired through computation.

3. Our Novel Personalized Recommendation Method for E-health Services.

3.1. Overview.

Our proposed method aims to offer citizens with satisfactory e-health services through the reference of trustable citizen recommenders and service providers according to their specific search requirements. Figure 1 shows the system architecture for our approach, which includes four basic modules: (1) e-health service categories matchmaking module. The service categories are found by comparing user e-health requests with service category traits, which are extracted from a clustered e-health service database. Service clustering is formed through the k-medoids [25] algorithm and stored by means of clusters; (2) trust-based rating prediction module. A category-based citizen-citizen local trust is first calculated through time-aware transformed mutual information and hesitation degree. Its concerned data are extracted from a citizen-service execution log, citizen-service rating database, and submitted user information. Predicted ratings based on trustable citizen neighbors are computed next and delivered to the recommended e-health services generation module; (3) category-based citizen-provider trust (trustworthiness) computation module. Provider trustworthiness is acquired and stored through a weighted sum of different influential factors selected by users, such as user satisfaction degree and working age, which are calculated with the data extracted from the health provider profile database and user feedback database; (4) recommended e-health services generation module. The final recommendation results are generated based on a two-step filtering process elaborated in Subsection 3.6, where the trust-based predicted rating and provider trustworthiness value are the two filtering conditions. Detailed calculation methods for these concepts, e.g., trust and hesitation degree, are described next.

3.2. Category-based citizen-citizen (local) trust computation.

To acquire different e-health service categories, various e-health services provided by different health institutions and professionals are clustered through the k-medoids algorithm by comparing the similarities between different e-health services. The similarity values are calculated by

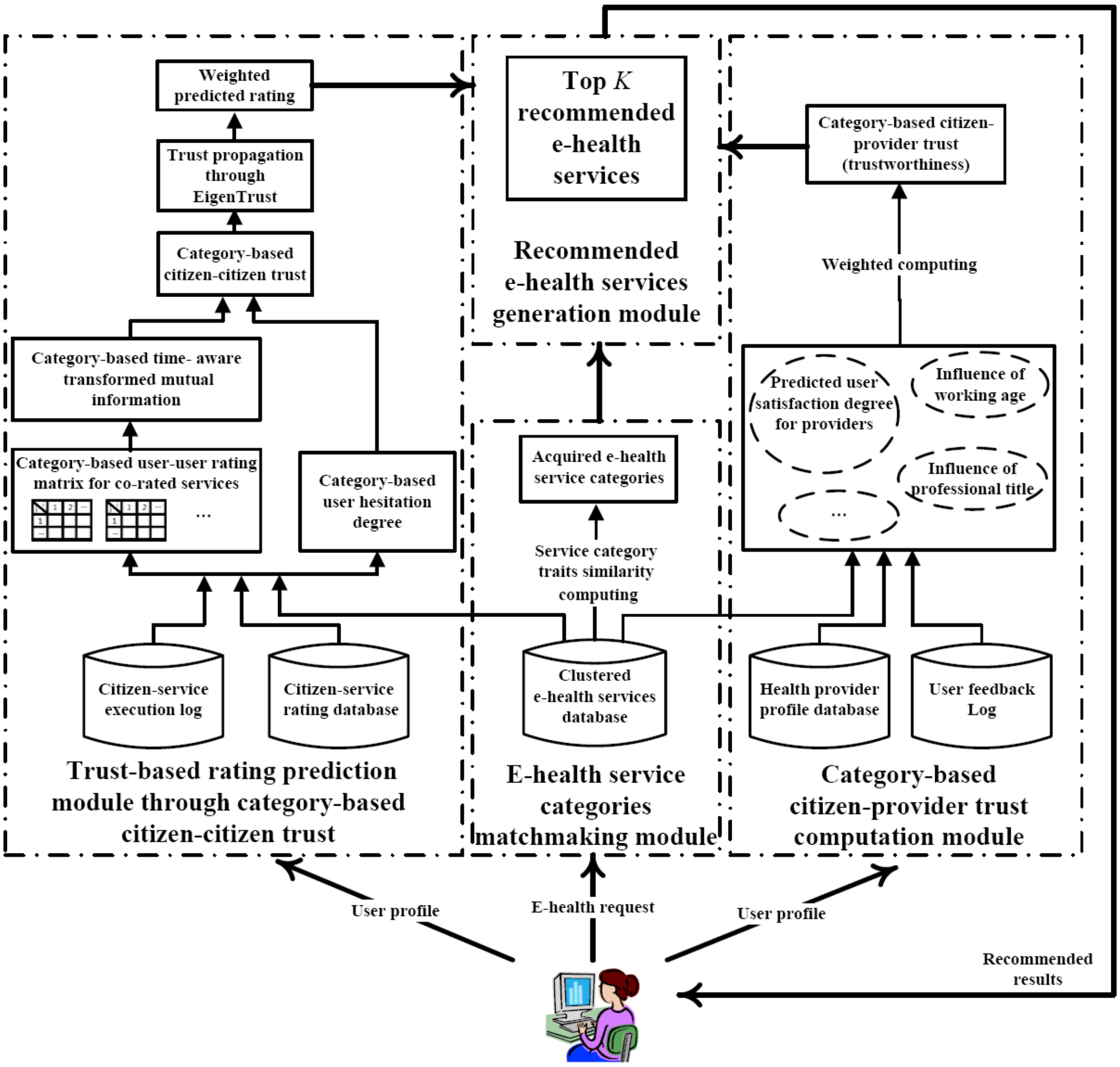


FIGURE 1. System architecture

the cosine measure [5] using the frequencies of important terms, including symptoms, related treatment, etc., extracted from the profiles of e-health services through the popular $tf \times idf$ method [26], which can be referred to our recent work on [27].

Because all service categories are formed through the clustering steps described above, we attempt to calculate the category-based trust between citizens next. Here, in the field of e-health, we decide to use the mutual information in information theory to define our trust. The entropy in the information theory measures the amount of uncertainty or information inherent in the source [28]. Assuming a discrete random variable A can obtain values of $\{a_1, a_2, a_3, \ldots, a_n\}$, its information entropy can be expressed as Equation (1) [22]:

$$H(A) = E\left(\log_2 \frac{1}{P(a_i)}\right) = -\sum_{i=1}^{n} P(a_i)\log_2 P(a_i) \quad s.t. \quad \sum_{i=1}^{n} P(a_i) = 1 \tag{1}$$

where $P(a_i)$ represents the probability of $A = a_i$, and $H(A)$ denotes the entropy of A that illustrates the amount of uncertainty or information from source A.

Then, the (average) mutual information $I(A; B)$ between random variables A and B can be expressed as Equation (2) [29], which can be explained as the acquired information of source A from source B or the decreased uncertainty of A given the condition of B.

$$I(A; B) = H(A) - H(A|B) = H(A) + H(B) - H(A, B) \tag{2}$$

where $H(A, B)$ is the joint entropy that represents the average uncertainty of the entire system, including A and B. Non-negativity and symmetry are the basic mathematical properties of mutual information, i.e., $I(A; B) = I(B; A) \geq 0$.

Thus, the citizen-citizen trust can be regarded as the decreased uncertainty/risk of user A when considering another user B as recommender. In other words, the trust between A and B, $t(A, B)$, is directly proportional to the mutual passing information between them, which can be illustrated as Equation (3).

$$t(A, B) \propto I(A; B) \tag{3}$$

The more the acquired information is, the more the feeling of uncertainty/risk decreases for an information receiver citizen toward the corresponding information source citizen. In addition, we introduce the element of rating time difference into the traditional mutual information measure, which is different from other researchers and more adaptable to the context of e-health.

However, in information theory, the information receiver cannot acquire the complete information sent by the information source because of interference in the communication channel or information source. In the context of e-health citizen communication, we consider the salient interference caused by the receiver. Because the system of e-health is a non-monetary and non-profit public platform, we assume that citizens prefer to provide help and more information to others, whereas receiver citizens might be prone to uncertainty and hesitation when making decisions because health service selection is rather important and involved with body health. Therefore, this type of hesitation becomes interference for information acquisition, which is inversely proportional to the citizen-citizen trust expressed as Equation (4):

$$t(A \to B) \propto \frac{1}{He(A)} \tag{4}$$

where $He(A)$ indicates the hesitation degree of citizen A, which is the information receiver, and $t(A \to B)$ is the trust of citizen A toward citizen B, which is the information source.

Therefore, the final relationship between trust and the influential element described above can be summarized as Equation (5):

$$t(A \to B) \propto \frac{1}{He(A)} \times I(A; B) \tag{5}$$

The detailed calculation of $I(A; B)$ and $He(A)$ is described in the next subsections.

3.2.1. *Time-aware transformed mutual information.* Suppose citizens A and B have co-rated N e-health services that belong to category C_p; meanwhile, the rating of each service wherein varies from 1 to R. Figure 2 shows the matrix of joint distribution for ratings co-rated by citizens A and B of e-health services within category C_p.

The matrix row represents ratings evaluated by citizen A, and the column indicates the ratings evaluated by citizen B. n_{ij} is defined by Hyoung [22] as the number of items with a rating of i by A and a rating of j by B, and $N = \sum_{i=1}^{R} \sum_{j=1}^{R} n_{ij}$; $n_{i\bullet}$ denotes the total number of items with a rating of i by A; $n_{\bullet j}$ denotes the total number of items with a rating of j by B. However, Hyoung's method ignores the influence of rating time

A \ B	1	2	3	...	R	
1	n_{11}	n_{12}	n_{13}	...	n_{1R}	$n_{1\bullet}$
2	n_{21}	n_{22}	n_{23}	...	n_{2R}	$n_{2\bullet}$
3	n_{31}	n_{32}	n_{33}	...	n_{3R}	$n_{3\bullet}$
...	n_{i1}	n_{i2}	n_{i3}	n_{ij}	n_{iR}	$n_{i\bullet}$
R	n_{R1}	n_{R2}	n_{R3}	...	n_{RR}	$n_{R\bullet}$
	$n_{\bullet 1}$	$n_{\bullet 2}$	$n_{\bullet 3}$	$n_{\bullet j}$	$n_{\bullet R}$	

FIGURE 2. The matrix of joint distribution for ratings co-rated by citizens A and B of e-health services within category C_p

difference between ratings for a certain item to trust. Because medical technology and therapeutic methods change rapidly in e-health, citizens prefer the recent suggestions from recommenders. Thus, here we insert rating time difference as an element for mutual information calculation. The computation of n_{ij} is transformed into Equation (6):

$$n'_{ij} = \sum_{n=1}^{n_{ij}} \frac{T}{Max\{|t_{A,n} - t_{B,n}|, T\}} \tag{6}$$

where n_{ij} indicates the number of e-health services with a rating of i by A and a rating of j by B; $t_{A,n}$ and $t_{B,n}$ denote the execution time of the nth e-health service in n_{ij} for A and B, respectively; T is a threshold within which no time difference is considered. Then, our transformed n_{ij}, i.e., new n'_{ij}, is defined as the time-aware numbers of e-health services with a rating of i by A and a rating of j by B.

With the new n'_{ij}, our new values of $n'_{i\bullet}$ and $n'_{\bullet j}$ can also be acquired through Equations (7) and (8), which are different from the original $n_{i\bullet}$ and $n_{\bullet j}$.

$$n'_{i\bullet} = \sum_{j=1}^{R} n'_{ij} \tag{7}$$

$$n'_{\bullet j} = \sum_{i=1}^{R} n'_{ij} \tag{8}$$

Then, the time-aware information entropy of citizens A and B can be expressed as Equations (9) and (10), respectively:

$$H'(A) = -\sum_{i=1}^{R} \frac{n'_{i\bullet}}{N'} \log_2 \frac{n'_{i\bullet}}{N'} \tag{9}$$

$$H'(B) = -\sum_{j=1}^{R} \frac{n'_{\bullet j}}{N'} \log_2 \frac{n'_{\bullet j}}{N'} \tag{10}$$

where $N' = \sum_{i=1}^{R} \sum_{j=1}^{R} n'_{ij}$.

The time-aware joint entropy of A and B is as Equation (11):

$$H'(A,B) = -\sum_{i=1}^{R} \sum_{j=1}^{R} \frac{n'_{ij}}{N'} \log_2 \frac{n'_{ij}}{N'} \tag{11}$$

The transformed time-aware mutual information between citizens A and B for e-health services in category C_p can be expressed as Equation (12):

$$\begin{aligned} I(A;B,C_p) &= H'(A)+H'(B)-H'(A,B) \\ &= -\sum_{i=1}^{R}\frac{n'_{i\bullet}}{N'}\log_2\frac{n'_{i\bullet}}{N'}-\sum_{j=1}^{R}\frac{n'_{\bullet j}}{N'}\log_2\frac{n'_{\bullet j}}{N'}+\sum_{i=1}^{R}\sum_{j=1}^{R}\frac{n'_{ij}}{N'}\log_2\frac{n'_{ij}}{N'} \\ &= \log_2 N'+\frac{1}{N'}\left(\sum_{i=1}^{R}\sum_{j=1}^{R}n'_{ij}\log_2 n'_{ij}-\sum_{i=1}^{R}n'_{i\bullet}\log_2 n'_{i\bullet}-\sum_{j=1}^{R}n'_{\bullet j}\log_2 n'_{\bullet j}\right) \end{aligned} \tag{12}$$

3.2.2. *Hesitation degree.* We assume people provide relatively similar ratings to resources in one category; the large diversity in service ratings within one category is supposed to be an individual's hesitation, i.e., the difficulty level of being influenced by other citizens.

Equation (13) calculates the rating fluctuation (RF) of e-health services in category C_p rated by citizen x, where $Count(x,C_p)$ represents the number of e-health services in category C_p used by citizen x, whereas $r_{x,n}$ is the nth rating; $\bar{r}$ is the average rating of e-health services in category C_p used by the citizen. Obviously, $\sqrt{\frac{\sum_{n=1}^{Count(x,C_p)}(r_{x,n}-\bar{r})^2}{Count(x,C_p)}}$ is the standard deviation for citizen x ratings within category C_p. The multiplier $\frac{1}{r_{\max}}$ is used to normalize the value of the RF degree $RF(x,C_p)$ to be within $[0,1]$, where $r_{\max}$ is the maximum rating.

$$RF(x,C_p)=\frac{1}{r_{\max}}\times\sqrt{\frac{\sum\limits_{n=1}^{Count(x,C_p)}(r_{x,n}-\bar{r})^2}{Count(x,C_p)}} \tag{13}$$

In addition, the uncertainty of citizen x in category C_p can also be expressed as the entropy toward it, shown as Equation (14):

$$H(x,C_p)=-\sum_{i=1}^{R}\frac{num(x,i,C_p)}{Count(x,C_p)}\log_R\frac{num(x,i,C_p)}{Count(x,C_p)} \tag{14}$$

where $num(x,i,C_p)$ represents the number of e-health services with a rating of i within category C_p used by citizen x. Furthermore, we make the logarithm to the base R that can restrict the entropy value to be no larger than one, as referred by Chandrashekhar and Bharat [30].

Thus, the hesitation degree in category C_p for citizen x can be the harmonic mean of the above two elements as expressed in Equation (15), which can synthesize and equalize the influence introduced by such elements.

$$He(x,C_p)=\frac{2\times H(x,C_p)\times RF(x,C_p)}{H(x,C_p)+RF(x,C_p)} \tag{15}$$

However, there is another problem from considering the aforementioned two influential factors, that is, citizens with high mutual information but few co-rated services still have a relatively high trust toward each other, which is not consistent with the actual situation. Thus, we introduce another reference, the Jaccard metric [31] expressed as Equation (16), to alleviate this issue, where $n(A,C_p)$ and $n(B,C_p)$ are the numbers of e-health services within category C_p used separately by citizens A and B.

$$Jaccard(A,B,C_p)=\frac{n(A,C_p)\cap n(B,C_p)}{n(A,C_p)\cup n(B,C_p)} \tag{16}$$

From the above, the final trust value of citizen A toward citizen B in category C_p can be calculated as Equation (17):

$$t(A \to B, C_p) = \frac{k \times I(A; B, C_p) \times Jaccard(A, B, C_p)}{He(A, C_p) + 1} \tag{17}$$

where k is an influential coefficient no larger than 1. In addition, in the case where $He(A, C_p)$ is equal to zero, which means there is no disturbance in information receiving, the trust value is simply $k \times I(A; B, C_p) \times Jaccard(A, B, C_p)$. Considering this, we put an added value "1" in the denominator to make the formula more reasonable and applicable to specific situations.

However, this type of directed trust is just local trust, which is not so accurate. Citizens or potential patients might prefer to consider other partner opinions or trust toward them rather than considering only a single opinion to finally decide trust. Therefore, next we calculate the global trust of the directed trusts indicated above to make them more systematic and exact.

3.3. **Category-based citizen-citizen trust propagation.** The measure of EigenTrust is to compute the converged global trust value. This measure is simple and does not depend on complex parameters or assumptions [32]. Its computation method is widely used in peer-to-peer networks. Here, we attempt to adopt this method [4] to the context of e-health for citizen-citizen trust propagation.

For a certain category C_p, the values of the above calculated trust that represent a citizen i toward another citizen j should be normalized first using Equation (18):

$$t'(i \to j, C_p) = \frac{t(i \to j, C_p)}{\sum_j t(i \to j, C_p)} \tag{18}$$

As in the area of e-health, citizens prefer to gather more information and are inclined to believe or accept suggestions from those partners with symptoms similar to theirs. If two citizens i and j have similar historical diseases and recent symptoms, the active seeker i accepts citizen j's recommendation more readily and trusts him/her more. Furthermore, people of similar age are more likely to share similar symptoms and gain more trust than others. Thus, the trust of citizen i toward j can be transformed into Equation (19).

$$t'(i \to j, C_p) = \alpha \times \sum\nolimits_k t'(i \to k, C_p) \times t'(k \to j, C_p) + (1 - \alpha) \times S_{ij} \tag{19}$$

Here, citizen i attempts to ask his/her direct partners for the credibility degree of citizen j, and then weights and combines the information acquired from all. $t'(i \to k, C_p)$ is the trust value of citizen i towards partner k and $t'(k \to j, C_p)$ is the trust value of partner k towards citizen j. $S_{ij} = \frac{Sim(i,j)}{\sum_j Sim(i,j)}$ is the normalization form that indicates the average health similarity degree, and $Sim(i,j) = \omega_1 \times SimHD(i,j) + \omega_2 \times SimSY(i,j) + \omega_3 \times SimA(i,j)$, where $SimHD(i,j)$, $SimSY(i,j)$, and $SimA(i,j)$ represent the similarity values of historical diseases, recent symptoms, and age between i and j, respectively. The similarity calculation method can be referred to our recent work in [27]. ω_1, ω_2 and ω_3 are the weighting factors with $\sum_{i=1}^{3} \omega_i = 1$ and $\omega_i \geq 0$. α is an adjustment coefficient within the interval $(0, 1)$ that is usually fixed as 0.85 in the EigenTrust measure.

Let T be the matrix of $[t'(i \to j, C_p)]$ and $\overrightarrow{t_{i,C_p}^{(n)}}'$ be the vector of the trust value of citizen i toward others within category C_p. $\overrightarrow{S_i} = [S_{i1}, S_{i2}, \ldots, S_{in}]^T$ can be the vector that contains the average health similarity values of user i toward others. Then, trust propagation can proceed as Equation (20):

$$\overrightarrow{t_{i,C_p}^{(n+1)'}} = \alpha \times T^T \times \overrightarrow{t_{i,C_p}^{(n)'}} + (1-\alpha) \times \overrightarrow{S_i} \tag{20}$$

After n iteration times, when $|t'(i \to j, C_p)^{(n+1)} - t'(i \to j, C_p)^{(n)}| \le \varepsilon$, $\overrightarrow{t_{i,C_p}^{(n+1)'}}$ is converged to a relatively stable value that contains the global trust values of citizen i toward others.

3.4. **Trust-based rating prediction through CF.** We assume that within category C_p, the predicted rating of a certain e-health service HS_i for citizen A, represented as $PR(A, HS_i)$, can be calculated through a weighted sum of the ratings shown as Equation (21) from the selected top M trustable partners ranked by the citizen-citizen trust values of A calculated above in this category while using this service.

$$PR(A, HS_i) = \frac{\sum_{j=1}^{M} R(U_j, HS_i) \times t'(A \to U_j, C_p)}{\sum_{j=1}^{M} t'(A \to U_j, C_p)} \tag{21}$$

where $R(U_j, HS_i)$ is the actual rating of the jth trustable rater U_j of A for the e-health service HS_i, and $t'(A \to U_j, C_p)$ represents the category-based citizen-citizen trust value of A towards U_j.

3.5. **Category-based citizen-doctor trust computation.** In the area of e-health, the trustworthiness of health providers is another important consideration for service selection, besides the recommendation of trustable citizens. In general, the trust of a citizen toward a health institution or a health professional depends mostly on citizen satisfaction. Certainly, there are still other influential elements such as working age, professional title and license. Here we will take working age as an example.

The degree of the satisfaction of provider D, $S(D, C_p)$, for providing services in category C_p can be regarded as the ratio of satisfied health transactions judged by citizens to all executed transactions provided by D in Equation (22):

$$S(D, C_p) = \frac{SatNum(D, C_p)}{TotNum(D, C_p)} \tag{22}$$

where $SatNum(D, C_p)$ and $TotNum(D, C_p)$ indicate the number of satisfied health transactions and all executed transactions offered by provider D, respectively. These values can be acquired through a health provider profile database, including execution logs and user feedback logs.

In addition, the influential degree of working age for provider D, $W(D)$, can be expressed as Equation (23), which is transformed from Huang et al. [33]:

$$W(D) = \frac{4}{\pi} \arctan\left(\frac{year(D)}{Max\{year(D)\}}\right) \tag{23}$$

where $year(D)$ indicates the working year of provider D, and $Max\{year(D)\}$ represents the maximum working year among all providers. The purpose of coefficient $\frac{4}{\pi}$ is to make $W(D)$ be in the interval $(0, 1]$.

Other computation methods are similar to the ones indicated above, and are not listed here for space. The final trustworthiness of health provider D in category C_p, $TW(D, C_p)$, is as shown by Equation (24):

$$TW(D, C_p) = \alpha_1 S(D, C_p) + \sum_{m=2}^{M} \alpha_m factor(m) \tag{24}$$

where α_m is the weighting factor for influential factor $factor(m)$ and M indicates the number of remaining factors; $\sum_{m=1}^{M} \alpha_m = 1$ and $\alpha_m \geq 0$.

3.6. **Recommendation.** As shown in Figure 1, the category-based global citizen-citizen trust and citizen-provider trust (provider trustworthiness) can be calculated and stored in the database once users have submitted their personal information and set required weights in their user profile. Suppose an e-health service seeker A is attempting to search an appropriate e-health service and has submitted his/her service requests; the recommended results are generated as follows.

Step 1: Service category matchmaking. The corresponding service categories are found easily after matching and comparing the searching keywords extracted from service requests with the category features stored in our clustered e-health service database using the text similarity measure introduced in our previous work [27]. The service categories with text similarity values larger than 0.5 can be retrieved.

Step 2: The top M services recommendation. Services are ranked by calculated predicted ratings within the retrieved categories from Step 1; meanwhile, the top M services are selected for further recommendation.

Step 3: Final recommendation. Because the trustworthiness of service providers $TW(D, C_p)$, i.e., the citizen-provider trust, is also an indispensable reference standard for recommendation, we choose the top K e-health services from the top M services through weighting and comparing the trustworthiness value of the providers of M services, as well as the predicted ratings, to make the final recommendation for citizen A. Here, we could select the top K services by ranking the value $\gamma_1 \times \frac{PR(A,HS_i)}{r_{\max}} + \gamma_2 \times TW(SP(HS_i), C_p)$ as the recommendation index, where $SP(HS_i)$ indicates the health provider of service HS_i, $r_{\max}$ is the maximum rating, and γ_1, γ_2 are the weighting factors with $\gamma_1, \gamma_2 \geq 0$, $\gamma_1 + \gamma_2 = 1$.

4. **Experiment and Case Study.**

4.1. **Data sets and evaluation metrics.** In order to prove the better performance of our novel CF method in the area of e-health services, we crawled 2,215 e-health consulting services from an interactive platform (www.120ask.com) provided by 689 doctors, as well as related feedback from 432 corresponding service users.

Because there is no rating information in this e-health consulting platform aside from user feedback (whether the user adopted the suggestion of a given doctor), we assume that the user-service rating value is 10 if the user adopted the services provided by the doctor. Otherwise, the rating value is 1. Furthermore, for the convenience of conduction, k is set to 0.8 because the influence of time-aware mutual information and user hesitation can be relatively high to trust in the e-health area.

The evaluation metrics in our experiment are Mean Absolute Error (MAE) and precision.

(1) MAE

MAE can indicate the precision of predicted rating by measuring the error between predicted ratings and actual ratings, which can be expressed as Equation (25) [31]:

$$MAE = \frac{\sum_{i=1}^{n} |PR_i - AR_i|}{n} \tag{25}$$

where PR_i and AR_i represent the predicted rating and actual rating of the ith user-service rating in the total n ratings, respectively.

(2) Precision

This metric can express the recommendation precision for services in the final recommendation list, which can be calculated by Equation (26) [10]:

$$Precision = \frac{N_{test} \cap N_{topK}}{N_{topK}} \tag{26}$$

where N_{test} and N_{topK} denote the items in the training set and top K recommendation list, respectively.

In the experiment, we randomly choose 20% user-service ratings as the test set, whereas the remaining data form the training set.

4.2. **Method comparison.** We compare our time-aware CF method with traditional CF and CF that uses pure mutual information through MAE and precision metrics.

Figures 3 and 4 show the precision and MAE comparison for different CF methods, respectively. Our novel time-aware CF method performs best compared with traditional CF and CF based on pure mutual information in both recommendation precision and rating prediction accuracy, which indicates the great significance of introducing time-aware numbers into pure mutual information in health trust models.

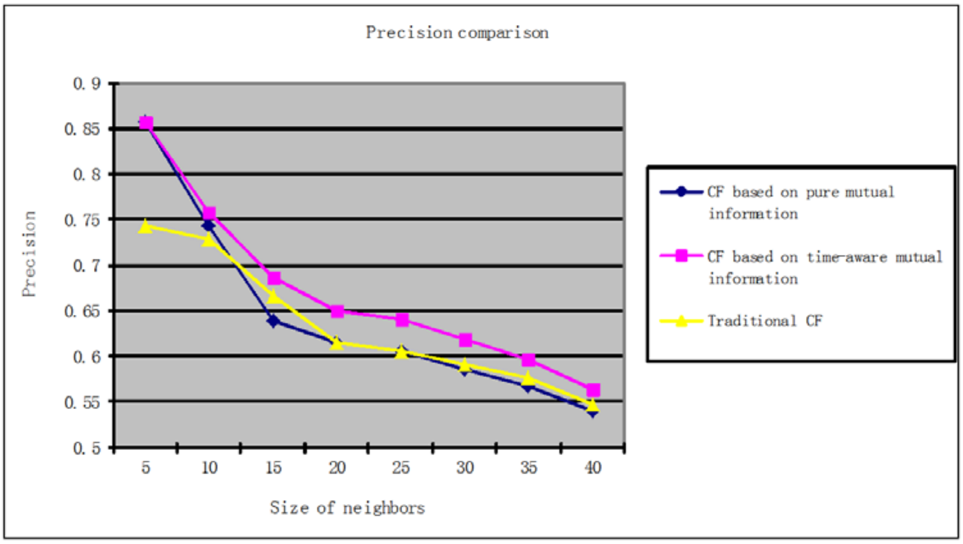


FIGURE 3. Precision comparison among different CF methods

4.3. **Case study.** In this section, we attempt to prove the effectiveness and practicality of our proposed method through a case study. A prototype system for e-health service recommendation based on our novel personalized recommendation algorithm was developed using the ASP.NET language and ExtJs framework for convenience of further testing and verification.

We illustrate the main functions of our prototype system, which contains 2,000 e-health services crawled from the Internet. Before starting the case study on service recommendation, the services are clustered through the k-medoids clustering method introduced in Subsection 3.2, and then classified by health experts into 80 e-health classes. Users complete their personal information and set the values of different weighting factors by answering questions from the "My profile" tab (Figure 5). In addition, users can check their history use in the "E-medical record" tab (Figure 6). The recommendation results are shown in the "Retrieved results" tab (Figure 7) according to user weight preference, personal information, and search categories selected by the users.

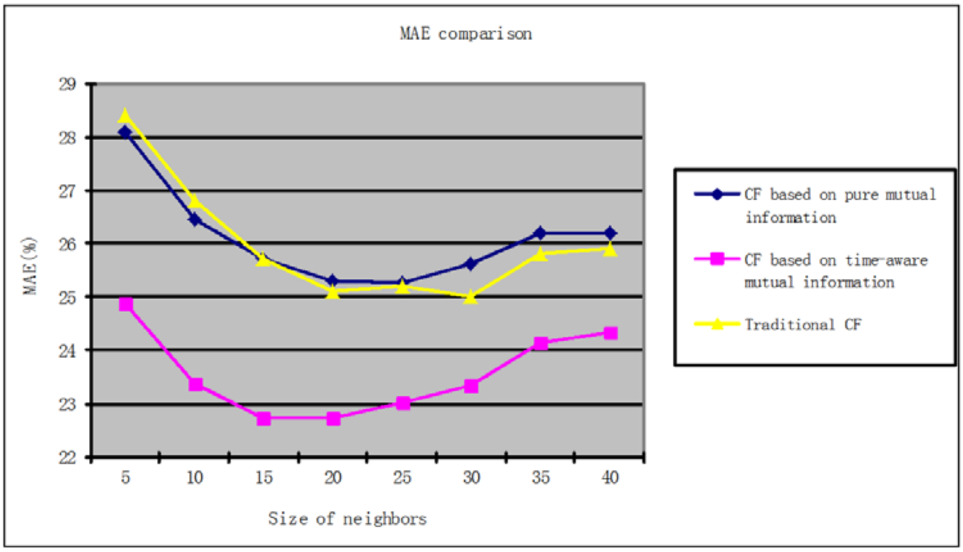


Figure 4. MAE comparison among different CF methods

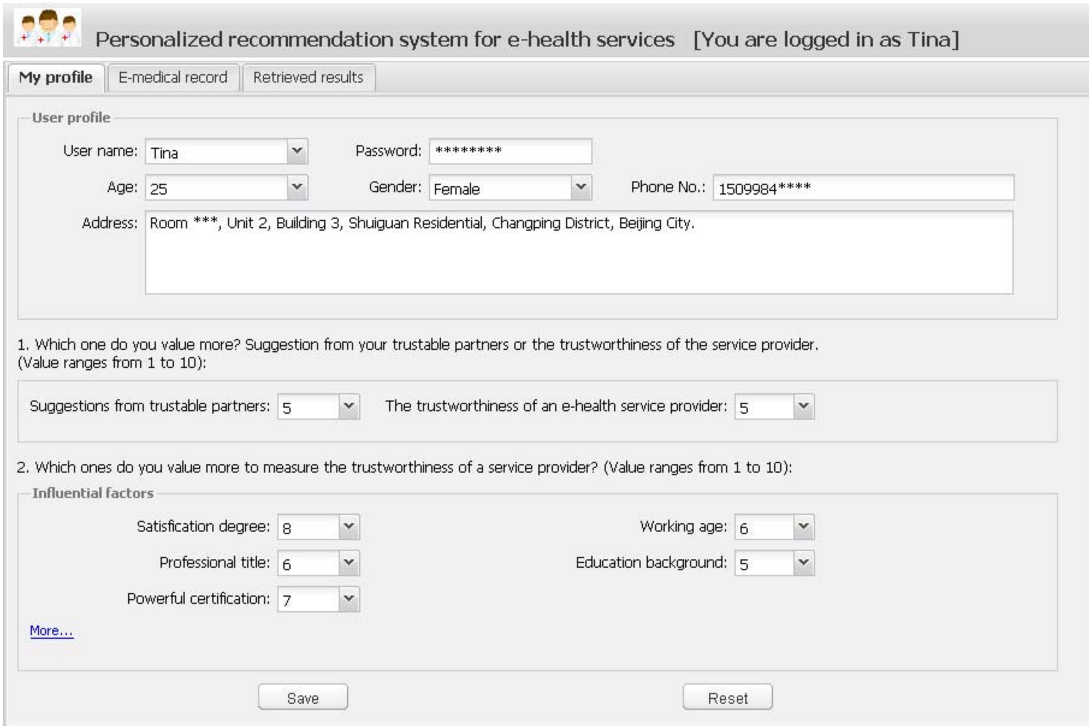


Figure 5. My profile

(1) Completing user profile and weights setting

The system interface for the "My profile" tab (Figure 5) is designed for new users to complete their personal information in the "User profile" tab, as well as weights preferences, or for existing users to modify their personal profiles and weights preferences by clicking the "Reset" button. Personal information includes username, password, age, gender, address, and phone number. The following are two questions for weight setting: (1) arrange the weights for trustable partners and providers to decide which can impact final recommendations more; (2) arrange the weights for influential factors to decide the provider trustworthiness calculation measure personalized for the user. The weights range from one to ten. In this case, user *Tina* has input her personal information and set weights for the recommendation index calculation, which indicates that she values equally the suggestion from trustable partners and provider trustworthiness.

(2) Checking e-medical records

The right window of Figure 6 shows the e-medical record checking interface with which users can check previously used e-health service information, including service date (Date), service name (Used service), provider name (Provider), symptoms (Symptoms), and service rating (Rating). Users can also check information in a selected period by clicking the "Record search" button. In this case, user *Tina* selected the period from *Jan. 1, 2012* to *Dec. 30, 2013*; seven previous records are returned, indicating that seven previous e-health service transactions occurred during this period.

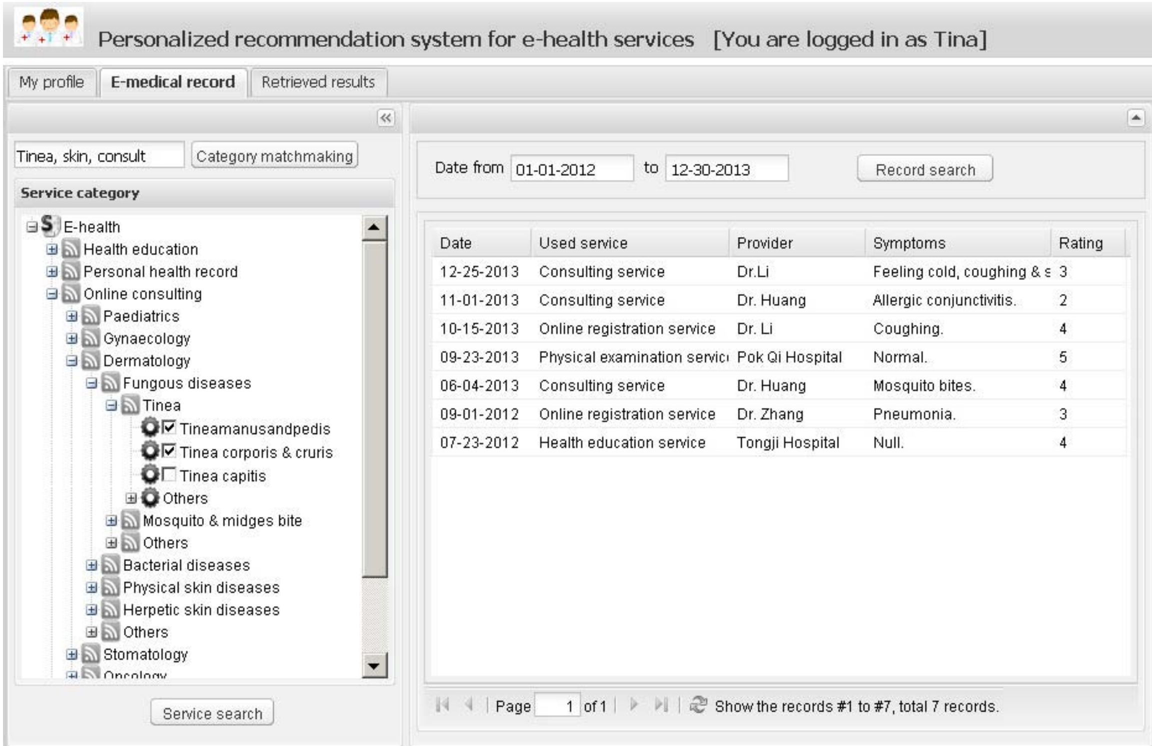


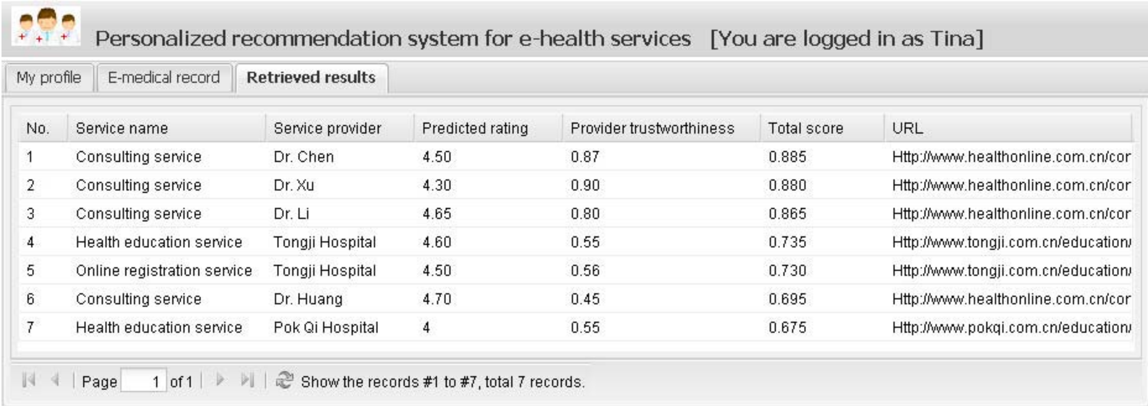
FIGURE 6. E-medical record

(3) E-health service recommendation

After user profile and weight setting have been input, users can search for their required e-health services.

The left window of Figure 6 shows the classes of e-health services named by health experts. The textbox at the top of the window is for class matchmaking. Users can input their health statements, symptoms, etc. into it, and the related classes are found with a selected checkbox (☑) shown next to the classes (see Figure 6) after clicking the "Category matchmaking" button. The checkboxes can be clicked to remove the selections, and different checkboxes can be selected, if users are not satisfied with the machine-matched results. The related categories that correspond to these classes are then found.

In this case, user *Tina* input the request "*Tinea, skin, consult*" into the search bar on the left window of Figure 6. As she clicks the "Category matchmaking" button, two classes of e-health services are found after category matchmaking. It is assumed that user *Tina* can reselect other classes manually if she is not satisfied with the service classes found by the machine. Here, the found service classes are appropriate for user *Tina*'s



No.	Service name	Service provider	Predicted rating	Provider trustworthiness	Total score	URL
1	Consulting service	Dr. Chen	4.50	0.87	0.885	Http://www.healthonline.com.cn/cor
2	Consulting service	Dr. Xu	4.30	0.90	0.880	Http://www.healthonline.com.cn/cor
3	Consulting service	Dr. Li	4.65	0.80	0.865	Http://www.healthonline.com.cn/cor
4	Health education service	Tongji Hospital	4.60	0.55	0.735	Http://www.tongji.com.cn/education/
5	Online registration service	Tongji Hospital	4.50	0.56	0.730	Http://www.tongji.com.cn/education/
6	Consulting service	Dr. Huang	4.70	0.45	0.695	Http://www.healthonline.com.cn/cor
7	Health education service	Pok Qi Hospital	4	0.55	0.675	Http://www.pokqi.com.cn/education/

FIGURE 7. Retrieved results

requirements. The recommended services are returned in the "Retrieved results" tab (see Figure 7). Information about these e-health services contains the service name, service provider, predicted rating calculated from trustable partners, provider trustworthiness, the total score (recommendation index from Subsection 3.6) and URL. For this case, the top seven recommended e-health services shown in Figure 7 satisfy user *Tina*'s service requests.

5. **Conclusions.** In this paper, we proposed a novel trust-based personalized recommendation method for e-health service recommendation to alleviate the adverse impact caused by false information disseminated from certain malicious information sources, including citizens, health professionals, or institutions.

The contributions of our work are as follows:

(1) A new model with new definitions of user-user/doctor trust is proposed to be adapted to the area of e-health using time-aware mutual information and a modified EigenTrust method.

(2) A novel trust-based CF method is used for e-health service recommendation instead of current impersonalized recommendation methods based on keywords or clicking rates.

(3) The trust here is category-based, which can help narrow selections for citizens and make the selections more professional, specific, and accurate.

Two types of trust, i.e., citizen-citizen trust and citizen-provider trust (provider trustworthiness) were considered as the two reference standards that act as filtering conditions in final service recommendation. The mutual information in information theory was used to measure the decreased uncertainty or passing information between two citizens in our calculation of category-based citizen-citizen trust combined with the new proposed concept of hesitation degree. Furthermore, we introduced the element of rating time difference into the computation of said mutual information to make it more reasonable to the real world. In addition, the EigenTrust method was adopted to propagate the local citizen-citizen trust to a global trust. The category-based citizen-provider trust regarded citizen satisfaction as the more important decisive influential factor, and other factors, such as working age, were less important to the trust relationship between citizens and service providers. The final recommended results were returned through two-step filtering, i.e., first filtering through predicted ratings calculated from the top M trustable citizen using CF, and second filtering through provider trustworthiness.

Though the introduction of mutual information can measure the uncertainty in our definition of trust that is more reasonable in the context of e-health, this method relies on the completeness of user-item rating matrix that cannot be applied in situations of extreme data sparsity. Thus, in the near future, we will focus on more influential factors to improve the trust model that can address the issue of extreme data sparsity better.

Acknowledgement. The work has been supported by China National Natural Science Foundation (No. 51375429, No. 71301142), Zhejiang Natural Science Foundation of China (No. LY13E050010), and Zhejiang Science & Technology Plan of China (No. 2014C33084).

REFERENCES

[1] G. Eysenbach, What is e-health? *Journal of Medical Internet Research*, vol.3, no.2, p.e20, 2001.
[2] D. M. Minich and J. S. Bland, Personalized lifestyle medicine: Relevance for nutrition and lifestyle recommendations, *The Scientific World Journal*, doi:10.1155/2013/129841, 2013.
[3] J. Lu, Q. Shambour, Y. S. Xu, Q. Lin and G. Q. Zhang, BizSeeker: A hybrid semantic recommendation system for personalized government-to-business e-services, *Internet Research*, vol.20, no.3, pp.342-365, 2010.
[4] S. D. Kamvar, M. T. Schlosser and H. Garcia-Molina, The eigentrust algorithm for reputation management in P2P networks, *Proc. of the 12th International Conference on World Wide Web*, Budapest, Hungary, pp.640-651, 2003.
[5] L. Mekouar, Y. Iraqi and R. Boutaba, An analysis of peer similarity for recommendations in P2P systems, *Multimedia Tools and Applications*, vol.60, no.2, pp.277-303, 2012.
[6] B. Yapriady and A. L. Uitdenbogerd, Combining demographic data with collaborative filtering for automatic music recommendation, *Proc. of the 9th International Conference, KES*, Melbourne, Australia, pp.201-207, 2005.
[7] Y. H. Xu, X. T. Guo, J. X. Hao, J. Ma, R. Y. K. Lau and W. Xu, Combining social network and semantic concept analysis for personalized academic researcher recommendation, *Decision Support Systems*, vol.54, no.1, pp.564-573, 2012.
[8] Y. Zhu, S. Zhang, Y. Wang, Y. H. Chen, W. Y. Zhang and X. Chen, A social network-based expertise-enhanced collaborative filtering method for e-government service recommendation, *Advances in Information Sciences and Service Sciences*, vol.5, no.10, pp.724-735, 2013.
[9] Q. Shambour and J. Lu, Government-to-business personalized e-services using semantic-enhanced recommender system, *Proc. of the 2nd International Conference, EGOVIS*, Toulouse, France, pp.197-211, 2011.
[10] X. T. Guo and J. Lu, Intelligent e-government services with personalized recommendation techniques, *International Journal of Intelligent Systems*, vol.22, no.5, pp.401-417, 2007.
[11] Q. Shambour and J. Lu, A framework of hybrid recommendation system for government-to-business personalized e-services, *Proc. of the 7th International Conference on Information Technology: New Generations*, Las Vegas, NV, pp.592-597, 2010.
[12] J. Lu, Q. Shambour and G. Zhang, Recommendation technique-based government-to-business personalized e-services, *Proc. of the 28th North American Fuzzy Information Processing Society Annual Conference*, Cincinnati, OH, USA, pp.1-6, 2009.
[13] C. H. Ju and C. H. Xu, A new collaborative recommendation approach based on users clustering using artificial bee colony algorithm, *The Scientific World Journal*, doi:10.1155/2013/869658, 2013.
[14] S. Zhang, C. Y. Xi, Y. Wang, W. Y. Zhang and Y. H. Chen, A new method for e-government procurement using collaborative filtering and Bayesian approach, *The Scientific World Journal*, doi:10.1155/2013/129123, 2013.
[15] W. Y. Zhang, S. Zhang, Y. G. Chen and X. W. Pan, Combining social network and collaborative filtering for personalised manufacturing service recommendation, *International Journal of Production Research*, vol.51, no.22, pp.6702-6719, 2013.
[16] F. E. 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, no.1, pp.57-74, 2008.
[17] F. Zhang, B. Long and G. Feng, A user trust-based collaborative filtering recommendation algorithm, *Proc. of the 11th International Conference, ICICS*, Beijing, China, pp.411-424, 2009.
[18] P. Bedi and R. Sharma, Trust based recommender system using ant colony for trust computation, *Expert Systems with Applications*, vol.39, no.1, pp.1183-1190, 2012.
[19] J. Cho, K. Kwon and Y. Park, Q-rater: A collaborative reputation system based on source credibility theory, *Expert Systems with Applications*, vol.36, no.2, pp.3751-3760, 2009.

[20] J. O'Donovan and B. Smyth, Trust in recommender systems, *Proc. of the 10th International Conference on Intelligent User Interfaces*, San Diego, CA, USA, pp.167-174, 2005.

[21] W. Du and J. L. Chen, The Bayesian network and trust model based movie recommendation system, *Proc. of 2012 International Conference of Intelligence Computation and Evolutionary Computation*, Wuhan, China, pp.797-803, 2012.

[22] D. K. Hyoung, Applying consistency-based trust definition to collaborative filtering, *KSII Transactions on Internet and Information Systems*, vol.3, no.4, pp.366-375, 2009.

[23] A. Alawneh, H. Al-Refai and K. Batiha, Measuring user satisfaction from e-government services: Lessons from Jordan, *Government Information Quarterly*, vol.30, no.3, pp.277-288, 2013.

[24] Y. L. Sun, W. Yu, Z. Han and K. R. Liu, Information theoretic framework of trust modeling and evaluation for ad hoc networks, *Selected Areas in Communications*, vol.24, no.2, pp.305-317, 2006.

[25] T. Velmurugan and T. Santhanam, Computational complexity between k-means and k-medoids clustering algorithms for normal and uniform distributions of data points, *Journal of Computer Science*, vol.6, no.3, pp.363-368, 2010.

[26] D. L. Lee, H. Chuang and K. Seamons, Document ranking and the vector-space model, *IEEE Software*, vol.14, no.2, pp.67-75, 1997.

[27] S. Zhang, Y. Zhu, Y. Wang, W. Y. Zhang and Y. Li, A demographic-based and expertise-enhanced collaborative filtering method for e-government service recommendation, *Journal of Computational Information Systems*, vol.10, no.6, pp.2463-2480, 2014.

[28] C. Piao, J. Zhao and J. Feng, Research on entropy-based collaborative filtering algorithm, *Proc. of the IEEE International Conference on e-Business Engineering*, Hong Kong, China, pp.213-220, 2007.

[29] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.

[30] H. Chandrashekhar and B. Bharat, Personalized recommender system using entropy based collaborative filtering technique, *Journal of Electronic Commerce Research*, vol.12, no.3, pp.214-237, 2011.

[31] Q. Shambour and J. Lu, A hybrid trust-enhanced collaborative filtering recommendation approach for personalized government-to-business e-services, *International Journal of Intelligent Systems*, vol.26, no.9, pp.814-843, 2011.

[32] D. Choi, S. Jin, Y. Lee and Y. Park, Personalized eigentrust with the beta distribution, *Electronics and Telecommunications Research Institute Journal*, vol.32, no.2, pp.348-350, 2010.

[33] Z. X. Huang, X. D. Lu, H. L. Duan and C. H. Zhao, Collaboration-based medical knowledge recommendation, *Artificial Intelligence in Medicine*, vol.55, no.1, pp.13-24, 2012.