A Personalized Trust-Aware Model for Travelogue Discovering

Shilong Zhang, Richong Zhang, Xudong Liu, Hailong Sun
School of Computer Science
Beihang University
Beijing, China
{*zhangsl, zhangrc, liuxd, sunhl*}*@act.buaa.edu.cn*

***Abstract*—Nowadays, many social tourism platforms, such as tripadvisor.com and lvping.com, provide tourists opportunities to share their experiences on tourism destinations, services and sites. The increasing number of these available opinions makes potential travelers impossible easily discovering helpful information from an immense number of lengthy travelogues. Therefore, it is vitally important to develop models and algorithms to assist potential tourists access useful travelogues. This paper proposes a travelogue discovering model that incorporates the implicit trust relations among the social tourism platform, with the aim of discovering the most suitable travelogue for travelers. In addition, the model generates personalized assistance for tourists. The empirical study confirms the effectiveness of our proposed model in discovering helpful travelogues.**

***Keywords*-Travelogue, Recommender System, Trust**

I. Introduction

Recently, travelers are increasingly accustomed to plan their trips by searching the Internet and browsing travel websites. With the popularity of web 2.0 concepts, travel websites, i.e. tripadvisor.com and lvping.com, allows travelers to become registered users, and encourages them to share travelogues, photos, videos and comments. Besides, many travel websites enable users to "follow" other users just like in twitter. A user follows others may because that he agrees with their opinion, enjoys their contents and probably hopes reading more contents generated by them.

Travelogues and other types of reviews can greatly facilitate travelers to know Points of Interest (PoIs). But the increasing availability of the shared contents makes discovering useful information from becomes so overloaded that picking helpful travelogues is a difficult and boring work. Traditional solutions include using search engine and collaborative filtering. Collaborative filtering algorithms collect users' information, look for similar users as neighbors, and recommend items which neighbors prefer. However, in the tourism-related communities, users usually share a small number of experiences. This fact induce the data sparsity and challenges the collaborative filtering in generating useful recommendations for tourists.

Trust-awareness [1] has been proved to be effective to improve accuracy of recommender system. However, in most cases, users of social network do not express how they trust others explicitly, so trust is not quantitative and may lead to mistakenly estimating the degree of the trustiness. Exploiting the similarity among users to build trust relationship and to propagate trust across a network is one of the possible choices to estimate the trust degree between users.

In this paper, we describe and evaluate a novel hybrid travelogue discovering model using both collaborative and trust-awareness to generate helpful travelogues to different users. Specifically, we estimate user similarities from their traveling experiences, and build trust network from their "follow" relationship. We also provide an algorithm for propagating trusts between users and propose strategies to deal with the isolated users that not connected in the propagated trust network.

The remaining of the paper is structured as follows. Section 2 discusses related works. Section 3 provides our model for measuring trustworthiness between users. Section 4 shows the experimental procedure and result. Finally, the conclusions are described in section 5.

II. Related Works

A. Recommender System

Recommender systems (RS) [2] assist users to make choices of items without sufficient personal experience of the alternatives. Recommender systems produce lists of recommendations mainly according to qualities and reviews made by users. Collaborative filtering (CF) [3] is a widely used technique for recommender systems. The collaborative filtering technique works in three steps: collecting opinions expressed by users, computing similarity among users, and recommending items with high rate from similar users. Many websites have applied recommender systems to improve their service. Travel websites or online travel agencies (OTA) encourage users to rate and review hotels, restaurants and tourism attractions. Personal information, such as age, sex, place of residence, income, travel preferences and experience, would be collected for analysis of similarity and other further uses. When someone visits the page of Summer Palace (Yiheyuan) at tripadvisor.com, the page will suggest that travelers who viewed Summer Palace also viewed Great Wall at Mutianyu, Forbidden City (Imperial Palace), Temple of Heaven (Tiantan Park), and Beijing Culture Exchange Center (The Hutong).

978-0-7695-4880-7/12 $26.00 © 2012 IEEE
DOI 10.1109/WI-IAT.2012.160

B. Tourism-related Knowledge Discovering

Besides recommending items according to reviews and ratings, other types of user generated contents, such as travelogues, photos and videos, contain rich knowledge for both travel websites and travelers. Since some photos and other data uploaded by travelers contain geographical information, such information can be mined to identify PoIs and apply a random walk algorithm on the PoIs to provide personalized PoIs recommendations [4]. Hao et al. [5] proposed a probabilistic topic model called Location-Topic (LT) model to discover topics from travelogues and meanwhile represent locations with appropriate topics. Mao et al. [6] explored location relevance classification and ranking techniques, and use a semi-supervised learning algorithm to address the issues of location relevance mining.

C. Trust-Aware Recommender System

Recommender systems suffer some weaknesses: data sparsity, cold start, and attacks made by unreliable users. One solution is to extend recommender systems with trust-awareness. As websites allow users to express their trust to other users, the trust relationships between users abstracted as trust networks or trust matrixes. Massa et al. [7] showed that trust network of users can increase the number of predictable ratings and decrease the error when compared with traditional collaborative filtering recommender systems. Guha et al. [8] made a further research on the propagation of trust and distrust. They concluded four ways of atomic propagation: direct propagation, co-citation, transpose trust and trust coupling, and the procedure of iterative propagation. The trust propagation model ignores similarity between users, while similarity between experiences of users may decrease the accuracy of classification or prediction.

Researches on travel facilities mainly focus on making recommendations of PoIs or itineraries, while recommending user generated contents based on social network from travel websites need relatively more work. We apply trust-aware recommender system to recommend helpful contents about PoIs personally.

III. Trust-Aware Travelogue Discovering Model

In this section we describe methods for mining trust information from travelogue communities.

A. Preparations

Travelogue platforms tend to allow registered users to "follow" other users as twitter.com does. A user follows another because he agrees with his opinions of attractions and like his travelogues. This fact indicates that the one-following-another relationship implies trust between those two users. We use F to denote the matrix of the *following* relationship; F_{ij} denotes whether user i follows user j. The values of F_{ij} are either 1 or 0. If user i follows user j, F_{ij} will be 1, no matter whether user j follows user i or not. In other words, this matrix can be seen as the adjacent matrix of the following network of a travelogue community.

To travelers, the similarity exists mostly among their traveling experience, furthermore, locations or attractions they has visited. We denote the matrix of experience by D and D_{ij} is whether user i has visited location j and has written travelogues about this location. The value of D_{ij} is 1 or 0. Obviously, the more same experience share by two users, the more similar they are to each other.

B. Analyzing users' similarity

General definitions of similarity (Pearson correlation coefficient, for instance) always regard similarity from user i to user j and that from user j to user i as the same, since similarity is considered to be symmetrical. But to analyze the similarity between two users i and j, sometimes i's similarity to j can be different from j's similarity to i. The reason will be discussed in section "Asymmetrical similarity".

1) Symmetrical similarity: Symmetrical similarity refers to traditional meaning of similarity. We take SS to be the matrix of symmetrical similarity, and SS_{ij} to be the similarity between user i and user j, where $SS_{ij} = SS_{ji}$ and the matrix SS is symmetrical. In this paper, the symmetrical similarity SS_{ij} is equal to the symmetrical similarity of D_i and D_j. The value of SS_{ii} is assumed to lie between 0 and 1. If $D_i = D_j$, then $SS_{ij} = 1$, which means that user i's experience is the same to user j's; if $D_i \cap D_j = \emptyset$, then $SS_{ij} = 0$, which means that user i's experience is completely different to user j's.

2) Asymmetrical similarity: In general, the similarity or difference between users by comparing their experience, but they ignore a significant fact in some cases. When considering how traveler j is similar to traveler i, it does not matter how many places j has visited where i has not. Therefore, only how many places both i and j has visited can affect how j is similar to i. We take AS to be the matrix of asymmetrical similarity; AS_{ij} is the similarity of user i to user j. As in the matrix of SS, the value of AS_{ii} is also assumed to lie between 0 and 1. If $D_i \cap D_j = D_j$, then $AS_{ij} = 1$, which means that user i's experience is a part of user j's; if $D_i \cap D_j = \emptyset$, then $AS_{ij} = 0$, which means user j share no same experience with user i. In most cases, $AS_{ij} \neq AS_{ji}$ and $AS_{ij}, AS_{ji} \geq SS_{ij}$.

3) Trust propagation with similarity: The following relationship indicates trusts between users and their followers. So trusts can generate from matrix F. As an example, Figure 1 illustrates the trust propagations. In this figure, we assume there are 4 users in F, where $F_{ij} = F_{kj} = F_{kl} = 1$, so i and k trust j, and k trusts l. Since i is one of j's followers, j may trust i in some degree. Both i and j are j's followers, so i may trust j in some degree, and the same to j. l is followed by one of j's follower k, therefore j may trust l in some degree, and the same to l.

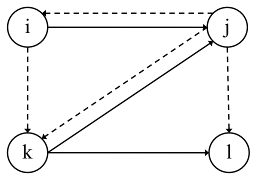


Figure 1. Example of basic elements. The solid lines indicate the *following* relationships, and the dotted lines indicate trust propagation.

Table I
USER i'S, j'S, k'S, l'S EXPERIENCES. A, B, C ARE LOCATIONS VISITED BY THESE USERS.

	i	j	k	l
A	√	√	√	×
B	√	√	×	√
C	√	×	√	√

Above all, trusts can be generated by 4 means: a user trusts whom he follows; a user trusts his followers; users who follow the same are trusted by each other; and users with the same followers are trusted by each other.

Next we need to consider the degree in which a user can trust others who he does not follow. People tend to believe those who has the same background knowledge with themselves, but to distrust those who know little they know. This fact indicats that trust among users is in proportion to the asymmetrical similarity of their experiences.

Assume that user i, j, k, l have visited one or more locations in A, B and C (as shown in Table 1). $D_i = \{A, B, C\}$; $D_j = \{A, B\}$; $D_k = \{A, C\}$; $D_i \cap D_j = D_j = \{A, B\}$; $D_k \cap D_j = \{A\} \subseteq D_j$. Both i and l follow j, but $AS_{kj} < AS_{ij} = 1$. So j would trust i more than k.

In order to simply our model, we take matrix S to contain following relationship and asymmetrical similarity. The values S_{ij} is defined as

$$S_{ij} = \begin{cases} AS_{ij}, F_{ij} = 1 \\ 0, F_{ij} = 0 \end{cases}$$

We take B to be the matrix of trusts generated by following. The values B_{ij} are between 0 and 1. Use function N(Matrix) to normalize matrix by row.

$$B = \alpha_1 F + \alpha_2 S^T + \alpha_3 N\left(SF^T\right) + \alpha_4 N\left(S^T F\right), \\ \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1 \tag{1}$$

We assume that trust can propagate along the follow-network. If user i trusts user j , and user j trusts user k, user i would probably trust k. Since user j's trust to user k is based on user j's experience, user i would not fully trust k if their experience has difference. To solve this problem, we use the symmetrical similarity to measure users' trusts to whom they do not trust directly. B' is define to be a matrix which combines trust and symmetrical similarity where $B'_{ij} = B_{ij}\dot{S}S_{ij}$.

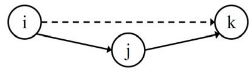


Figure 2. Trust propagations along the follow network.

$$Tp = \sum_{k=1}^{K} \gamma^k B'^{(k)} \tag{2}$$

Tp is a matrix to restore values of trustworthiness computed by propagation and γ is a discount factor to penalize long steps of propagation. We note that the value of K should not be too big, otherwise the result of propagation would be unreliable.

C. Trustworthiness between isolated users

The trust propagation algorithm can be used to effectively compute trustworthiness among some users, while other user cannot be analyzed within K steps of iterative propagation. We call these users "isolated users" in this paper. There are three major factors that can affect trustworthiness of user j to user i: importance of j among all users, similarity between j's trusters and i's, and similarity between j's trustees and i's.

The matrix B can be viewed as a trust network, and a graph walking algorithm can be proposed to predict the importance or trustworthiness of each node of the network. PageRank [9], famous for the Google search engine, is one of the algorithms which computes users' global trustworthiness. In this paper, we choose the PageRank algorithm running on matrix B to predict importance of isolated users. We define I_j to be the importance of user j computed by the PageRank algorithm.

We have measured similarity between single users, and such algorithm can be extended to a set of users. Algorithm 1 shows the pseudocode of the calculation of similarity between user i's and user j's trustors' experiences. $SS(D_i, D_j)$ is a function to compute symmetrical similarity between two arrays. So indicates the Similarity of trustOrs; the values So_{ij} is between 0 and 1. This algorithm can also be used to compute Se, which indicates the Similarity of trustEes.

The trustworthiness of user j to user i is proportional to importance of j and the average similarity of their trustors and trustees.

$$Ti_{ij} = I_j \cdot \frac{So_{ij} + Se_{ij}}{2} \tag{3}$$

Finally, we propose a model of the trustworthiness T_{ij} by combination of Tp and Ti.

$$T_{ij} = \begin{cases} Tp_{ij}, Tp_{ij} > 0 \\ Ti_{ij}, Tp_{ij} = 0 \end{cases} \tag{4}$$

Algorithm 1 The algorithm for computing trustworthiness between isolated users

Require: Input: An $n \times n$ matrix B denotes trusts generated by following, an $n \times m$ matrix denotes relationship between n users and m locations, an integer n denotes number of users, an integer m denotes number of locations, i and j denotes two users.
Output: similarity of i's trustors and j's trustors.

```
for all user, u, ∈ [0, n − 1] do
    for all location, l, ∈ [0, m − 1] do
        if B_ki > 0 then
            D_il = D_il + D_ul
        end if
        if B_uj > 0 then
            D_jl = D_jl + D_ul
        end if
    end for
end for
So_ij = SS(D_i, D_j)
```

IV. Evaluations

The travelogue and review data set was crawled from lvping.com, a famous Chinese travel website running for more than 10 years. We collected 168,489 travelogues and 24,853 reviews of them published before May 2012, and extracted 48,700 registered users. 2653 users of them follows others or are followed by others, and there are 4748 follow-relations among the 2653 users. We choose 2653 users for this study. These users have written 2098 travelogues with reviews, and more than 10,000 reviews to those travelogues. Considering one user may review the same travelogue for more than one time, we select the first review which shows the attitude towards the travelogue by the reviewer. Reviewers' attitude, or sentiment, expresses whether a travelogue is thought "helpful" or "unhelpful". We obtain this information by labeling reviews with emotional words and evaluating the results manually. In this way, 6626 emotional reviews are selected to be used in following steps. The users in the dataset have reviewed the travelogues for 5084 times, and these "helpful" or "unhelpful" opinions will be used to build the classifiers.

The helpfulness of a travelogue is largely decided by following features:

Text features are derived from the travelogue text. We consider the number of words. Some languages like Chinese need word segmentation. We count the proportions of punctuation and emotional words, which include nouns, verbs, adjectives and interjections). Besides, the number of photos in the travelogue is important.

Social features are readers' behaviors and attitude towards travelogues. Typical travelogue websites count how many users have read each travelogue. On typical travelogue websites, registered users can express their opinion with polarity, thank the authors or add the travelogues into their favorite. Users post reviews to each travelogue to express their opinion, so we obtain helpfulness of travelogues to each reviewer by extracting sentiment expressed by each review).

Table II
Prediction Performance of Different Feature Sets

	Precision	F-Measure
Text	.474	.556
Text+Review	.751	.737
Text+Review+Trust	.799	.743

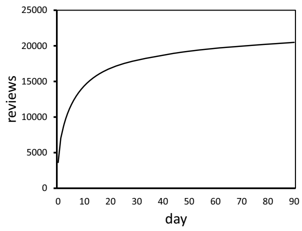


Figure 3. Reviews increase after travelogues published.

Trust-aware features are based on trustworthiness of authors and reviewers. The trustworthiness of authors and reviewers can be measured by elements in matrix. We use trustworthiness of author, the average trustworthiness of reviewers who make positive reviews and the average trustworthiness of reviewers who make negative reviews.

We try several classifier methods including Naïve Bayes, SVM and decision tree. Classifiers are used to classify users' opinion of travelogues into two type: helpful or unhelpful. We count the number of "helpful" items: 1731 of 5084 items are "helpful", and the percentage is 34%. We use FudanNLP[1] to segment Chinese sentences into lists of words. To demonstrate the effectiveness of our trust-aware model, we compare it against two models that use text features only and use text features and sentiment of reviews. In this paper, we only report the results generated by J48 [10] decision tree due to the length limit.

Obviously, users differ about the helpfulness of the same travelogues, so it is almost impossible to predict helpfulness accurately without information or relations between users. Table II shows that the sentiment of reviews can significantly improve the accuracy of prediction, as well as the number and proportion of positive and negative reviews, while trustworthiness seems to work little. Since reviews are not all posted soon after the travelogues are published, we analyze the variation of accuracy over time.

Figure 3 shows the distribution of days of reviews posted after travelogues. The number of reviews posted in 30 days after travelogues were published is 17,982, which is 72.4% of the whole 24,853 reviews. The percentage of reviews

[1] http://code.google.com/p/fudannlp/

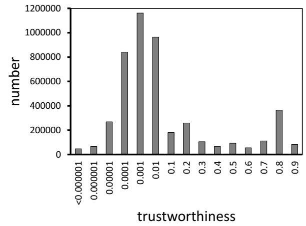
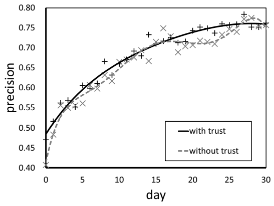
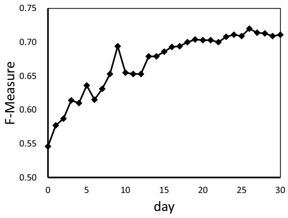


Figure 4. Distribution of trustworthiness values.



(a) Variation of precision of models with and without trustworthiness.



(b) Variation of F-measure.

Figure 5. Variation of result

reaches 80% 68 days after travelogues were published.

Figure 4 shows the distribution of trustworthiness among users. Because there are only 4748 direct trust relations, the most trust are propagated or computed by the importance and similarity. We ignore 2,372,689 trustworthiness equal to 0. The rest of trustworthiness contains 71.8% less than 0.1, partly because these value are calculated by more than one step of propagation.

Finally we consider the variation of accuracy of classifiers within 30 days after travelogues were published. Figure 5a shows the improvement taken by trustworthiness. After travelogues were published, the accuracy of our model can reach above 60% in one week, and 70% in two weeks. The two regression curves show that trustworthiness can not only improve the performance of predicting helpful travelogues, but also make this process smooth or stable. In Figure 5b, we observed that the average accuracy of the precision and the recall also increase rapidly in first 2 weeks. Accuracy of the classifier without trustworthiness fluctuates especially when reviews are not rich enough.

V. CONCLUSION

In this paper we proposed a hybrid recommender system by exploiting trust propagation and collaborative filtering that can predict helpfulness of travelogues better than traditional quality-based classification technique and collaborative filtering model. We use symmetrical and asymmetrical similarity of experiences to describe similarity between users. To overcome the data sparsity, trust propagation method, an integrated algorithm of PageRank and similarities are used to mine hidden trust relationships between users. Our results shows that trust-awareness can increase the accuracy of travelogue helpfulness classification especially for recent posted travelogues, and decrease average error during one month. This model can be easily switched to a cross-site model, because users from different websites can be viewed as isolated users. To demonstrate the applicability of our proposed model, we plan to apply this model to a vertical search engine for the field of tourism to analyze users and to rank travelogues from different websites.

VI. ACKNOWLEDGEMENT

This work was supported by China 863 program (No. 2012AA011203), National Natural Science Foundation of China (No. 61103031), Specialized Research Fund for the Doctoral Program of Higher Education (No. 20111102120016), the State Key Lab for Software Development Environment (No. SKLSDE-2012ZX-12), and the Fundamental Research Funds for the Central Universities (No. YWF-12-LXGY-023) and (No. YWF-12-RHRS-016).

REFERENCES

[1] R. Guha, "Open rating systems," December 03 2002. [Online]. Available: http://citeseer.ist.psu.edu/694373.html;

[2] P. Resnick and H. R. Varian, "Recommender systems," *Communications of the ACM*, vol. 40, pp. 56–58, March 1997.

[3] D. Goldberg, D. Nichols, B. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry," *Communications of the ACM*, vol. 35, no. 12, pp. 61–70, December 1992.

[4] C. Lucchese, R. Perego, F. Silvestri, H. Vahabi, and R. Venturini, "How random walks can help tourism," in *ECIR*, 2012, pp. 195–206.

[5] Q. Hao, R. Cai, C. Wang, R. Xiao, J.-M. Yang, Y. Pang, and L. Z. 0001, "Equip tourists with knowledge mined from travelogues." in *WWW*, M. Rappa, P. Jones, J. Freire, and S. Chakrabarti, Eds. ACM, 2010, pp. 401–410.

[6] M. Ye, R. Xiao, W.-C. Lee, and X. Xie, "On theme location discovery for travelogue services." in *SIGIR*, W.-Y. Ma, J.-Y. Nie, R. A. Baeza-Yates, T.-S. Chua, and W. B. Croft, Eds. ACM, 2011, pp. 465–474.

[7] P. Massa and P. Avesani, "Trust-aware collaborative filtering for recommender systems." in *CoopIS/DOA/ODBASE (1)*, ser. Lecture Notes in Computer Science, R. Meersman and Z. Tari, Eds., vol. 3290. Springer, 2004, pp. 492–508.

[8] R. V. Guha, R. Kumar, P. Raghavan, and A. Tomkins, "Propagation of trust and distrust," in *WWW*, 2004, pp. 403–412.

[9] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web." Stanford InfoLab, Technical Report 1999-66, 1999.

[10] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementation*. San Francisco: Morgan Kaufmann, 2000.