ARTICLE IN PRESS

Neurocomputing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Attraction recommendation: Towards personalized tourism via collective intelligence

Junge Shen, Cheng Deng, Xinbo Gao*

Department of Electronic and Engineering, Xidian University, Xi'an 710071, China

ARTICLE INFO

Article history:
Received 31 March 2015
Received in revised form
31 July 2015
Accepted 9 August 2015
Communicated by L. Shao

Keywords:
Heterogeneous information
Collective intelligence
Attraction recommendation
Personalization

ABSTRACT

Travel recommendation systems can tackle the problem of information overload and recommend proper attractions on the basis of users' preferences. Most existing travel recommendation systems utilized travel history, yet neglected the low frequency of tourism and the flexible styles of attractions in different cities, which will cause the inaccuracy in both collaborative filtering recommendation and content-based recommendation. To deal with this issue, we propose a novel personalized travel recommendation framework by leveraging explicit user interaction and multi-modality travel information. As far as we known, it is the first time that attractions are recommended by user interaction and collective intelligence in a unified framework. Specifically, we first collect heterogeneous travel information by multi-user sharing, which is regarded as collective intelligence to provide reliable references by other travelers. Second, valuable knowledge is mined from collective intelligence in order to filter out the noisy data and make travel information structured. Then, personalized attraction similarity (PAS) model is designed to suggest attractions through fusing heterogeneous information with weighted adaptation and simultaneously considering explicit user interaction. Finally, context information such as the user's location is well adopted to refine the recommendation that may influence the user's choice at a particular moment. Experimental results on pseudo-relevance data and real-world data demonstrate that our method gains promising performance in terms of effectiveness as well as efficiency.

1. Introduction

With attraction the improvement of people's daily life, tourism has become more and more popular. Moreover, with the rapid development of Internet technology and the rise of social media, the users' requirements for the quality of travel service have become more and more high. However, it is challenging that valuable information can be quickly and correctly picked out from massive travel information. For generic, most existing commercial tourism websites show the must-see attractions in the location city on the basis of user ratings to travelers. When travelers wish to explore the places where they have not previously been to, it is rather difficult for them to schedule a perfect trip in view of personal interests and characteristic tourism. Consequently, in order to satisfy users' personal requirements and content-based recommendation can be introduced for personalized attraction recommendation. Meanwhile, when the attractions are so appealing, travelers will take souvenir photos, write comments and make scores. Thus, the heterogeneous information uploaded by travelers can be considered as their travel preferences and experiences, namely *collective intelligence*. Moreover, considering massive travel information, an intelligent website or system should take advantage of collective intelligence for content-based personalized attraction recommendation. Therefore, it is more desirable to mine knowledge from heterogeneous collective intelligence and combine personalization in the coming intelligent travel recommendation system.

The traditional dominant travel recommendation approaches are roughly divided into two categories: collaborative filtering recommendation [1,2] and content-based recommendation [3,4]. Travelers may travel once or twice a year even less on average, so travel-based user data is very sparse. Although collaborative filtering recommendation is much easier to implement than content-based recommendation, it will cause cold-start problem which depends on productive users' behaviors and profiles. Consequently, collaborative filtering approaches [5,6] are not appropriate for sparse data in travel situation. By comparison, content-based approaches can handle the sparse user data but only can cope with single-modality information instead of heterogeneous information. And some works [7,8] focus on visual information classification and the visual-based classification problem is well-suited for the attraction recommendation, because visual information can vividly

* Corresponding author.
E-mail address: xbgao@mail.xidian.edu.cn (X. Gao).

http://dx.doi.org/10.1016/j.neucom.2015.08.030
0925-2312/

represent the attractions. In reality, travel information contains massive heterogeneous information [9], including text, image and numerical value. In the case of travel recommendation, the content-based recommendation should be adopted in order to solve the problem with heterogeneous information from social media. Aiming to make full use of abundant heterogeneous resources in social media, the knowledge of collective intelligence will be explored to address the cold-start problem.

To further establish the personalized recommendation, the user information is collected either explicitly or implicitly. Explicit collection [10] means that user providing information actively, where the users should answer the questions provided by the interface for interaction. Implicit collection [11] is a passive way to mine users' interests from user historical behavior and context information. However, implicit way is not available for travel recommendation, because different cities have a diversity of architectural styles so that travel histories are not the best way as prior knowledge. For example, there are many different styles of parks around the world. The visual characteristics are not uniform in parks in different cities or even in the same city. Obviously, implicit information, such as user history, is not suitable for travel recommendation.

According to above analysis, due to the intrinsic characteristics of tourism, content-based recommendation with user explicit feedback, inevitably, is more suitable for travel personalized attraction recommendation. Therefore, we formulate a novel framework of travel attraction recommendation with personalization which consists of four principal modules, such as collective intelligence collection, knowledge extraction, PAS-model and user interaction. As shown in Fig. 1, in order to learn experiences of other travelers, we first collect travel heterogeneous information as collective intelligence from various travel-related websites on the Internet. Photos with metadata are crawled from Flickr [12] which are searched by the GPS location of each attraction. Meanwhile, official travelogues from Wikitravel and comments from Tripadvisor [13] are searched by the name of the attraction. In the same manner, ratings by travelers from Tripadvisor are tailed up. Then, knowledge is multi-modality descriptions of attractions, which is extracted from collective intelligence in different aspects, i.e., content-based, semantic-based and social-based. And then, given user interaction to avoid data sparsity and cold-start problem, the personalized attraction similarity model (PAS-model) is established with a combination of knowledge fusion to recommend attraction in a comprehensive view. In the model, each aspect of knowledge can construct graph-links between attractions with appropriate similarity measure. To realize personal recommendation, the user can choose favorite and unfavorite attractions as positive and negative labels in an explicit way, and the recommendation problem is considered as graph-based classification. Candidate recommended attractions are classified by graph-based multi-modality attraction information fusion in the way of weighted adaptation. Finally, users' current situation is utilized properly as context information to optimize the candidate attractions that can influence the user's choice under a particular

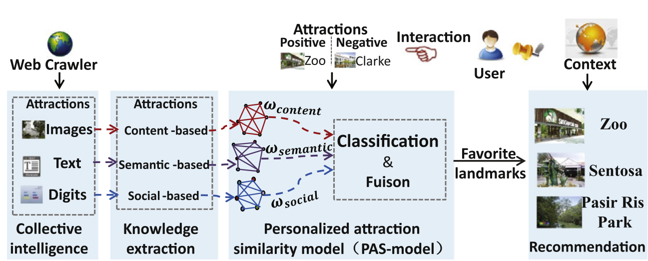

Fig. 1. The flowchart of the personalized recommendation framework.

condition. To give an intuitive way, personalized attraction recommendation is shown as an attraction ranking list in our system.

Our contributions are summarized as: (a) personalized attraction recommendation with explicit interaction is first composed in the personalized travel recommendation by analyzing collective intelligence from social media; (b) PAS-model is designed in a unified way to solve the recommendation problem which can mine the intrinsic links between modalities of heterogeneous information and fuse heterogeneous collective intelligence with weight adaptation; (c) context information are considered to refine the recommended attractions to simulate the particular situation to predict user's favors. The rest of the paper is organized as follows. Section 2 briefs related work. Then, in Section 3, we introduce our personalized attraction recommendation framework. The details of our framework are introduced in Section 3. Experiments and discussions are presented in Section 4. The final is conclusion in Section 5.

2. Related work

In an unfamiliar city, travelers want to visit both popular and favorite attractions. Travel attraction recommendation infers what the users' preferences and shows interesting and popular attractions for users to plan trips for them. Intelligent attraction recommendation on the basis of travel information and user's information is a hot topic. In general, dominant methods are classified as collaborative-filter (CF)-based, content-based methods.

On one hand, users' traveling histories, groups of users and user-location relationships are used by collaborative-filter-based methods. In [9], Markov model and topic model are combined to predict preference attractions based on user's traveling history. Based on collaborative filtering, [14] mined knowledge from GPS data to discover locations, and activities and a collective matrix factorization is utilized to recommendation. Cost-aware collaborative filtering [15] crawls travel logs from a travel company, and then represents cost factors associated with different travel packages. A latent factor model can join the cost factors together for recommendation.

On the other hand, content-based recommendations are produced by mining the travel information. In content-based methods, some works of travel recommendation utilized geo-tagged photos in social websites. Ref. [16] clustered a large amount of geotagged photos based on geo-location. Then, query provided by users can match the similar attractions based on text or images with the assumption that users will like similar attractions. Ref. [17] exploited the context information of photos, including textual tags, geolocation, images and the similarity of users. Similarities of users are calculated to predict user's favorite attractions. Then, a ranking algorithm is employed to show ranking attractions to users. When visiting to a city, the work of [9] proposed personalized recommend attractions with the Bayesian network techniques on the basis of users behavior and users' relationships. An expert tourist guide is presented by [18], which adapts not only user's travel history, but also considers travel time and users' preferences. In [19], based on locations traveling history and geotagged contributed images, user profiles and attributes are taken into account in personalized travel recommendation by a probabilistic Bayesian learning framework. Then, demographics are mined for personalized attraction and route recommendation.

It is observed that each category has its merits in recommendation. However, travel data has its special characteristic, such as sparsity and variety. In other words, when thinking about all the hassles of traveling, most people has been travelled once or twice in their whole lives. Thus, few travel histories can be obtained and collaborative filtering may be not easy to implement. As

Please cite this article as: J. Shen, et al., Attraction recommendation: Towards personalized tourism via collective intelligence, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.08.030

above-mentioned, different cities have different styles, for example, two parks in different cities may be quite different in vision. So, traditional methods for calculating the attractive similarity based on visual may not always get a good performance. Unlike current methods, our recommendation model considers the real world applications, obtaining users feedbacks explicitly and recommends attractions based on contents. Different from the previous approaches that most of the works focused on mining information from one principal modality and other modalities are considered as context information. Although few works focused on fusing multi-modality, they are ignored to mine latent relations among multi-modality. Thus, our method focuses on multi-modality fusion with latent relations to recommend attractions.

3. Personalized travel attraction recommendation

3.1. Problem definition

The purpose of our method is to fuse multi-modality of collective intelligence to recommend attractions based on explicit user feedback. We have the assumption that the recommended attractions are similar to user feedbacks. And travel attraction recommendation system is developed as input a set of famous attractions in each city and collective intelligence from social media. First, heterogeneous information is collected from social media as collective intelligence. Second, content-based, semantic-based and social-base knowledge can be mined from collective intelligence to show multi-modality descriptions of attractions. Then in PAS-model, laplacian graphs are constructed to represent the relations between attractions based on different modalities. Classification technique is employed to explore similar attractions based on user interaction, and then the similarities between all attractions are computed by the multi-modality fusion. Finally, the attractions are classified into favorite and unfavorite groups with multi-modality fusion. To harvest the results with freshness and surprise, the context information is introduced for final recommendation and the final outputs of the travel recommendation are possible attractions that the user is likely to visit.

Let $A=(A^1, A^2, \ldots, A^n)$ be a set of attractions in the user's location city, where A^i denotes an attraction. Heterogeneous travel information includes photos, comments and user ratings, which composed the collective intelligence. Thus, an attraction contains the knowledge from collective intelligence, which can be recorded as a 3-dimensional tuple: $A^i=\langle M^i, T^i, D^i \rangle$, where images, texts and ratings are collected by social media. Here, M^i denotes the image set of the attraction A^i, and content-based knowledge are extracted from the image set of the attraction, which is denoted as content-based feature vector, Co^i . Likewise, T^i denotes the text set of the attraction A^i and D^i denotes the digit set of the attraction A^i, where semantic-based features and social-based knowledge are extracted from them separately and are denoted as Se^i and So^i respectively.

Context information consists of current city, user's location with latitude and longitude, which can be described as a three-dimensional triple: $U=\langle C, L, T \rangle$, where C is the user's location city, L denotes the user's location with latitude and longitude, T is current time. The interaction with the user in our framework is regarded as user's favorite attractions which are labeled as positive samples, $A=(A^i, y^i=1)_i^p$, where p represents the number of user positive feedback. With the positive samples, we aim to solve the recommendation problem as classification problem to estimate the category of unlabeled samples. Laplacian graph can be built to measure the similarities between attractions with loosely labeled samples, thus various perspectives of collective intelligence can build different knowledge graphs on the basis of multi-modality feature vector, including Co^i, Se^i, So^i. Here, each node is represented as an attraction and the edge is the similarity between attractions. Then, G denotes a graph set of a city, where G^{Co}, G^{Se}, G^{So} denote content-based, semantic-based and social-based graphs respectively. Therefore, how to classify the attractions according to positive samples and how to fuse collective intelligence based on different graphs are significant problems to be solved. Finally, the recommendation problem is considered as loosely labeled graph-based classification problem with multi-modal fusion. Thus, a set of recommended attractions are classified on the basis of heterogeneous collective intelligence with user feedbacks. When adding context information, a list of recommendation attractions is presented to users. To further explain our model, the notations and their definitions of elements are shown in Table 1.

3.2. Heterogeneous information preprocessing

3.2.1. Data collection

We collect a large number of heterogeneous travel information from tourism-related social websites. The crawler database is accomplished by the deadline of "2014-08-30", which is crawled based on the famous attractions from six tourist cities, such as Beijing, Xi'an, Singapore, London, Paris, and New York. Top famous attractions are selected from each city, and the attractions are defined as queries to collect data. Fig. 2 shows the number of attractions in different cities. Moreover, the data are collected by the name of attractions. Table 2 shows the summary of our data collection. From Flickr, photos with metadata are crawled based on textual matching with the query of attraction-names. Each photo consists of time, location, attraction, and User ID. From the website of tripadvisor, user comments about the candidate attractions are downloaded, and user rating about each attraction is recorded as

Table 1
A list of key notations.

Notation	Definition
$A=(A^1, A^2, \ldots, A^n)$	A denotes the attraction set in an objective city, and A^i denotes the i-th attraction
$M_i=(M_1^i, M_2^i, \ldots, M_m^i)$	M^i denotes the image set of the i-th attraction, M_1^i is the image in the set.
$T_i=(T_1^i, T_2^i, \ldots, T_m^i)$	T_i denotes the text set of the i-th attraction, T_1^i is the document in the set.
$D_i=(D_1^i, D_2^i, \ldots, D_m^i)$	D_i denotes the digit set of the i-th attraction, D_1^i is the rating of the traveler.
$G=(V, E, w)$	G denotes a graph, and V, E, w denote the set of vertices, the set of edges and the weights of the edges.
Co^i	Content-based feature vector of attraction A^i
Se^i	Semantic-based feature vector of attraction A^i
So^i	Social-based feature vector of attraction A^i
G^{Co}	The graph constructed by content-based feature
G^{Se}	The graph constructed by semantic-based feature
G^{So}	The graph constructed by social-based feature

Please cite this article as: J. Shen, et al., Attraction recommendation: Towards personalized tourism via collective intelligence, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.08.030

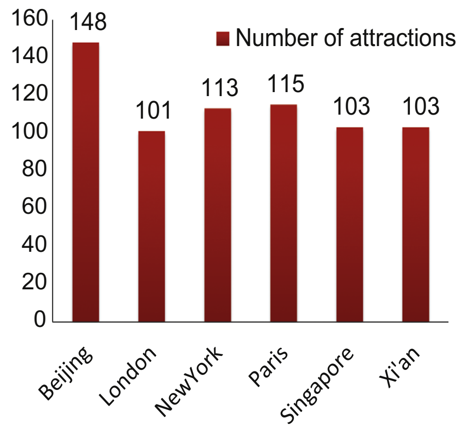

Fig. 2. The number of attractions in each city.

Table 2
The summary of our data collection.

Cities	Flickr		Tripadvisor	
	Photos	Tags	Comments	Ratings
Beijing	89282	89198	10578	23108
London	107592	107585	89521	115997
NewYork	121794	121793	105301	146354
Paris	124786	124783	71778	134935
Singapore	82290	82287	26812	32974
Xi'an	47257	45214	14004	19443

the statics of collective preferences. In addition, Wikitravel is the website with official travel information, so the information from Wikitravel could help to remove the noises on social websites.

3.2.2. Knowledge extraction from collective intelligence

Knowledge extraction will be introduced in this section. The core problem is to mine collective intelligence in different modalities. Content-based, semantic-based and social based knowledge can give a comprehensive view of attractions. Also, different modalities should choose appropriate measures to calculate the similarities of attractions. Thus, the details are introduced in the following independent aspects:

Content-based knowledge (CoK): Visual features are extracted to describe the content of the attractions in an intuitive view. However, a large number of photos from social media belong to an attraction, and lots of photos from social websites contain noises since many people expose themselves or other importance instead of the attraction itself. Thus, the photos should be clustered to eliminate the noisy photos, and global features including GIST [20] and color moment [21] are cascaded as one vector and k-means are used for clustering. In this paper, the number of clustering is 20. The dimensionality of GIST is 255, and so is color moments.

And according to common experience, we only analyze the top five clusters of photos to mine the content-based knowledge. Moreover, global feature [21,22] can represent the entire photo, but cannot static the characteristics of the sets of photos. For this reason, the local feature [23] is adopted to depict the attractions, so bag of visual words(BovW) [24] is supposed to represent the attraction, which is considered as content-based structured information. Visual feature extraction in view of photos employs BovW to represent the visual style in an attraction from photos after eliminating noise. To begin with, SIFT [23] is available to extract interest points of each photo to present abundant information. And then, the work of [25] is employed to cluster the interest points, and then clustering center is regarded as visual word and here the number of clustering values is 1024. All visual words from a photo set are composed of a visual word dictionary. Moreover, each image corresponds to a visual word distribution, and thus, the visual feature histogram of an image is obtained. Basic LDA [26] is adopted to extract topic model, and finally attraction A^i can be regarded as visual topic probability distribution, which is denoted as $P^M_{A^i}$. The number of visual topics is denoted as t^M, which is set to 15 empirically. Thus, the visual topic probability distribution can be used to measure the content-based similarity. Kullback–Leibler(KL) divergence is available for similarity measure in this situation,which is shown as follows:

$$D^M_{KL}(P^M_{A^i} \parallel P^M_{A^j}) = \sum_{i^M}^{t^M} P^M_{A^i}(i^M) \ln \frac{P^M_{A^i}(i^M)}{P^M_{A^j}(i^M)}. \tag{1}$$

Since the symmetrical KL divergence (SKLD) is adopted to measure the divarication of content-based structured information. Thus,

$$D_{Content} = D^M_{KL}(P^M_{A^i} \parallel P^M_{A^j}) + D^M_{KL}(P^M_{A^j} \parallel P^M_{A^i}). \tag{2}$$

Semantic-based knowledge (SeK): The user comments were generated by text when travellers have visited the attractions. Each attraction is presented by a collection of comments which contains much high-level semantic knowledge. Thus, attractions are represented by bag of words which are defined by travel-related words. Specific process is as follows, meaningless words are removed which referred to a predefined travel-related word vocabulary. The stemming approach is used to process the words in reserve and the words after stemming are employed to mine the semantics. After preprocessing of the textual information in each attraction, the word dictionary can be represented by words statics. All textual information of each attraction can be seen as a document and word distribution of a document can be obtained. Here, we introduce basic LDA [26] to extract textual topic model and the textual topic probability distribution can measure the semantic-based similarity, and also KL divergence is shown as follows:

$$D^T_{KL}(P^T_{A^i} \parallel P^T_{A^j}) = \sum_{i^T}^{t^T} MP^T_{A^i}(i^T) \ln \frac{P^T_{A^i}(i^T)}{P^T_{A^j}(i^T)}, \tag{3}$$

where $P^T_{(A^i)}$ is the textual topic probability distribution of the attraction A^i, t^T is the number of textual topics which is also set to 15. The symmetrical form of distance is shown as

$$D_{Semantic} = D^T_{KL}(P^T_{A^i} \parallel P^T_{A^j}) + D^T_{KL}(P^T_{A^j} \parallel P^T_{A^i}). \tag{4}$$

Social-based knowledge (SoK): The social websites provide rating item for travelers to record the personal experience, and represented as comprehensive scores, which can reflect traffic, service, cost-effective and so on. Though the form of the social-based information is fairly easy, it is the most important aspect for the users to make traveling decisions with collective intelligence. Consequently, the scores are collected from two famous travel websites, such as Yahoo and tripadvisor, which are denoted as S_1 and S_2 respectively. And then the ratings are normalized to represent the social influence. Since rating is extracted to describe attractions, in essence, the social-based similarity measure of attractions is to access the distance of ratings in a fixed vector space. Euclidean distance is an efficient way to measure the visual feature vector, as follows:

$$D_{Social} = D_{Eu}(R^i, R^j) = \exp\left(-\frac{\parallel R^i - R^j \parallel}{\sigma^2}\right), \tag{5}$$

Please cite this article as: J. Shen, et al., Attraction recommendation: Towards personalized tourism via collective intelligence, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.08.030

where R^i is the social-based feature of the attraction A^i, and σ is the distance constant, which is set to 0.2.

3.3. Personalized attraction similarity model (PAS-model)

Our previous sections show the process of collective intelligence mining from social media. Therefore, aiming to recommend attractions, we regard the recommendation problem as a classification problem, and we propose a model with the name of personalized attraction similarity (PAS) model. In the model, graph learning is introduced and multi-modality heterogeneous information is fused by graphs. The attractions of user feedback is considered as labeled data to training a classifier, and the rest of candidate attractions in the location city are unlabeled data. In consequence, manifold regularization [27] with Laplacian Support Vector Machines [28] is employed to learn the unlabeled attractions based on user feedbacks in single modality, which is shown as follows:

$$\min_{f \in H_m} \frac{1}{l}(1-y_i f(x_i))_+ + \gamma_A \|f\|_K^2 + \gamma_I \|f\|_I^2, \tag{6}$$

where the reproducing kernel Hilbert space (RKHS) is denoted as H_m, the hinge loss is employed in basic Laplacian support vector machine (LapSVM) [28] which is denoted as $(1-y_i f(x_i))_+$ to classify the unlabeled attractions, $\|f\|_K^2$ is the cost regularization to penalize the complexity of classifier, $\|f\|_I^2$ controls the smoothness along the manifold assessed from unlabeled attractions, γ_A and γ_I are the weights to balance the loss function and regularization. Further, $\|f\|_I^2$ is normalized graph Laplacian regularizer, which can be approximated by graph laplacian. Given labeled data $\left\{(A^i, y_i)\right\}_{i=1}^{l}$ and unlabeled data $\left\{(A^j)\right\}_{j=l+1}^{l+u}$, it turns to solve the optimization problem:

$$\min_{f \in H_m} \frac{1}{l}(1-y_i f(A^i))_+ + \gamma_A \|f\|_K^2 + \frac{\gamma_I}{(l+u)^2}\mathbf{f}^\top L\mathbf{f}, \tag{7}$$

where $\|f\|_I^2 = \frac{1}{(l+u)^2}\mathbf{f}^\top L\mathbf{f}$. L is denoted as $L = I - D^{-(1/2)}WD^{(1/2)}$, and W is a similarity matrix and w_{ij} indicates the similarity between two attractions in one modality, and here, d_{ii} sums the i-th row of the similarity matrix W and D is the diagonal matrix.

In our model, separate graphs, including content-based, social-based and semantic-based graphs, are constructed to represent the relations among attractions based on heterogeneous information. In total, three modalities will be joined together for a combination seamlessly. Our PAS model employs ensemble learning with manifold regularization [29] to fuse heterogeneous travel information. Consequently, the learned ensemble manifold regularization framework can take the calculation of personalized attraction similarity as a problem of classification in view of multi-modality knowledge. The modalities will be joined together for a linear combination, where the weights of modalities will be learnt and the unlabeled attractions can be classified into users favorite or un-favorite attractions. Accordingly, heterogeneous collective intelligence are fused based on user feedbacks, and the classification function with ensemble manifold regularization [29] can be represented by

$$\min_{f \in H_m} \frac{1}{l}(1-y_i f(A^i))_+ + \gamma_A \|f\|_K^2 + \gamma_I \sum_{k=1}^{K} \mu_k \|f\|_I^2 + \gamma_R \|\mu\|^2,$$
$$s.t. \sum_{k=1}^{m} \mu_k = 1, \mu_k \geq 0, \quad \text{for } k = 1, \ldots, m,$$

where an ensemble manifold regularization term turns out to be $\sum_{k=1}^{K} \mu_k \|f\|_I^2$, and μ_k are the weights to join the K modalities. γ_R is the parameter to balance other items and $\|\mu\|^2$ can avoid μ_k overfitting to a single modality.

From the optimization problem, the variables of the categories of unlabeled data and the weights of combining different modalities need to be solved. An alternating optimization will be adopted for optimization solving. Particularly, f and μ are updated alternatively to optimize the function.

First, we fix μ and optimize f. The unconstrained primal problem can be changed into a constrained problem with the slack variable $\xi_i \geq 0$. According to [7], we can get that $f(A)' = \sum_{i=1}^{n} \alpha_i' K(A^i, A)$, where K is considered as the gram matrix.

Then, we introduce nonnegative Lagrange multipliers for inequality constraints and bias is also introduced which is common in SVM derivation. Thus, the function can be derived by the solution of LapSVM, which is proposed by Belkin [27].

Second, the coordinate descent-based algorithm is adopted to learn the weights μ with a fixed f. In each round, two elements are selected to update and the others are fixed. In particular, if the two weights are selected, the sum of the two weights will be invariant in this iteration. Consequently, μ will be solved according to the rules in [29]:

$$\begin{cases} \mu_i' = 0, \mu_j' = \mu_i + \mu_j, & \texttt{if } 2\gamma_R(g_i+g_j)+(\alpha_j-\alpha_i) \leq 0 \\ \mu_i' = \mu_i + \mu_j, \mu_j' = 0, & \texttt{if } 2\gamma_R(g_i+g_j)+(\alpha_i-\alpha_j) \leq 0 \\ \mu_i' = \dfrac{2\gamma_R(g_i+g_j)+(\alpha_j-\alpha_i)}{4\gamma_R}, \mu_j' = \mu_i + \mu_j - \mu_i', & \text{otherwise} \end{cases} \tag{8}$$

The process will be iterated in all pairs until the objective function does not decrease. Since $g_k = (\gamma_I/(l+u)^2)\mathbf{f}^\top L_k \mathbf{f}$ keeps the smoothness in the k-th modality, and a smaller g_k measures the consistency of previous manifold. Since PAS-model can integrates heterogeneous travel information and the weights learning can be capable of feature selection in different modalities. The complexity of the PAS-model is $O(l+u^3)$. Additionally, we only process the attractions in the location city for high computational efficiency. In each city, about 100 attractions are selected for recommendation. Thus, we can have a high efficiency in real application.

3.4. Rank to recommendation in the view of user context

One of the most important components in recommendation is to rank attractions in the view of user context. The model will find a lot of available attractions, but the user wishes to have an intuitive recommendation. Thus, an attraction ranking list is a good way for recommendation. As previous introduction, the hard classification can be turned into the output of probability. The learned function f can be changed into probability, and the probability of the attraction A^i is defined as

$$P^A(A^i) \approx \frac{1}{1+\exp(-f(A^i))}. \tag{9}$$

Moreover, context information is also an important component in recommendation. The user location can be obtained and the ranking of attractions is influenced by the geo-distance to determine the recommendation. The probability of user-wish attraction somehow is determined by geo-distance between user and attractions. The probability of distance-context between the user location $g(U)$ and the attraction-location $g(A)$, is defined by

$$P^G(A^i) \approx \frac{1}{1+\exp(-|g(U)-g(A)|)}. \tag{10}$$

After calculating the probability of attractions, the ranking can be estimated by both collective intelligence and context-based probabilities, we formulate the final ranking score of an attraction A as

$$S(A) = \alpha_s P^A(A^i) + (1-\alpha_s)P^G(A^i) \tag{11}$$

Please cite this article as: J. Shen, et al., Attraction recommendation: Towards personalized tourism via collective intelligence, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.08.030

where $0 < \alpha_s < 1$ and according to experience, we set α_s to 0.8 in order to balance content and context. Both content-based and context-based information can influence the final ranking, but content-based information plays a principle role in recommendation. And context can assist in enhancing user experience. Thus, we set empirical value at 0.8. If the score of one attraction is higher than the other, the attraction is appropriate for recommendation. Therefore, the final recommended attractions will be obtained which not only considered personalized must-see attractions, but also jointed the context information of the current user.

4. Experiments

4.1. Evaluation methodology

For our travel attraction recommendation framework, we are intended to evaluate the framework from five aspects, including accuracy, surprise, computational efficiency and freshness . Following sections will introduce the details.

4.1.1. Accuracy: Pseudo-relevant sample verification of PAS-model

The experiments are performed on the crawled dataset to verify the results affected by multi-modality fusion. And we will select a set of pseudo-relevant samples as labeled data to simulate the process of recommendation. The purpose of PAS-model is to recommend similar attractions which are similar to users' feedback. A set of popular words is predefined to label relevant attractions and the relevant top attractions are selected as the pseudo-relevant labeled samples. A set of queries are selected, such as "*Art*", "*Recreation*", "*Museum*", "*Zoo*", "*Gardens*", "*Historic*","*Education*", "*Malls*", "*Parks*", "*Mountain*","*Peaceful*", "*Happy*", "*Hot*", "*Walking*", "*Hiking*". To evaluate the model by pseudo feedback, the ground-truth should be labeled by local travel enthusiasts who are familiar with the attractions. The enthusiasts are asked to distinguish the relevance of the attractions based on the pseudo-labeled word. And the attractions for each kind are divided into two categories: relevant and irrelevant. In each word round, we randomly selected p positive attractions from relevant ones of each kind and p negative attractions are also randomly selected. Our PAS-model can classify the attractions in the located city. And similar attractions can be obtained according to the labeled ground truth as users feedback. Then, precision should also be calculated to evaluate our PAS-model. Consequently, *precision@k* is employed to estimate the accuracy and it measures the relevance of the returned results. *Precision@k* is defined as the proportion of true positive results (TR) in k attractions. We compare the following methods:

- *Popular rating* (*PR*): In this method, popularity ratings are estimated by users and an initial ranking list can be obtained by TF-IDF. And then, a new ranking list can be obtained.
- *Manifold regularization based on visual information* (*MR-V*): Manifold regularization framework [27] can learn with labeled and unlabeled attractions. In this method, the graph is constructed by visual information, where the feature named CoK is used here.
- *Manifold regularization based on textual information* (*MR-T*): In this method, we also employ manifold regularization framework with textual feature (SeK) to recommend similar attractions in view of user's favorite.
- *PAS-model with no weights learning based on heterogeneous features* (*PAS-nw*): Heterogeneous features (CoK, SeK, SoK) are all used to construct graphs respectively, and the weights of features are treated equally for final optimization.
- *PAS-model with weights learning based on heterogeneous features* (*PAS-w*): Weights learning is proposed in the model and the weights are varied with labeled data.

Table 3
The precision@10 results between PAS-model and baseline approaches in cities of our dataset.

Query	*PR*	*MR-V*	*MR-T*	*PAS-nw*	*PAS-w*
Art	0.68	0.85	0.92	0.84	**0.86**
Recreation	0.54	0.54	**0.56**	0.53	**0.56**
Museum	0.88	0.93	**0.96**	0.92	0.95
Zoo	0.74	0.76	0.72	0.71	**0.82**
Gardens	0.81	**0.83**	0.82	0.80	**0.83**
Historic	0.59	0.66	**0.72**	0.62	0.65
Education	0.66	0.71	0.73	0.69	**0.76**
Malls	0.80	0.83	**0.89**	0.84	0.88
Parks	0.89	0.91	0.94	0.90	**0.95**
Mountain	0.76	0.84	0.89	0.82	**0.91**
Peaceful	0.51	**0.66**	0.61	0.58	0.59
Happy	0.31	0.38	**0.54**	0.41	0.53
Hot	0.25	**0.53**	0.42	0.25	0.36
Walking	0.67	**0.73**	0.58	0.62	0.72
Hiking	0.68	0.74	**0.79**	0.73	0.75

Table 3 gives a clear show for the *precision@*10 results between PAS-model and baseline approaches in cities of our dataset. Experimental results show that the proposed PAS-model can make a better similarity measure in view of heterogeneous collective intelligence. Here, in each query, we randomly picked two positive labeled attractions and two negative ones according to the ground truth. After query-based searching, the popular rating ranking in famous websites can only show the most high-score attractions to users and ignore the content-based and semantic-based users interests to develop a personalized strategy. As a result, a special user for other hobbies will be not satisfied. MR-V and MR-T are requested to find similar attractions by visual or textual information on collective intelligence. But both of the methods consider partial view of attraction. Thus, performance will be influenced by different queries. For example, the query word historic can be described well by visual information instead of textual, thus the performance of MR-V by the query historic is better than MR-T's. Also, PAS-model fuses multi-modality of collective intelligence and can mine latent relations between different modalities. Our model fuses heterogeneous information in view of collective intelligence and can mine latent relations between different modalities. When we treated the three aspects of information equally in PAS-nw, and it is just to fuse heterogeneous information without weight learning. From the results, they do not focus on the important characteristics to represent candidate attractions. The performance of PAS-w shows better precision, because the adaptive weights can find the most similar attractions in fusion of the heterogeneous information.

4.1.2. Surprise: Real data verification

In travel attraction recommendation, tourism contains occasional mutations and randomness, because the data is sparse. Thus, the number of attractions that the user has been visited will influence the accurate estimation. For example, if the user only travels two attractions in a city, the recommendation with real data verification may be a great deal of randomness. We will evaluate the accuracy with the number of attractions that users visited and the top-k feedback. We collect photos from Flickr [12] and the metadata of the photos. Suppose that if a user uploaded the photos of the attraction, it is supposed that the user has been to this attraction. Table 4 gives a summary of the number of users employed in different cities. In order to eliminate the noise of

Please cite this article as: J. Shen, et al., Attraction recommendation: Towards personalized tourism via collective intelligence, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.08.030

Table 4
An overview of the number of users in the dataset.

Cities	The number of users	
	Total	Visited attractions ≥ 5
Beijing	4254	325
London	16955	1259
New York	16173	1304
Paris	14465	1391
Singapore	11577	504
Xi'an	3546	164

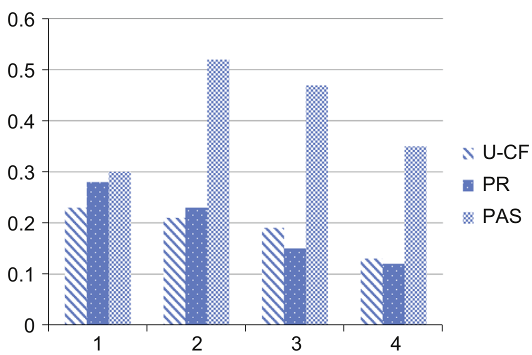

Fig. 3. Performance comparison in different number of positive attractions when precison@10.

photos, users who uploaded less than five attractions are deleted from the dataset. And in this way, we got user travel history from photos uploading by users and 80 users are used to test our model in each city. For each individual user, the attractions are formed by two parts, where the visited attractions are as one part, another part is candidates of attractions. In our experiments, the visited attractions are divided into two parts. One part is viewed as labeled data, and the other part is mixed into the candidates as unlabeled data. The experiment on real data verification aimed at the comparisons among popular rating (PR), user-based collaborative filtering recommendation algorithm (U-CF) and our PAS-model. Here, U-CF is a method to find the users who has visited the same attractions. And the user is supposed to prefer the other attractions that similar users have been to. The percentage of the feedbacks which are visited by travelers is defined as precision. Therefore, P@N evaluates the positive results of the first N. The evaluation metric is the proportion statistics of the correct results from the recommended top-n. The higher P@N means more available results returned and the higher performance our system is. For each method, the value of precision@10 is calculated respectively when positive labeled data which are selected from visited attractions changes from 1 to 4.

Fig. 3 illustrated the *precision@*10 for different methods on our travel heterogeneous dataset. From Fig. 3, it is noted that PAS-model is superior to user-based collaborative filtering and popularity rating. This is because that user-based collaborative filtering only thought the relations between users. However, the travel data is very sparse which is not suitable for collaborative-based recommendation. Moreover, user-based collaborative filtering ignores the content-based relations of attractions. As for popularity rating, it only considers rating of attraction, which only one modality of heterogeneous information. PAS-model considers content-based attraction recommendation and can well suit for sparsity travel data. In addition, multi-modality heterogeneous information is fused in our model. Thus, PAS-model performances well in comparisons of two commonly used methods. Moreover in PAS-model, we have the conclusion that when the

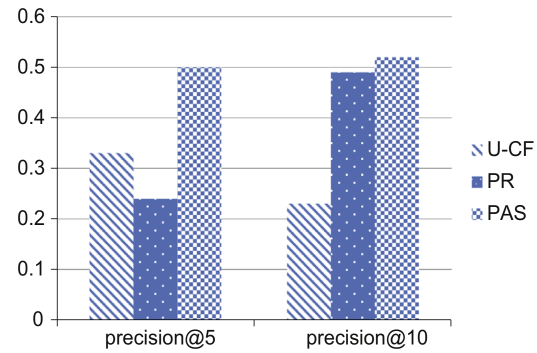

Fig. 4. Performance comparison in top n when labeled two positive attractions.

number of the labeled positive attractions equals two, the performance is best. This is because that, in real world, most people came to a new city, and they usually wish to visit famous and popular attractions. Therefore, personalized preferences are not really necessary. But to some extent, personalized preferences can influence the performance because that the performance of two labeled attractions is better than one labeled. Thus, if the user just wants to have a good tourism in a new city with our recommendation system, two relative feedbacks are wished to select. And it is also accordant with actual situations that users will drop their satisfaction when asking to label too many attractions. Furthermore, we have analyzed the real travel history, and the attractions have relations on location.

Fig. 4 shows the performance comparison when *precision@*5 and *precision@*10. In U-CF, *precision@*5 shows better performance, that is because most users wished to visits famous attractions and some users may have similar travel attractions in their tourism. In general, users who have two or more similar attractions will have an overlap in other attractions. Furthermore, in our PAS-model, it is concluded that the performance of *precision@*5 is little worse than *precision@*10. Thus, if more attractions are advised to users, the performance will be better. And it also provides more choices for users to make decisions.

4.1.3. Computational efficiency

Efficiency is the main element for a recommendation model, which is interpreted as space-efficiency and time-efficiency. Since space-efficiency is considered as accuracy of the model, we mainly discuss time-efficiency as computational efficiency. Moreover, computing divides itself two parts: off-line computing and on-line computing. Off-line computing represents data crawler, and data preprocessing which can be ready for the core algorithm. The core algorithm recommends attractions in view of data preprocessing and thus, the online efficiency is the most important for time efficiency. In each location city, about 100 attractions are employed for attraction recommendation. Thus, in the small data scale, we have a high efficiency for online-learning. The average running is employed to evaluate the online computing as the system efficiency. All experiments were conducted on a 3.2 GHz CPU(Intel Core $i5-3470$) and 8 GB memory. In the location city, there are only about a hundred attractions and the online time is about 0.0015 s on average.

4.1.4. Freshness

Experiments show that the proposed personalized recommendation framework can make a better recommendation based on heterogeneous collective intelligence. The popularity rating ranking in famous websites can only show the most popular and famous attractions to users and ignore the personalized interests to develop personalized recommendation strategy. As a result, it recommends applicable to certain users, a special user for other

Please cite this article as: J. Shen, et al., Attraction recommendation: Towards personalized tourism via collective intelligence, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.08.030

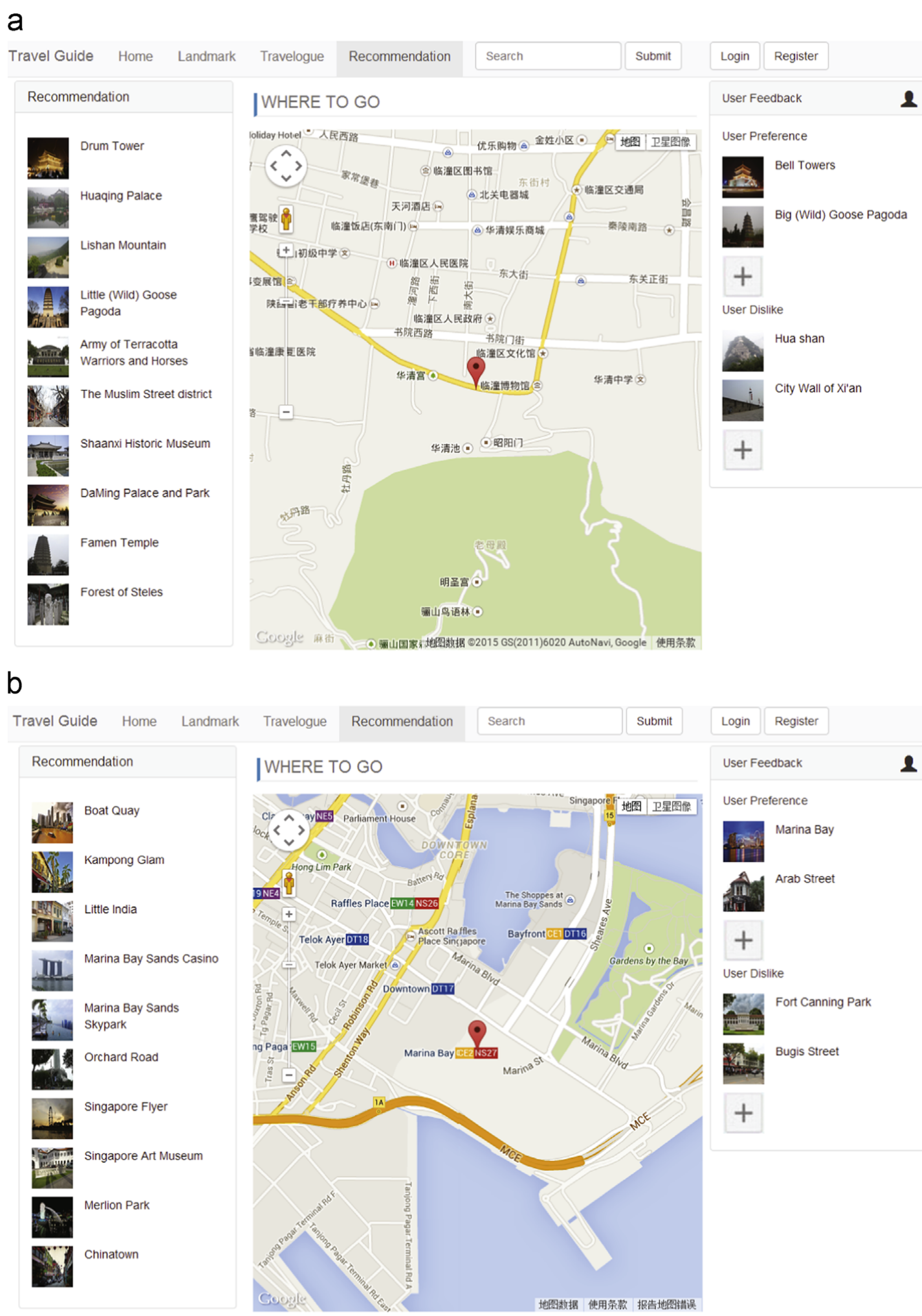

Fig. 5. User's interface is shown and two visual examples are given. (a) The user is in Xi'an. (b) The user is in Singapore.

hobbies will be not satisfied. If we guessed correctly what the users like, we will bring a great user experience, and the system will be thought to be smart and amazing. Thus, recommendation algorithms should also consider context information. For instance, we can obtain peoples locations to refine travel attraction recommendation. When recommending the attractions to the users, context will be included to improve the results, because the users prefer to visit nearby attractions in real life. In our framework, the PAS model places an important part in recommendation, but in real world, users will consider the contextual information. Thus, PAS-model with context will performance better than PAS-model, because we considered the geo-location for further recommendation in our framework. In a word, we considered not only accuracy, but also freshness and surprise. We wish users to be surprised with system, but it is rather difficult to define a accurate metric.

4.1.5. Visual examples

Fig. 5 shows two visual examples of our personalized travel recommendation. The figure shows the interface of the experimental environment. The system can collect the current location and show the located city on the map. The user can input their favorite and un-favorite attractions on the right of the interface. If the user does not wish to interact with the system, the system will show them the results which are ranked by popularity, to avoid cold-start problem. As showed in Fig. 5, the user gave his interaction and the attractions are recommended to users. Our system can not only leverage collective intelligence and user's preference, but also support the collection of context to refine the recommendation. Thus, our system can provide convenient for users and show its intelligence, which can guess the users' favorites.

5. Conclusion

It is desperately needed to recommend relevant attractions to users with the consideration of travel information and user interaction. In this paper, we present a novel framework of personalized travel attraction recommendation by leveraging explicit user interaction with the system and heterogeneous travel information. Aiming to learn from tourists who had just been there, collective intelligence is first gathered from large amount of user-generated content in social media. Second, different aspects of knowledge can be mined from collective intelligence for denoising data and structuring heterogeneous information. Then, we introduced a personalized attraction similarity (PAS) model unifies travel information fusion and users feedbacks to recommend candidate attractions. Finally, the framework also considered context information to improve the results of recommendation, which can correspond to the user's actual situation. Experimental results show that our method outperforms several approaches with the benefits of collective intelligence and explicit user interaction.

The main purpose of our proposed method is to combine users' feedbacks and collective intelligence. For future work, it will take full advantage of other context information to improve users presence and experiences. Thus, the framework can be implemented on the mobile to make the interaction faster and collect more context instead of current experimental environment. Besides, exploring other auxiliary information from collective intelligence is another issue to be solved in the future.

Acknowledgment

This work is supported by the National High Technology Research and Development Program of China (2013AA01A602), the National Natural Science Foundation of China (Nos. 61125204 and 61432014), the Program for New Century Excellent Talents in University (NCET-12-0917), the Fundamental Research Funds for the Central Universities (Grant Nos. BDZ021403 and K5051302019), the Key Science and Technology Program of Shaanxi Province, China (2014K05-16) and Program for Changjiang Scholars and Innovative Research Team in University (No. IRT13088).

References

[1] B. Sarwar, G. Karypis, J. Konstan, J. Riedl, Item-based collaborative filtering recommendation algorithms, in: Proceedings of the 10th International Conference on World Wide Web, ACM, Hong Kong, May, 2001, pp. 285–295.
[2] J.L. Herlocker, J.A. Konstan, J. Riedl, Explaining collaborative filtering recommendations, In: Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work, ACM, Philadelphia, PA, USA, December, 2000, pp. 241–250.
[3] C.D. Manning, P. Raghavan, H. Schütze, Introduction to Information Retrieval, vol. 1, Cambridge University Press Cambridge, Cambridge, British, 2008.
[4] D. Billsus, M.J. Pazzani, User modeling for adaptive news access, User Model. User-Adapt. Interact. 10 (2–3) (2000) 147–180.
[5] S.K.L, Al Mamunur Rashid, G. Karypis, J. Riedl, Clustknn: a highly scalable hybrid model memory-based cf algorithm, Proceeding of WebKDD.
[6] B. Sarwar, G. Karypis, J. Konstan, J. Riedl, Application of dimensionality reduction in recommender system—a case study, Technical Report, DTIC Document, 2000.
[7] F. Zhu, L. Shao, Weakly-supervised cross-domain dictionary learning for visual recognition, Int. J. Comput. Vis. 109 (1–2) (2014) 42–59.
[8] L. Shao, L. Liu, X. Li, Feature learning for image classification via multiobjective genetic programming, IEEE Trans. Neural Netw. Learn. Syst. 25 (7) (2014) 1359–1371.
[9] T. Kurashima, T. Iwata, G. Irie, K. Fujimura, Travel route recommendation using geotags in photo sharing sites, In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, ACM, Toronto, Canada, October, 2010, pp. 579–588.
[10] G. Jawaheer, M. Szomszor, P. Kostkova, Comparison of implicit and explicit feedback from an online music recommendation service, in: Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion In Recommender Systems, ACM, Barcelona, Spain, September, 2010, pp. 47–51.
[11] M. Balabanovic, An interface for learning multi-topic user profiles from implicit feedback, In: AAAI-98 Workshop on Recommender Systems, 1998.
[12] Flickr ⟨https://Flickr.com⟩.
[13] Tripadvisor ⟨https://tripadvisor.com⟩.
[14] V.W. Zheng, Y. Zheng, X. Xie, Q. Yang, Collaborative location and activity recommendations with gps history data, In: Proceedings of the 19th International Conference on World Wide Web, ACM, North Carolina, USA, April, 2010, pp. 1029–1038.
[15] Y. Ge, Q. Liu, H. Xiong, A. Tuzhilin, J. Chen, Cost-aware travel tour recommendation, In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, San Diego, CA, August, 2011, pp. 983–991.
[16] L. Cao, J. Luo, A. Gallagher, X. Jin, J. Han, T.S. Huang, A worldwide tourism recommendation system based on geotagged web photos, In: 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), IEEE, Dallas, TX, April, 2010, pp. 2274–2277.
[17] K. Jiang, P. Wang, N. Yu, Contextrank: personalized tourism recommendation by exploiting context information of geotagged web photos, In: 2011 Sixth International Conference on Image and Graphics (ICIG), IEEE, Hefei, Anhui, China, August, 2011, pp. 931–937.
[18] I. Memon, L. Chen, A. Majid, M. Lv, I. Hussain, G. Chen, Travel recommendation using geo-tagged photos in social media for tourist, Wirel. Pers. Commun. (2014) 1–16.
[19] A.-J. Cheng, Y.-Y. Chen, Y.-T. Huang, W.H. Hsu, H.-Y.M. Liao, Personalized travel recommendation by mining people attributes from community-contributed photos, in: Proceedings of the 19th ACM International Conference on Multimedia, ACM, Scottsdale, Arizona, USA, November, 2011, pp. 83–92.
[20] A. Oliva, A. Torralba, Modeling the shape of the scene: a holistic representation of the spatial envelope, Int. J. Comput. Vis. 42 (3) (2001) 145–175.
[21] A.K. Jain, A. Vailaya, Image retrieval using color and shape, Pattern Recognit. 29 (8) (1996) 1233–1244.
[22] A.K. Jain, R.C. Dubes, et al., Algorithms for Clustering Data, vol. 6, Prentice-Hall Inc., Upper Saddle River, NJ, USA, 1988.
[23] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[24] L. Fei-Fei, P. Perona, A bayesian hierarchical model for learning natural scene categories, In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, CVPR 2005, vol. 2, IEEE, Los Alamitos, 2005, pp. 524–531.
[25] J. MacQueen, et al., Some methods for classification and analysis of multivariate observations, In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, Oakland, CA, USA, 1967, pp. 281–297.
[26] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent dirichlet allocation, J. Mach. Learn. Res. 3 (2003) 993–1022.
[27] M. Belkin, P. Niyogi, V. Sindhwani, Manifold regularization: a geometric framework for learning from labeled and unlabeled examples, J. Mach. Learn. Res. 7 (2006) 2399–2434.
[28] L. Gómez-Chova, G. Camps-Valls, J. Munoz-Mari, J. Calpe, Semisupervised image classification with laplacian support vector machines, IEEE Geosci. Remote Sens. Lett. 5 (3) (2008) 336–340.
[29] B. Geng, D. Tao, C. Xu, Y. Yang, X.-S. Hua, Ensemble manifold regularization, IEEE Trans. Pattern Anal. Mach. Intell. 34 (6) (2012) 1227–1233.

Junge Shen received the B.Sc. degree in electronic engineering from Xidian University in 2010. Currently she is pursuing her Ph.D. degree at School of Electronic Engineering, Xidian University. Her current research mainly focuses on multimedia and computer vision.

Cheng Deng received the B.Sc., M.Sc., and Ph.D. degrees in signal and information processing from Xidian University, Xi'an, China. He is currently a Full Professor with the School of Electronic Engineering at Xidian University. His research interests include computer vision, multimedia processing and analysis, and information hiding. He is the author and coauthor of more than 40 scientific articles at top venues, including IEEE T-MM, T-SMC, T-IP, Signal Processing, Neurocomputing, ICCV, CVPR and AAAI.

Xinbo Gao received the B.Eng., M.Sc. and Ph.D. degrees in signal and information processing from Xidian University, China, in 1994, 1997 and 1999 respectively. From 1997 to 1998, he was a research fellow in the Department of Computer Science at Shizuoka University, Japan. From 2000 to 2001, he was a postdoctoral research fellow in the Department of Information Engineering at the Chinese University of Hong Kong. Since 2001, he joined the School of Electronic Engineering at Xidian University. Currently, he is a Professor of Pattern Recognition and Intelligent System, and Director of the VIPS Lab, Xidian University. His research interests are computational intelligence, machine learning, computer vision, pattern recognition and wireless communications. In these areas, he has published 5 books and around 150 technical articles in refereed journals and proceedings including IEEE T-IP, T-CSVT, T-NN, T-SMC, Pattern Recognition etc.

He is on the editorial boards of several journals including Signal Processing (Elsevier), and Neurocomputing (Elsevier). He served as general chair/co-chair or program committee chair/co-chair or PC member for around 30 major international conferences. Now, he is a Fellow of IET and Senior Member of IEEE.

Please cite this article as: J. Shen, et al., Attraction recommendation: Towards personalized tourism via collective intelligence, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2015.08.030