11-2016

Landmark reranking for smart travel guide systems by combining and analyzing diverse media

Junge SHEN
Xidian University

Jialie SHEN
Singapore Management University, jlshen@smu.edu.sg

Tao MEI
Microsoft Research Asia

Xinbo GAO
Xidian University

DOI: https://doi.org/10.1109/TSMC.2016.2523948

Published in IEEE transactions on Systems, Man, and Cybernetics, 2016 Nov, Vol. 46 (11), p.1492-1504.
http://doi.org/10.1109/TSMC.2016.2523948

Landmark Reranking for Smart Travel Guide Systems by Combining and Analyzing Diverse Media

Junge Shen, Jialie Shen, *Member, IEEE*, Tao Mei, *Senior Member, IEEE*, and Xinbo Gao, *Senior Member, IEEE*

***Abstract*—Advanced networking technologies and massive online social media have stimulated a booming growth of travel heterogeneous information in recent years. By employing such information, smart travel guide systems, such as landmark ranking systems, have been proposed to offer diverse online travel services. It is essential for a landmark ranking system to structure, analyze, and search the travel heterogeneous information to produce human-expected results. Therefore, currently the most fundamental yet challenging problems can be concluded: 1) how to fuse heterogeneous tourism information and 2) how to model landmark ranking. In this paper, a novel landmark search system is introduced based on a newly designed heterogeneous information fusion scheme and a query-dependent landmark ranking strategy. Different from the existing travel guide systems, the proposed system can effectively combine the heterogeneous information from multimodality media into a landmark reranking list via a user's query. Experimental results conducted on a large travel information collection illustrate the advantages of the proposed system in terms of both effectiveness and efficiency.**

***Index Terms*—Heterogeneous multimedia analysis, information fusion, landmark reranking, travel guide.**

I. Introduction

With the explosive growth of social media and development of Web 2.0, large amounts of travel information are being uploaded per minute on travel websites. Nowadays, the idea of travel and the way how we travel have changed a lot. Self-guided tours are becoming more and more popular. Therefore, the high-quality travel information is necessary for travelers to plan their holidays.

Manuscript received July 3, 2015; revised August 26, 2015; accepted September 26, 2015. Date of publication March 15, 2016; date of current version October 13, 2016. This work was supported in part by the National High Technology Research and Development Program of China under Grant 2013AA01A602, in part by the National Natural Science Foundation of China under Grant 61125204 and Grant 61432014, in part by the Fundamental Research Funds for the Central Universities under Grant BDZ021403 and Grant K5051302019, in part by the Key Science and Technology Program of Shaanxi Province, China, under 2014K05-16, and in part by the Program for Changjiang Scholars and Innovative Research Team in University under Grant IRT13088. This paper was recommended by Associate Editor B.-F. Wu.

J. Shen and X. Gao are with the School of Electronic Engineering, Xidian University, Xi'an 710071, China (e-mail: xbgao@mail.xidian.edu.cn).

J. Shen is with the School of Information Systems, Singapore Management University, Singapore.

T. Mei is with the Media Computing Group, Microsoft Research Asia, Beijing 10080, China.

Travel guide systems are developed to recommend travel information to interested users. Although most existing commercial travel guide systems [1]–[3] can provide travel-related information effectively, it is inefficient for travelers to pinpoint the useful information from dazzling travel information. Consequently, smart travel guide systems have been introduced to offer travelers tour's essence and help them travel wisely. However, designing a smart travel guide system is a difficult task for the following reasons. The travel information is uploaded by tourism editors and travelers in casual ways. Moreover, the travel information is spread to different social media with different modality. The travel information bears heterogeneous unstructured forms, and consequently slows down the development of smart travel guide systems. Current commercial travel websites only tell the tourists what must-see landmarks there are. The landmarks are generally ranked by user ratings without considering users' specific requirements, therefore failed to be intelligent. As a result, landmark reranking is important to improve intelligent travel guide systems.

Table I lists several main categories of travel guide systems which relate to landmark ranking. In order to have representative views of landmarks, these systems collect tourism information from diverse media, including texts, images, and digits. We can thus classify these systems into two categories in terms of the modality of the tourism information. One category is single-modality information mining. For example, diversified landmark (DLM) [4] aims to get detailed views of landmarks in a city. This method utilizes the visual information extracted from social media to analyze landmarks, and shows the candidate landmarks with representative views. Another category focuses on multimodality heterogeneous information fusion. Diversified landmark-table of content [5] establishes a search and browsing system, which learns from community photographs and forms an overview of the landmarks by reranking images. Kennedy and Naaman [6] explored the travel information of landmarks where meaningful tags, visually representative features, and geographic features are extracted to generate descriptive views of a landmark without noise. In [7], a system is designed to mine user-contributed travelogues and photographs for virtual tours. This system can not only recommend popular places with comprehensive aspects, but also give representative views of landmarks. It can be seen from the aforementioned systems that the research stream appears to involve heterogeneous information. Along with this trend,

TABLE I
SUMMARY OF THE MAIN CATEGORIES OF TRAVEL GUIDE SYSTEMS

System	Physical representation	Cross-domain	Heterogeneous	Query dependent	Semantic related
DlMSearch [4]	Landmarks Ranking	Flickr	Visual	Yes	No
DL-TOC [5]	Images reranking of landmarks	Flickr	Textual+Visual	No	No
Diversified [6]	Landmark view representation	Flickr	Textual+Visual	No	No
Hao *et al's* [7]	Landmark view generation	Flickr	Textual+Visual	No	Yes
W2GO [8]	Landmarks Ranking	Flickr+Yahoo	Textual+Visual	No	Yes
Ours	Landmark ranking list	Flickr+tripadvisor+wiki	Textual+Visual+Digital	Yes	Yes

W2GO [8] analyzes landmarks by using social information such as the geo-tagged photographs in Flickr and the travel reviews collected from Yahoo Travel guide [1]. The tags and reviews of travel photographs are adopted to calculate the popularity of landmarks which are suitable for users to choose. W2GO [8] provides the top three results in a city, and it recommends a landmark ranking list without considering the diverse personal needs. Moreover, it does not take account of landmark search. Therefore, it is inconvenient for users to seek their favorite landmarks. For instance, most people may have questions before visiting a destination, for example:

"I will arrive at Beijing this weekend, and I want to visit some cultural relics. But I am unfamiliar with that city."

"I want to take my child to Singapore. It would be good to have appropriate landmarks suggested for children."

As aforementioned, it is a pressing issue for travel guide systems to add the functionality of search at landmark-level to improve user experiences. Existing travel guide websites (see [1], [3]) cannot produce satisfied search results for users. The common reason is that they only rely on text search. When the query word is mountain, the users' intention might not be the landmark with its name having mountain, but the associated travel information such as travelogues and photographs, have the characteristics of mountain. So both texts and images are useful for searching the landmark based on its characteristics. Additionally, when a user wants to search the popular landmarks in the visiting city, the popularity ratings become useful. Therefore, in a smart travel guide system, heterogeneous tourism information should be thoroughly exploited, and furthermore, the semantic gap between heterogeneous information and users' needs should be bridged.

To address the above issue, our smart travel guide system is considered as a search problem instead of a recommendation problem. Users express their intentions as queries and the system makes use of representative heterogeneous information to search the candidate landmarks. Moreover, we regard this travel guide search problem as a ranking problem. Specifically, the ranking is performed in a query-dependent way, and the landmark-order is affected by the proposed diverse information fusion strategy (which is introduced in Section IV-B).

The main goal of ranking and reranking is to maximize the satisfaction and minimize the information load. The ranking methods involve queries or contexts as constraints to rank the landmarks, and generate the ranking results based on textual, visual, or both features. The methods on landmark ranking can be classified into global ranking [9]–[11] and local ranking [12], [13]. Global ranking recognizes a landmark mainly as a whole, which synthesizes the heterogeneous information of the landmark and then conduct ranking. By contrast, the local ranking usually gives a partial view of a landmark, such as showing a representative view via photographs. In our travel guide system, landmarks are expressed by heterogeneous information, including texts, images, and digits. The landmarks are ranked by fusing different aspects of heterogeneous tourism information. In addition, the graph-based ranking is utilized to measure the similarity between the landmarks.

Heterogeneous information fusion contains two levels of fusion issues [14]–[17]: 1) early and 2) late fusion. Early fusion focuses on exploring the features of different modalities and fusing different features at the feature-level. Similarly, in the proposed landmark ranking system, feature fusion is conducted among the heterogeneous information. The fusion algorithms used in our system are inspired by [18], which mines latent semantic features from visual features. In that work, the travel information is composed of visual and textual information, both of which relate to landmarks. For this reason, our previous algorithm [18] suits well to mine the relations between these modalities. However, the difference of the current work from [18] is that we mine the latent topics from heterogeneous information instead of visual features. The challenges are shown as follows.

1) Travel multimedia information is heterogeneous and unstructured in different dimensions, and therefore is difficult to analyze. Structuration and feature extraction of the heterogeneous information are necessary yet challenging.
2) The community-contributed heterogeneous information can offer more reliable information of different modalities, which is, however, difficult to analyze for the fusion purpose.

In the proposed system, latent semantic analysis (LSA) is employed to mine the latent links among visual and textual information. Specifically, we calculate the latent links via multitopical latent features, by which the semantic similarity of landmarks can be measured. To the best of our knowledge, this is the first attempt to use multitopical latent features in landmark ranking systems, and also the first attempt to integrate textual and visual information into travel heterogeneous information fusion strategy.

Considering the above ideas, we construct a novel landmark ranking system which is a branch of travel guide systems. This system can implement the query-dependent landmark

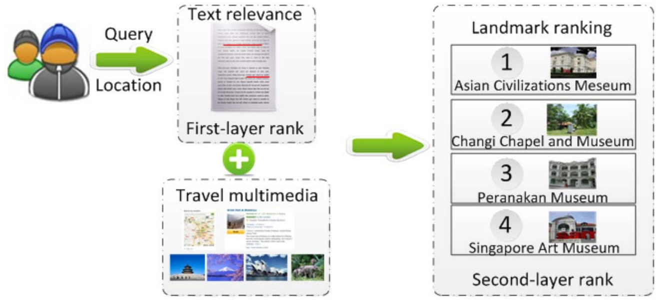


Fig. 1. Information flow of our landmark ranking system.

ranking based on heterogeneous tourism information fusion. Fig. 1 illustrates the information flow of our landmark ranking system. Given a query, an initial ranking list is obtained based on textual matching. A second-layer ranking list is then produced by our system. Since heterogeneous information provides multiple views of landmarks, travel multimedia, including visual information, textual information, and ratings, is involved. To analyze heterogeneous travel information, the proposed system is assessed on a large collective dataset which contains the famous landmarks in five cities. For each landmark, photographs from Flickr [19], reviews and user ratings from Tripadvisor [2], and textual introduction from Wikitravel [3] are used to construct our system. Experimental results demonstrate that our system is well established as a smart landmark ranking system, and the algorithm used in our system has a pleasing performance.

Our main contributions include: 1) comprehensive heterogeneous travel information is collected from social media; 2) heterogeneous information is fused to compensate the deficiency of each individual modality; and 3) the appropriate landmarks can be ranked in a query-dependent way and then reranked based on the heterogeneous information fusion strategy.

The rest of this paper is organized as follows. Section II briefs the related work. Then, in Section III, we identify the overview of our travel guide system. The details of our system are introduced in Section IV. Experiments and discussion are presented in Section V. Section VI gives the conclusion.

II. Related Work

In recent years, various travel guide systems have been developed [20] based on prolific online travel multimedia. Wikitravel [3] is the first travel guide interactive website in the era of Web 2.0. It can provide users with timely travel information at the landmark-level. The development of social media stimulates the boom of multimedia sharing, and a large volume of accessible travel multimedia can be spread widely, such as travelogues, YahooTravel [1], and Flickr [19]. Thus, travel multimedia is an exciting source for the research of travel guide systems.

One type of travel guide systems focuses on the inner-representation of landmarks. For example, VirtualTour [22] is an online travel system, which aims to provide high-quality image representations from famous photograph sharing websites. DiverseSearch [5] analyzes the geo-referred photographs in Flickr and diversifies the search results to visualize the landmarks. Personalized landmark summary is generated based on user queries in personalized-MM [22], which utilizes the travel multimedia, such as texts, images, and videos collected from cross-domain social media. In [23], a novel framework for image classification is proposed, which utilizes the labeled landmark images to construct a 3-D model and discovers the hot region images of the landmarks, and then classifies the unlabeled images into the landmark categories.

Another type of travel guide systems targets at searching or recommendation problems based on travel multimedia. For instance, by analyzing the photographs from Flickr and the knowledge from YahooTravel [1], W2GO [8] recommends the top three landmarks in a city and gives summaries of the landmarks to users. photograph2Trip [24] explores the travel destinations and routes between landmarks based on the geo-tagged photographs and travelogues, and makes the travel route plan for travelers. gTravel [24] is a social interactive travel system, which assists travelers to plan trips and to share travel information. Ji *et al.* [25] presented a mobile landmark search framework in which the photographs can be transmitted in a compact way and then multiviews of the landmarks can be shown to users. Moreover, in Min *et al.*'s [31] work, a landmark can be recognized and searched based on low quality photographs. The recognized landmark is shown via a 3-D view formed by photograph collections.

A travel guide system should be able to leverage travel heterogeneous information and provide the relevant results based on queries or users' profiles. Thus, ranking and reranking are crucial steps in search or recommendation systems. Currently, many reranking methods have been proposed, including classification-based [26], [27], [46], clustering-based [13], and graph-based [28], [29] methods. In the landmark ranking process of Gao *et al.*'s [8] work, the landmarks are ranked by analyzing the cross-domain travel knowledge. However, this method considers a recommendation problem, which shows the top popular landmarks without user interaction. A personalized travel recommendation [22] utilizes both locations and geo-tagged photographs to rerank popular landmarks. Another line of research concerns the ranking based on the interpretation of landmarks. Visual ranking is an attempt in travel guide systems. For example, Kennedy and Naaman [6] proposed to represent the landmarks by visual features, which can be used for reranking. In order to have a quicker view of landmarks, Ren *et al.* [30] learned from the community photographs, and created an outline of content as a summary of the landmarks by reranking the image search results. Ye *et al.* [4] proposed the DLMSearch System which employs an image query to retrieve the diverse landmarks without geo-tags, and ranks the obtained images to summarize the interested landmark. In addition, an important application in travel guide systems is route recommendation, which can be seen as a landmark ranking problem. Lu *et al.* [24] employed photographs to discover the landmarks, and the proper routes among these landmarks are recommended by the ranking principles. The differences between our method and the above mentioned literatures are: 1) we propose a travel guide system which focuses on heterogeneous information fusion to help users make decisions and 2) the ranking model is proposed based on heterogeneous information to satisfy users' needs.

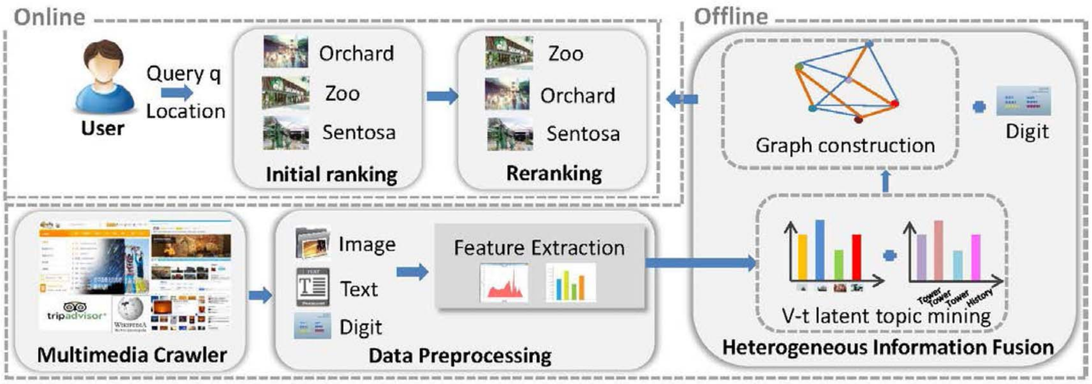


Fig. 2. Framework of the travel landmark reranking based on heterogeneous information fusion.

III. Our Travel Guide System

A. System Overview

Concerning the development in social media and the problems in tourism, a system is constructed to improve the status quo after the analysis of why current systems lag behind. Our proposed landmark ranking system can assist in searching for landmarks with heterogeneous information, which provides the convenience of users. The general framework of our travel landmark search system is illustrated in Fig. 2. First, travel guide information is crawled from cross-domain social media as a data collection. Second, images, travelogues, and ratings are structured and analyzed in data preprocessing where the latent knowledge can be mined. Then, given a city location, a user can input a concise word (e.g., historical) about the characteristics of landmarks to meet his/her specific demands. The system will rank the initial landmark list based on textual information. Finally, we will further analyze travel heterogeneous information from cross domains, and the similarities between the landmarks are measured according to different kinds of heterogeneous information. The fusion of the heterogeneous information from user-generated contents is conducted to refine the preliminary results. Moreover, our system will use the graph-based ranking model based on heterogeneous information fusion to rank a set of candidate landmarks. The following sections will detail each step.

B. Problem Definition

Our landmark ranking system is formulated as a decision making system [32], [33], which is performed by combining heterogeneous information for landmark ranking and reranking. Our system is based on the fact that heterogeneous information uploaded to social media are valuable information sources for analyzing and ranking the landmarks in a query-dependent way.

Each landmark contains travel heterogeneous information and can be seen as a multimedia entity, denoted as L_i. Specifically, L_i can be represented as a collection of visual, textual, and digital information, that is, $L_i = \{M_i, T_i, D_i\}$, where M_i, T_i, and D_i represent the sets of images, texts, and digits, respectively. These definitions facilitate us to formulate a multimedia-entity ranking problem, which can be solved by the graph-based ranking model. Regarding this, an undirected

TABLE II
List of Key Notations

Natation	Definition
$\mathcal{L} = (L_1, L_2, \dots, L_n)$	$\mathcal{L}$ denotes the landmark set, and L_i denotes the *i-th* landmark
$\mathcal{M}_i = \left(M_1^i, \dots, M_m^i, \dots\right)$	$\mathcal{M}_{\mathrm{i}}$ denotes the image set of the *i-th* landmark
$\mathcal{T}_i = (T_1^i, \dots, T_m^i, \dots)$	$\mathcal{T}_{\mathrm{i}}$ denotes the text set of the *i-th* landmark
$\mathcal{D}_i = (D_1^i, \dots, D_m^i, \dots)$	$\mathcal{D}_{\mathrm{i}}$ denotes the digit set of the *i-th* landmark
$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{w})$	$\mathcal{G}$ denotes a graph, and $\mathcal{V}$, $\mathcal{E}$, and $\mathcal{w}$ denote the set of vertices, the set of edges and the weights of the edges, respectively.
f_i^v	Visual feature vector of the *i-th* landmark
f_i^t	Textual vector of the *i-th* landmark
$r = (r_1{}', r_2{}', \dots, r_n{}')^T$	Initial ranking scores of the landmarks
$r = (r_1, r_2, \dots, r_n)^T$	Reranking scores of the landmarks

graph is constructed according to the similarities between different images, and the final ranking result is obtained based on the optimization of the graph. In the following explanations, we employ the notations and definitions of the elements listed in Table II.

The travel guide system should satisfy the following rules.

1) Heterogeneity: The similarities between the landmarks should be measured based on different kinds of information. The dissimilarities between landmarks can be defined as

$$\mathrm{Dist}\left(L_i,\ L_j\right) = \sqrt{(F_i - F_j)^2} \tag{1}$$

where F_i denotes the feature of the ith landmark.

2) Ranking: The ranking problem can be formulated via the Laplacian regularization technique [26], which is shown as

$$Q(q, r) = R(r) + \mu \mathrm{Dist}\left(r, r'\right) \tag{2}$$

where r' is the initial ranking list based on the query q, and r is the reranking list. $R(r)$ is the loss function that ensures the similar landmarks to be ranked together, and $\mathrm{Dist}(r, r')$ is the penalty function that controls the differences between the initial ranking list and the reranking list. The definition of $\mathrm{Dist}(r, r')$ is introduced in Section IV-D.

Following these rules, the problem of heterogeneous information fusion is addressed by generating a mixed feature via LSA [34]. Inspired by [18], [34], and [35], textual and visual features are integrated into the multitopical latent features. In addition, the problem of creating the reranking list is formulated as the minimization of $Q(q, r)$ based on the ranking assumptions.

IV. System Details

A. Heterogeneous Information Mining

The collection of travel heterogeneous information is crawled from cross-domain social media. The famous landmarks are extracted from the landmark list of Wikitravel [3]. The images with tags and geo-tags are crawled from Flickr, which are regarded as the visual view of landmarks. Simultaneously, to form the textual view of landmarks, travelogues and reviews are downloaded from Wikitravel and Tripadvisor. Moreover, user ratings and landmark click ratings are also used as the digital information to compute the popularities of landmarks. All these information will be structured for further processing.

1) Image Feature Extraction: The image set $\mathcal{M}_i = (M_1^i, M_2^i, \ldots, M_m^i, \ldots)$ is regarded as the visual view of landmark L_i. To exploit visual contents in each image set, bag of visual words [36] is employed for landmark visual description. The first step in building a bag-of-words representation is to extract visual features from each image. In our case, we apply dense sampling with the vertical and horizontal step sizes of 10 pixels, and the image pyramid is created by a scale factor of 1.2. A scale-invariant feature transform (SIFT) descriptor [37] is computed for each region of 41×41 pixels such that the obtained local features are scale-invariant and rotation-invariant. In this way, L_i is represented by an $n_v \times 1$ feature vector f_i^v, where $f_i^v(w^v, 1)$ denotes that landmark L_i expresses the word w^v. While SIFT features are employed in this paper, we emphasize that any additional features could be used for specific purposes, such as gist and color moments.

2) Text Feature Extraction: The text set $\mathcal{T}_i = (T_1^i, T_2^i, \ldots, T_m^i, \ldots)$ composes textual information. To building the vocabulary, we list all occurred words (all texts) which have been used more than once and then, we filter all the words by checking whether they are known by Wordnet [38], which is a lexical database of English. If a word exists in Wordnet, it will be assimilated into the final vocabulary, otherwise it will be discarded. In our model, landmark L_i is represented by an $n_t \times 1$ textual feature vector f_i^t, where $f_i^t\left(w^t, 1\right)$ denotes that landmark L_i conveys the word w^t.

3) Popularity Extraction: Besides the content-based textual and visual features of heterogeneous information, popularity and user ratings are also important and noteworthy. Popularity is determined by two components: 1) the number of the photographs uploaded by users about landmark L_i (denoted by n_i^v) and 2) the number of the online comments about L_i (denoted by n_i^t). To normalize the popularity score, we define the maximal number of the photographs as Max^v, and the maximal number of online comments as Max^t. The popularity score of landmark L_i is calculated as

$$s_i^{\text{pop}} = \alpha^{\text{pop}} \frac{n_i^v}{\text{Max}^v} + \beta^{\text{pop}} \frac{n_i^t}{\text{Max}^t} \tag{3}$$

where α^{pop} and β^{pop} are the weighting factors.

4) User Ratings: User ratings can be utilized to assess the inner satisfaction of a landmark. Typically, a user rating is determined by the number of positive views in Tripadvisor (denoted as n_i^{pos}), and the average rating of landmark L_i (defined as R_i). The following equation defines user rating score of landmark L_i:

$$s_i^r = \alpha^r \frac{R_i}{\text{Max}^r} + \beta^r \frac{n_i^{\text{pos}}}{N^{\text{all}}} \tag{4}$$

where Max^r is the maximal value of the ratings, N^{all} is the total number of the ratings, and α^r and β^r are the weighting factors.

B. Heterogeneous Information Fusion

The landmarks are described by multisources including not only textual and visual information, but also digital information, and hence the heterogeneity appears. Moreover, the aforementioned low-level features of the landmarks cannot express the high-level semantics. It is thus necessary to exploit high-level topic concepts from heterogeneous travel information. In this section, we will introduce how to extract multitopical latent features by fusing textual and visual features, and how to dig the latent relations between them. And then, we will discuss how to combine the content-based information and the digit-based information.

1) Multitopical Latent Feature Mining: Multitopical latent feature mining intends to find a dimensionality-reduced expression of multiple features. LSA [39] is competent for this task, which cannot only create vector-based representations of documents, but also can exploit the low-dimensional concept space. Mathematically, an $n \times m$ matrix is denoted as $\mathbf{B}$, which can be disposed by singular vector decomposition

$$B = \text{TSD}^T \tag{5}$$

where both $\mathbf{T}$ and $\mathbf{D}$ have orthonormality, and $\mathbf{S}$ is the diagonal matrix which contains the singular values of $\mathbf{B}$. To reduce the dimensionality of the concept representations, we select k largest singular values and the corresponding singular vectors, which compose a new matrix $\mathbf{S_0}$. The matrix B can be approximated as

$$B \approx B^{'} = T_0 S_0 D_0^T = \Gamma D_0^T. \tag{6}$$

The corresponding optimization problem is expressed as

$$\min_{\Gamma, D_0} \left\| B - \Gamma D_0^T \right\|_F^2 + \gamma \|D_0\|_F^2 \tag{7}$$

where $\| \cdot \|_F$ is the Frobenius norm, γ is a positive value, and the matrix $\mathbf{D}_0 \in R^{m \times k}$.

Suppose we have two topical feature matrices. The one is an $n \times p$ visual matrix, denoted as $\mathbf{B}$, and the other is an $n \times l$ textual matrix, denoted by $\mathbf{C}$. p and l represent the dimensions of visual and textual features, respectively. Both $\mathbf{B}$ and $\mathbf{C}$ are expected to share the same representation Γ [35] such that the similar structures of the matrices can be preserved. Therefore, the optimization problem [35], [40] can be written as follows:

$$\sum_{\Gamma, D_B, D_C} \left(\left\| B - \Gamma D_B^T \right\|_F^2 + \gamma \left\| D_B \right\|_F^2 + \lambda \left(\left\| C - \Gamma D_C^T \right\|_F^2 + \gamma \left\| D_C \right\|_F^2 \right) \right) \tag{8}$$

where $\lambda = 0.5$ is the constant to balance the losses for two features, and γ is set to 0.2.

C. Query-Dependent Initial Ranking

A query-word dataset about the travel information is established based on Wordnet [38]. First, the stop words are eliminated by employing [41]. Then, the query is refined by an automatic error detecting and correcting process for word spelling [42]. Finally, the query is mapped to the query-word dataset. In the mapping, the query is matched with the textual information of landmarks which is collected from the tags of Flickr and the travelogues of Wikitravel [3]. The probability of the query is determined by using the language model of [43]. As a result, the initial ranking list with textual query dependency is generated.

D. Graph-Based Reranking With Heterogeneous Information

A connected graph [28], [29], [44] is constructed to reveal the similarities between the landmarks. We have the assumption that the similar landmarks are likely to be ranked together in the query-dependent way. Given a query, if the neighbors of a landmark in the graph are similar to the query, the landmark is regarded as relevant to the query. In addition, the reranking sequence should be compatible with the initial ranking sequence, since textual cues exploited by the initial ranking cannot be ignored.

The accuracy of the ranking results could be improved by involving travel heterogeneous information, such as user's comments and photographs. Here we formulate the reranking task as a graph-based reranking problem, where the content-based similarities between landmarks are measured by multitopical latent features. Specifically, the k-NearestNeighbor (kNN) graph is used, which has a great ability to capture the local structures of data. Even though the kNN graph needs an expensive construction cost, our system only ranks popular landmarks, the number of which is very limited. Suppose there are n landmarks $\mathcal{L} = (L_1, L_2, \ldots, L_n)$, which have been ranked by the initial text search based on a query. Let $G = (V, E)$ be a graph, where each node in V represents a landmark and each edge weight ω in E represents the similarity between two landmarks. As listed in Table II, the initial ranking score vector is $r = (r_1, r_2, \ldots, r_n)^T$ and the reranking score vector is $r' = (r'_1, r'_2, \ldots, r'_n)^T$. $r(L_i, q)$ denotes the ranking score of L_i when the query is q. We formulate the weight between L_i and L_j by kNN as

$$\omega_{ij} = e^{-\|\Gamma_i - \Gamma_j\|^2 / 2\sigma^2} \tag{9}$$

where the heat kernel is controlled by σ, $\Gamma_i(\Gamma_j)$ denotes multitopical latent feature of the landmark $L_i(L_j)$. $\mathbf{W} = \{\omega_{ij}\}$ is an $n \times n$ matrix, and $D_{i,i} = \sum_j \omega_{i,j}$ composes a diagonal matrix. Assuming that the similar landmarks will be ranked together, the loss function is then represented by

$$R(r) = \sum_{i,j=1}^{n} \omega_{ij} \left\| \frac{r(L_i, q)}{\sqrt{D_{i,i}}} - \frac{r(L_j, q)}{\sqrt{D_{j,j}}} \right\|. \tag{10}$$

Another assumption is that the initial ranking list should constrain the reranking list, which leads to the following penalization:

$$\mathrm{Dist}\left(r, r'\right) = \sum_{i=1}^{n} \left\| r_i - r'_i \right\|. \tag{11}$$

Finally, given a query q, the whole loss function for optimization is defined as

$$Q(r, q, G) = \frac{1}{2} \sum_{i,j=1}^{n} \omega_{ij} \left\| \frac{r(L_i, q)}{\sqrt{D_{i,i}}} - \frac{r(L_j, q)}{\sqrt{D_{j,j}}} \right\| + \mu \sum_{i=1}^{n} \left\| r_i - r'_i \right\| \tag{12}$$

where μ is the regularization parameter to balance the ranking loss and the penalization.

The minimization of $Q(r, q, G)$ results in an optimal r as a closed-form solution [29]

$$r = (\mathbf{I} - \varepsilon \mathbf{S})^{-1} r' \tag{13}$$

where $\mathbf{I}$ is an identity matrix of the size $n \times n$, $\varepsilon = (1/1 + \mu)$ and $\mathbf{S} = \mathbf{D}^{-(1/2)} \mathbf{W} \mathbf{D}^{(1/2)}$ is the normalized Laplacian matrix. This solution represents the content-based reranking score vector r, and we denote it as S_{content} in the below.

User ratings and popularity also influence the users for decision making, and result in a different ranking score vector S_{rating}. Here, $S_{\text{rating}} = s^{\text{pop}} + s^r$, where s^{pop} and s^r are calculated using (3) and (4). To fuse digit-based travel knowledge, S_{rating} should be considered in the content-based reranking list. Intuitively, more popular and favorite landmarks should be ranked higher, but this should not affect the content-based reranking significantly. Thus, a parameter ρ is introduced to adjust the reranking list slightly. After normalization, the reranking score is calculated as

$$(1-\rho) S_{\text{content}} + \rho S_{\text{rating}} \tag{14}$$

where $0 \leq \rho \leq 1$. The content-based reranking can be seen as a weak reranker afforded by each partial view. Heterogeneous information fusion can also be considered as a way of producing a strong feature by collecting the weak ranking information. In this sense, ρ is important for balancing S_{content} and S_{rating}. If $\rho = 0$, we obtain the content-based reranking. Conversely, $\rho = 1$ results in the ranking based on user ratings and popularity. In this paper, we focus on an important composite reranking with the empirical value of $\rho = 0.2$.

V. Experiments

A. Design of Dataset

We build the test collection from the social media website Flickr, and the travel guide websites Wikitravel, Yahoo Travel guide, and Tripadvisor. Five cities are selected which include Singapore, Beijing, New York, Paris, and London. In each city, 120 top-popular landmarks are identified via referring the landmark lists published on travel guide websites. The collections of heterogeneous information are described as follows.

1) Text Data: We crawled all travelogues of the selected landmarks from Wikitravel, and all comments from Tripadvisor, to form our text-based dataset.

2) Image Data: Images of each landmark are collected from Flickr and Tripadvisor. The name of the landmark is employed as the query to search images. The images of each landmark are crawled from Flickr by using the Flickr's public application programming interface, and 4000 top ranked photographs are saved. The algorithm to retrieve the photographs concerns the similarities between the query and the to-be-matched photographs. To enrich the image data, we also collected all photographs from Tripadvisor.

3) Query: The landmark search is simulated by enquiring landmark queries with different kinds of users' needs. Each query is correlated with specific characteristics of landmarks, which could simulate real-world retrieval situations. Users can give queries to satisfy their needs. After eliminating the stop words, the query-word list is collected based on a diverse set. The diverse set includes two subsets of words, one of which includes the characteristics of landmarks, while the other contains the presentations of landmarks. Text, including tags of the photographs from Flickr and comments from Tripadvisor, are used to calculate the word frequency statistics. We select the travel-related words with high frequency as queries. In specific, the characteristics of landmarks are the words like "house," "park," "zoo," "garden," "beach," "bridge," "children," "library," "mall," "water," "theater," "mountain," "clubs," "island," and "tower." The presentational words of landmarks are listed as "art," "entertainment," "museum," "historic," "national," "peaceful," "ancient," "happy," "great," "hot," "military," "walking," "natural," and "hiking."

4) Ground-Truth: To evaluate the performance, we build the ground-truth set manually. Three subjects (two females and one male), who had either been in the target city for more than two years or been travelled to the target city as travel hobbyists, were invited to label the ground-truth. The detailed labeling procedure is described as follows. The subjects did not know what kinds of information could reflect the ranking results, and they were just required to judge whether the returned landmarks were related to the query. An introduction was conducted before the labeling work such that each subject understood the procedure. The subjects were asked to read the briefings of the landmarks in each city to make sure that they were familiar with the landmarks. The subjects should search the given queries one by one. In each search round, the subjects were asked to label the returned top 15 landmarks with the degree of the relationship between the query and each landmark. The degrees were divided into three levels including "very relevant," "relevant," and "irrelevant." For example, when the query is park, a list of landmarks was given by the system. The subjects would grade each landmark according to the predefined levels.

B. Methodology

1) Precision and Recall: The percentage of the feedbacks in the system which are labeled as positive landmarks is defined as precision. The percentage of user-labeled landmarks which are correctly detected by our system, is defined as recall. Labels with relevant and very relevant are considered as positive, and the irrelevant are considered as negative. Precision and recall are defined in terms of true positive results (TR), wrong positive results (WR), and labeled true results (LR), which can be calculated, respectively, as

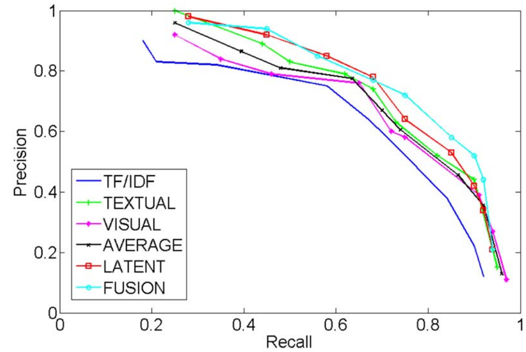


Fig. 3. Precision-recall curves of different settings.

$$\text{precision} = \frac{\text{TR}}{\text{TR} \bigcap \text{WR}} \tag{15}$$

$$\text{recall} = \frac{\text{TR}}{\text{LR}}. \tag{16}$$

2) Top-k Accuracy: Top-k accuracy should also be estimated to evaluate our ranking system. Normalized discounted cumulative gain (NDCG) [45], which measures the relevance of the returned results, is employed to estimate the top-k accuracy. The ground-truth is the manually labeled feedback. NDCG is defined as

$$\text{NDCG}@k = C_k \sum_{i=1}^{k} \left(2^{\text{rel}(i)}/\log_2(1+i)\right) \tag{17}$$

where rel(i) is the scaled relevance level of the ith ranked image, and C_k is a constant to normalize the value of NDCG@k.

C. Results and Analysis

To evaluate the proposed framework, we study from two aspects: 1) effectiveness and 2) efficiency. Both objective and subjective experiments are conducted to evaluate the effectiveness of our system. The efficiency of our system is also discussed. Examples are provided to illustrate our system visually in the final analysis.

1) Effectiveness of the Multitopical Latent Feature Generation: An effective feature generation plays an important role in the performance of the system. By employing LSA, multitopical latent features are generated from heterogeneous information. The performance of the fused features is compared with that of each single view feature, under the following baseline settings.

a) Ranking with term frequency/inverse document frequency (TF/IDF): Given a query, term frequency and inverse document frequency rank the landmarks' relevance based on the tags.

b) Graph-based reranking: The graph-based ranking is to learn a ranking function, which is defined by the relevance between queries and landmarks.

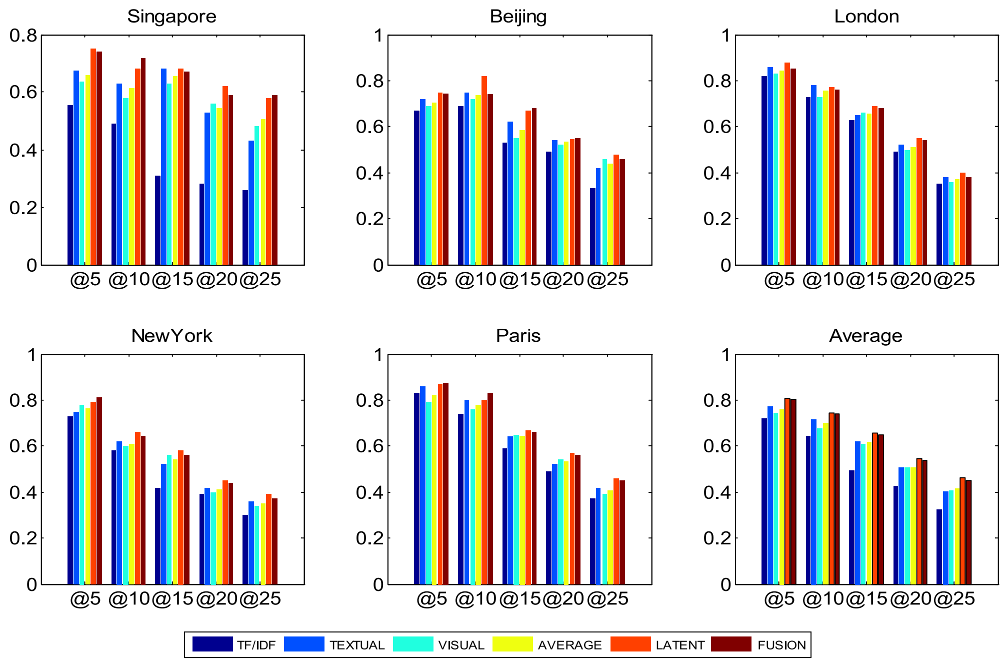


Fig. 4. NDCG@{5–25} performances of our system.

1) *Reranking With Textual Information (TEXTUAL):* We employ the textual information of Tripadvisor to estimate the reranking list with graph learning.
2) *Reranking With Visual Information (VISUAL):* We use the visual information (i.e., images) from Web sources to rerank the landmarks.
3) *Reranking With Averaged Fusion (Average):* The textual information and the visual information are averaged to test the linear feature fusion.
4) *Reranking With Latent Content (LATENT):* In this method, the multitopical latent features are learned by our model and used for the graph-based reranking.
5) *Reranking With Heterogeneous Fusion (FUSION):* In this method, both the latent features and the digital information are considered to rerank the landmarks.

The precision-recall curves in Fig. 3 illustrate the performance comparison of different settings. All reranking curves show better performances than the baseline curve of TF/IDF. The FUSION setting performs better than the others in general. The LATENT-curve and the FUSION-curve are superior to the TEXTUAL-curve and the VISUAL-curve. This indicates that the multimodality fusion is better than any single modality. The performance of LATENT illustrates that heterogeneous information fusion based on content and digital knowledge plays a significant role on the ranking performance. This means that the multitopical latent features contain not only the textual and visual information, but also the latent relations between them. Moreover, the LATENT-curve occasionally performs better than the FUSION-curve. The reason is perhaps that the popularity is uncorrelated with the contents of the textual and visual information.

Fig. 4 illustrates the NDCG@{5–25}, where each index is the average value of candidate queries in different cities. The heterogeneous fusion and the content-based reranking perform better than other algorithms. However, we note that heterogeneous fusion is not always better than the multitopical latent features ranking. This can be explained by the fact that, when the unrelated content information is added, the top-ranked landmarks are not always very relevant to the query. Despite this, most results imply that the landmark reranking based on the heterogeneous information fusion carries more useful information than the other settings. Therefore, our method results in better search accuracy. It is shown that the NDCG@15 accuracy is lower than NDCG@5, because a city only contains 120 landmarks and candidate landmarks based on the query usually have no more than ten landmarks generally.

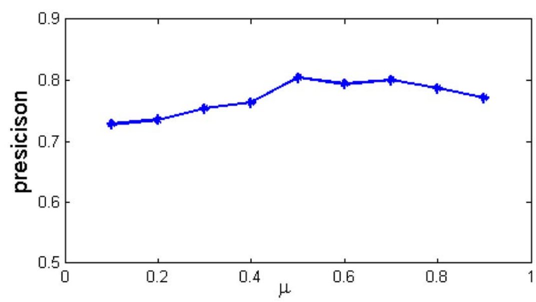
To make a detailed comparison, Table III shows the NDCG@5 results of the designed queries by using our method and the competitors. As observed, the multitopical latent feature achieves better performance than any other single feature, such as the textual feature or the visual feature. This indicates that the multitopical latent features can also describe the latent relations between different features. Notably, it also performs better than the linear combination of the textual and visual information. We see that several queries (e.g., park, beach, and bridge) are easy to extract from the textual information, so their accuracies are higher than that of others. In addition, some queries, such as happy and popular, are abstract, in which case the textual information cannot reveal semantics. Therefore, extracting the multitopical latent features is a good manner to exploit high-level topics. The fusion settings perform well in most queries, but perform poorly in several cases. Two reasons may explain this phenomenon. One is that travel information crawled from the websites is always arbitrary and contains noises which may influence the analysis on landmarks. For example, given a query *children*, we could obtain the landmarks which are suitable for children based on text matching. However, the uploaded photographs hardly have visual cues on the elements of children. Consequently, the textual feature performs better than others for the query *children*.

TABLE III
NDCG@5 RESULTS OF DIFFERENT SETTINGS IN THE CITY OF SINGAPORE, WHERE THE BEST RESULT IN EACH QUERY IS MARKED IN BOLD

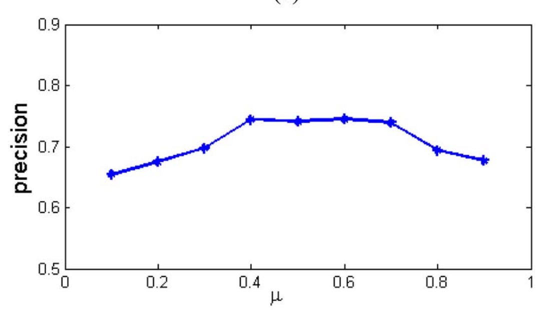
Query	Type	Initial ranking	Graph-based reranking				
			TEXTUAL	VISUAL	AVERAGE	LATENT	FUSION
house	*C*	0.48	0.63	0.73	0.76	0.82	**0.78**
parks	*C*	0.87	0.89	0.92	0.905	0.93	**0.95**
zoo	*C*	0.70	0.69	0.73	0.71	**0.83**	0.82
garden	*C*	0.80	0.79	0.81	0.80	**0.84**	0.83
beaches	*C*	0.56	0.79	0.74	0.765	**0.80**	0.79
bridges	*C*	0.60	0.69	0.57	0.63	**0.75**	0.70
children	*C*	0.43	**0.58**	0.34	0.46	0.43	0.46
library	*C*	0.53	0.59	**0.68**	0.635	0.63	0.60
malls	*C*	0.78	0.86	0.82	0.84	0.87	**0.88**
water	*C*	0.84	0.91	0.84	0.875	**0.92**	0.88
theater	*C*	0.41	0.59	0.37	0.48	**0.64**	0.58
clubs	*C*	0.92	0.96	0.59	0.775	0.96	**0.97**
island	*C*	0.93	**0.96**	0.86	0.92	0.95	0.96
tower	*C*	0.29	0.38	0.57	0.475	**0.62**	0.53
museum	*C*	0.94	0.93	0.92	0.925	0.94	**0.95**
art	*P*	0.84	0.83	0.54	0.685	0.85	**0.86**
recreation	*P*	0.53	0.54	0.53	0.535	**0.56**	**0.56**
historic	*P*	0.58	0.60	0.64	0.62	**0.67**	0.65
national	*P*	0.92	0.69	0.33	0.51	**0.93**	0.90
peaceful	*P*	0.49	0.53	**0.64**	0.585	0.63	0.59
ancient	*P*	0.25	0.34	0.40	0.37	**0.56**	0.50
happy	*P*	0.32	0.45	0.37	0.41	0.51	**0.53**
great	*P*	0.18	0.81	0.84	0.825	0.84	**0.93**
hot	*P*	0.20	0.19	0.32	0.255	**0.37**	0.36
military	*P*	0.67	0.75	0.37	0.56	**0.81**	0.74
walking	*P*	0.31	0.53	0.72	0.625	**0.74**	0.72
hiking	*P*	0.20	0.75	0.71	0.73	0.74	**0.75**
natural	*P*	0.27	0.68	0.84	0.76	**0.88**	0.85
education	*P*	0.24	0.67	0.72	0.695	0.73	**0.76**
mean	/	0.554	0.676	0.637	0.660	**0.75**	0.737

Another reason is that some queries cannot describe the landmarks' content completely or do not have the specific characteristics. For example, national can be found from the textual words, but it cannot be well described by, for example, the visual feature. The landmarks related to the word *national* are different from each other. As a result, a good latent feature cannot be obtained from the multimodality fusion. Furthermore, it is noteworthy that there is a leap in the FUSION ranking list which performs worse than the LATENT ranking, because the digits from social-media cannot describe landmarks on semantic levels. Thus, heterogeneous fusion is probably not the best approach for the content-based evaluation metric. However, the ranking based on heterogeneous fusion is essential in real practice, because the ratings can improve user experiences.

As mentioned in our optimization problem, the parameter μ can affect the reranking results. Fig. 5 shows the effects of μ for presicsion@5 and presicsion@10. While determining the value of μ does not have sufficient theory foundations, the experiments inform us that the best possible value can be found within $0.4 < \mu < 0.6$. Therefore, we set μ to 0.55 in all experiments.



(a)



(b)

Fig. 5. Effects of μ in our reranking system. (a) Precision@5. (b) Precision@10.

2) Effectiveness of the Landmark Ranking System: The proposed landmark ranking system integrates travel heterogeneous information and ranks landmarks in a query-dependent way. More importantly, landmarks are reranked based on heterogeneous information fusion and semantic bridging. To demonstrate the superior search accuracy of our system, Tripadvisor (www.tripadvisor.com) is selected as a competitor. In Tripadvisor, the algorithm used is similar to TF/IDF, and the ranking of landmarks is performed by matching the

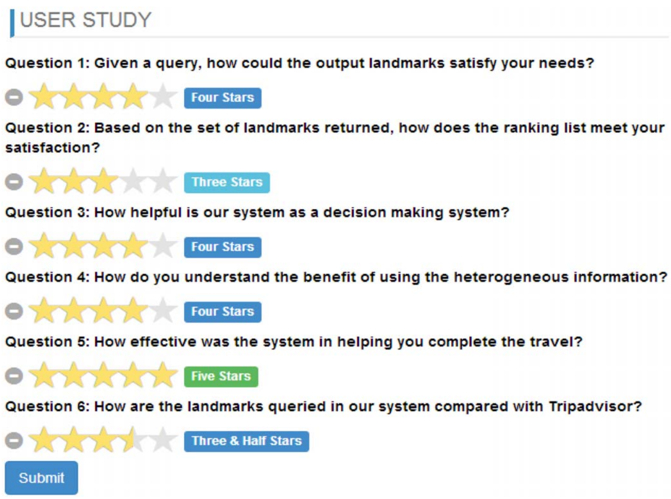


Fig. 6. Interface of user study with six questions.

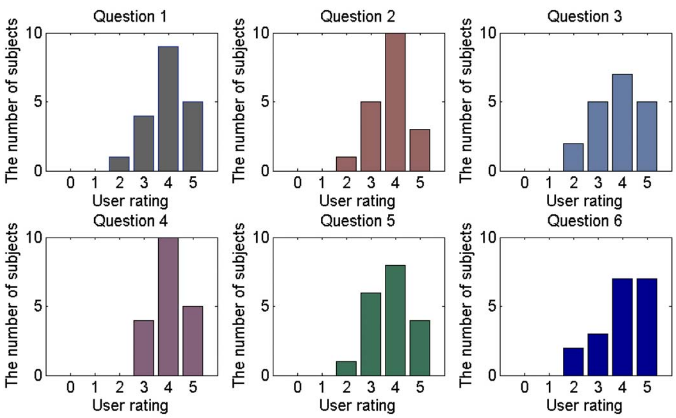


Fig. 7. User study on the evaluated questions.

queries with the textual information. For evaluation, we conduct a user study which conveys several questions to the participated users, as shown in Fig. 6.

In user study, 19 subjects (ten females and nine males), who had either been in the target city for more than two years or traveled to the target city as travel hobbyists, were invited to participate in the evaluation. The subjects were asked to search with short queries which can characterize the styles of distinct landmarks. In each task, the subjects would judge whether each of the top k returned results satisfied their needs. To quantify the relevance between the query and the returned results, the subjects were asked to grade the relevance score. The score ranges from 0 to 2, where 0, 1, and 2 mean irrelevant, relevant, and very relevant, respectively. To evaluate our system, the subjects were required to answer the questions listed in Fig. 6, by providing for each question a rating score ranging from 0 to 5. Table IV shows the explanations of the rating scores. Fig. 7 shows the scores of the questions 1–6 which are graded by all subjects. Question 1 shows that the returned landmarks can satisfy most subjects by query-dependency. Question 2 indicates that the ranking sequence can conform to the subjects. The subjects can reach a decision with the help of our system according to question 3. In question 4, the subjects are requested to experience two kinds of ranking results, which are ranked by content and by heterogeneous information

TABLE IV
EXPLANATIONS OF THE RATINGS

Ratings	Explanations
5	Very good. I'd like to make a decision via the ranking list.
4	Good. I'd like to take into account if visiting.
3	No feelings. I have no feeling about the system, but don't oppose others to choose it.
2	Not very well. It is not a good choice to select this.
1	Bad. I don't like this, and will tell others to give it up.
0	Very bad. I really hate this.

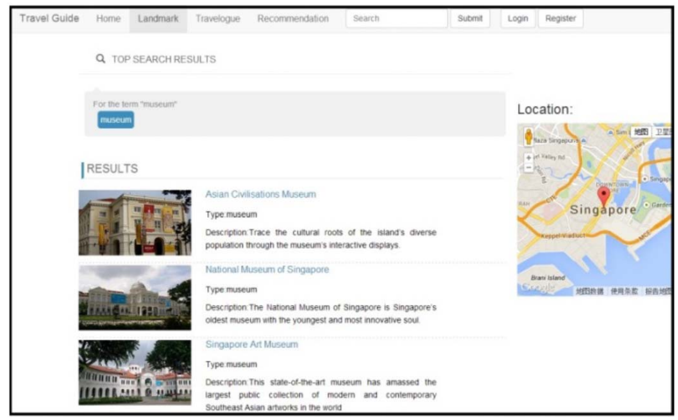


Fig. 8. User interface of our experimental environment.

fusion, respectively. As mentioned in the results of NDCG, although the content-based reranking is more accurate than the heterogeneous fusion-based reranking, user rating knowledge can give users more satisfaction. Question 5 concerns about the effectiveness of our system, where our system is supported by 73.6% subjects (with Scores ≥ 4) but unsupported by 10% subjects (with Scores ≤ 2). In question 6, given a query, if the query cannot match any words with the textual information, no information will be returned to Tripadvisor. However, our system can produce the reranking list based on heterogeneous information even without the textual information matching. As shown in Fig. 7, most subjects voted high scores (with Scores ≥ 4) to our system. These results indicate that a satisfactory system is obtained and users can search information quickly at the landmark-level to satisfy their requirements.

3) Efficiency of the Landmark Ranking System: Efficiency is a key factor in the evaluation of our system, which includes offline and online efficiency. The offline computations represent the heterogeneous information crawling and mining, which are time consuming. In the task of multitopical latent features mining, due to the sparse property of the matrices, the computational complexity only relates to the nonzero elements. Thus, the computational complexity is $O(\varepsilon k)$, where k is the dimension of the latent feature and ε is the number of nonzero elements. Considering the gradient computation of $O(nk^2)$, the final complexity is $O(\varepsilon k + nk^2)$. While the latent feature learning costs plenty of time at the offline procedure, this saves much time for the online applications. The online efficiency is regarded as the most important factor for efficiency. We have 120 landmarks for searching in each city. Considering the online heterogeneous fusion, the reranking computational efficiency is $O(n)$. Hence, the number of the landmarks influences the computational efficiency linearly.

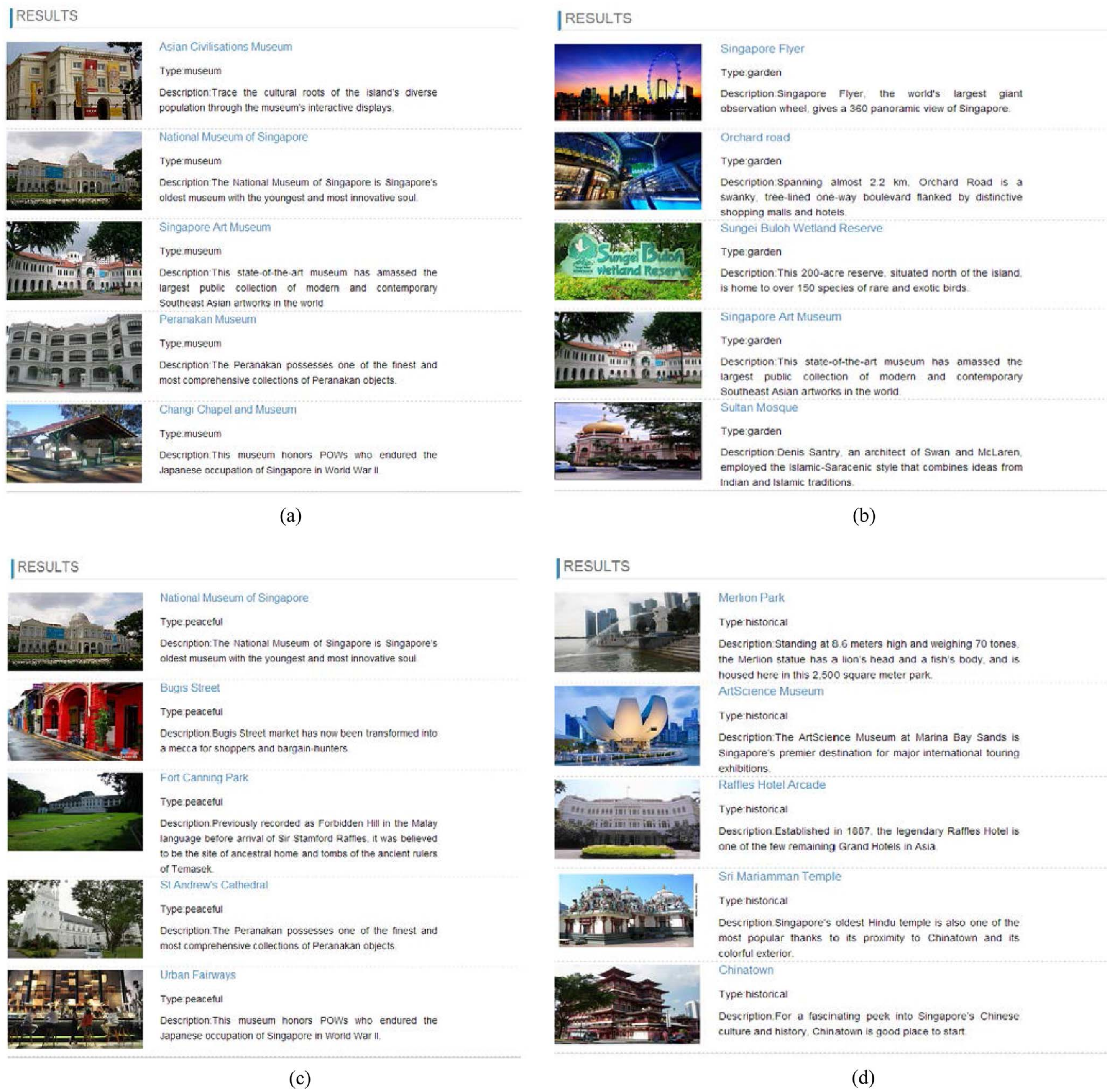


Fig. 9. Top relevant results returned by different queries on NDCG@5 when the location is Singapore. (a) Query: museum. (b) Query: garden. (c) Query: peaceful. (d) Query: historical.

The average CPU running time is tested to evaluate the online efficiency of our system. All experiments are conducted on the platform equipped with a 3.2 GHz CPU (Intel Core i5-3470) and 8 GB memory. In a target city, the online time cost is about 10^{-3} s when there are about a hundred landmarks to be searched.

4) Examples: We further show the expressive performance of our landmark ranking system. Our system differs from the other travel guide systems in that it is not restricted by popularity of the landmarks only, but also by user-queries. Given a query, the proposed system can provide the relevant landmarks which are first ranked by text matching and then reranked based on travel heterogeneous information fusion. The resultant ranking list can preserve the landmarks which conform both the user-query and travel heterogeneous information. In our experiments, the query *museum* and the query *art* are two kinds of queries. One kind is about characteristics and the other is an issue of presentations.

The interface of our system is illustrated in Fig. 8. In this experimental environment, a user can input the query of his/her interests when he/she wishes to seek favorite landmarks in a visiting city. The system ranks and reranks the landmarks based on heterogeneous information, and represents the landmarks to users. Each landmark has an explanation with photographs and concise texts. On the right side, the geo-locations of the returned landmarks can be revealed on the map. Fig. 9 demonstrates the top five results which are obtained with different queries. As shown in the returned landmarks, our system can produce very relevant results, which demonstrates the superior performance of our system. Even though the textual information plays an important role in query-based retrieval, cross-domain diverse media makes the ranking more qualified, in which case the landmarks with similar styles can be ranked together.

Query-words may have high semantic meanings. As shown in Fig. 9(a) and (b), the query-words, *museum* and *garden* easily get reasonable results. Especially when searching for *garden*, even though the results are not shown as expected the names of the landmarks are in the garden style, the top five landmarks contain the visual factor of garden. This is because the visual information has a greater effect on determining the results, that is, the landmarks with similar visual representations are reranked together. In the perspective of heterogeneous information fusion, the reranking of the landmarks is determined not only by the textual information, but also by the visual elements. Therefore, the multitopical latent features have a major impact on the visual information in this kind of query (i.e., garden).

The popularity of the landmarks can influence the reranking results. It causes the landmarks which are popular to be ranked higher, but the landmarks which are not popular to be ranked lower. The Singapore Art Museum is a typical example in the search of *garden*. In addition, the photographs of this landmark are collected from Flickr, and exhibit the square shape, which is consistent with the profiles of gardens. In other words, the contents of photographs are correlated with the elements of gardens from the perspective of visual information. Therefore, both popularity and visual information have positive effects in such a matching case (i.e., "Singapore Art Museum" and garden).

In Fig. 9(c) and (d), the query words are abstract and difficult to be used for retrieval. However, when the query is peaceful, landmarks with this characteristic are discovered. This indicates that our latent features are suitable for mining the abstract concepts. Besides owing to the heterogeneous fusion, the landmark "Raffles Hotel Arcade" is placed in a top position in the query "historical." All these examples demonstrate that our system can bridge the semantic gap between user's intention and travel information, and therefore performs promisingly.

All of the four visualized examples illustrate that the performance of the query-dependent travel guide system is pleasing. On one hand, the multitopical latent features contain not only the textual and visual information but also the latent relationships between them. On the other hand, popularity can improve system performance and can help users make decisions for travel. We observe that the proposed system produces satisfactory results in the content-based landmark search, especially, based on the fusion strategy, which differs from the conventional text-based search.

VI. Conclusion

Travel guide systems aim to provide friendly assistance to users for intelligent travel. An important yet challenging problem for a smart tour guide is how to improve user experiences. In this paper, we have presented a novel query-dependent landmark ranking system based on heterogeneous travel information fusion to facilitate a smart travel guide. First, given a query, the proposed system gets the initial ranking list of landmarks via text matching. Second, the landmarks are reranked based on the proposed heterogeneous travel information fusion scheme. In detail, the multitopical latent features are mined which can not only fuse the textual and visual information seamlessly, but also represent the latent relations between the multimodality features. User ratings are employed as a measure of social popularity to get the final heterogeneous information fusion. Finally, both subjective and objective evaluations indicate the advantages of the proposed system. In future work, we will concentrate on how to improve the efficiency of the system, and how to fuse heterogeneous information. Moreover, videos, as an important source to represent the landmarks, could be introduced for heterogeneous multimedia mining in the future version. Additionally, the context information is also valuable to investigate to implement personalized ranking in a mobile application, and voice commands can be integrated to facilitate landmark searching.

References

[1] (Aug. 30, 2013). *Yahoo Travel*. [Online]. Available: https://travel.yahoo.com

[2] (Aug. 30, 2013). *Tripadvisor*. [Online]. Available: https://tripadvisor.com

[3] (Aug. 30, 2013). *Wikitravel*. [Online]. Available: https://www.wikitravel.com

[4] J. Ye *et al.*, "DLMSearch: Diversified landmark search by photo," in *Proc. 20th ACM Int. Conf. Multimedia*, Nara, Japan, 2012, pp. 905–908.

[5] Q. Hao *et al.*, "TravelScope: Standing on the shoulders of dedicated travelers," in *Proc. 17th ACM Int. Conf. Multimedia*, Beijing, China, 2009, pp. 1021–1022.

[6] L. S. Kennedy and M. Naaman, "Generating diverse and representative image search results for landmarks," in *Proc. 17th Int. Conf. World Wide Web*, Beijing, China, 2008, pp. 297–306.

[7] Q. Hao *et al.*, "TravelScope: Standing on the shoulders of dedicated travelers," in *Proc. 17th ACM Int. Conf. Multimedia*, Beijing, China, 2009, pp. 1021–1022.

[8] Y. Gao *et al.*, "W2Go: A travel guidance system by automatic landmark ranking," in *Proc. Int. Conf. Multimedia*, Florence, Italy, 2010, pp. 123–132.

[9] M. Collins, "Ranking algorithms for named-entity extraction: Boosting and the voted perceptron," in *Proc. Assoc. Comput. Linguist. Conf.*, Philadelphia, PA, USA, 2002, pp. 486–496.

[10] X. Tian *et al.*, "Bayesian video search reranking," in *Proc. 16th ACM Int. Conf. Multimedia*, 2008, pp. 131–140.

[11] H. Zitouni, S. Sevil, D. Ozkan, and P. Duygulu, "Re-ranking of Web image search results using a graph algorithm," in *Proc. Int. Conf. Pattern Recognit.*, Tampa, FL, USA, 2008, pp. 1–4.

[12] L. S. Kennedy and S.-F. Chang, "A reranking approach for context-based concept fusion in video indexing and retrieval," in *Proc. 6th Int. Conf. Image Video Retrieval*, Amsterdam, The Netherlands, 2007, pp. 333–340.

[13] W. H. Hsu, L. S. Kennedy, and S.-F. Chang, "Video search reranking via information bottleneck principle," in *Proc. 14th Annu. ACM Int. Conf. Multimedia*, Santa Barbara, CA, USA, 2006, pp. 35–44.

[14] T. Takehara, S. Miki, N. Nitta, and N. Babaguchi, "Extracting context information from microblog based on analysis of online reviews," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Melbourne, VIC, Australia, 2012, pp. 248–253.

[15] R. Yan and A. G. Hauptmann, "The combination limit in multimedia retrieval," in *Proc. 11th ACM Int. Conf. Multimedia*, Berkeley, CA, USA, 2003, pp. 339–342.

[16] E. M. Voorhees, N. K. Gupta, and B. Johnson-Laird, "Learning collection fusion strategies," in *Proc. 18th Annu. Int. SIGIR Conf. Res. Develop. Inf. Retrieval*, Seattle, WA, USA, 1995, pp. 172–179.

[17] Q. Wu, Z. Wang, F. Deng, Z. Chi, and D. D. Feng, "Realistic human action recognition with multimodal feature selection and fusion," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 43, no. 4, pp. 875–885, Jul. 2013.

[18] J. Shen, T. Mei, Q. Tian, and X. Gao, "Image search reranking with multi-latent topical graph," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Beijing, China, 2013, pp. 1–4.

[19] (Aug. 30, 2013). *Flickr*. [Online]. Available: https://www.flickr.com/

[20] J. Shen, Z. Cheng, J. Shen, T. Mei, and X. Gao, "The evolution of research on multimedia travel guide search and recommender systems," in *Proc. 20th Int. Conf. MultiMedia Model. (MMM)*, Dublin, Ireland, Jan. 2014, pp. 227–238.

[21] F. Jing, L. Zhang, and W.-Y. Ma, "VirtualTour: An online travel assistant based on high quality images," in *Proc. 14th Annu. ACM Int. Conf. Multimedia*, Santa Barbara, CA, USA, 2006, pp. 599–602.

[22] X. Wu, J. Li, Y. Zhang, S. Tang, and S.-Y. Neo, "Personalized multimedia Web summarizer for tourist," in *Proc. 17th Int. Conf. World Wide Web*, Beijing, China, 2008, pp. 1025–1026.

[23] X. Xiao, C. Xu, J. Wang, and M. Xu, "Enhanced 3-D modeling for landmark image classification," *IEEE Trans. Multimedia*, vol. 14, no. 4, pp. 1246–1258, Aug. 2012.

[24] X. Lu, C. Wang, J.-M. Yang, Y. Pang, and L. Zhang, "Photo2Trip: Generating travel routes from geo-tagged photos for trip planning," in *Proc. Int. Conf. Multimedia*, Florence, Italy, 2010, pp. 143–152.

[25] R. Ji *et al.*, "Location discriminative vocabulary coding for mobile landmark search," *Int. J. Comput. Vis.*, vol. 96, no. 3, pp. 290–314, 2012.

[26] R. Yan, A. Hauptmann, and R. Jin, "Multimedia search with pseudo-relevance feedback," in *Proc. 2nd Int. Conf. Image Video Retrieval (CIVR)*, Urbana, IL, USA, Jul. 2003, pp. 238–247.

[27] Y. Liu, T. Mei, G. Qi, X. Wu, and X.-S. Hua, "Query-independent learning for video search," in *Proc. IEEE Int. Conf. Multimedia Expo*, Hanover, Germany, 2008, pp. 1249–1252.

[28] X. Tian *et al.*, "Bayesian video search reranking," in *Proc. 16th ACM Int. Conf. Multimedia*, Vancouver, BC, Canada, 2008, pp. 131–140.

[29] H. Zitouni, S. Sevil, D. Özkan, and P. Duygulu, "Re-ranking of Web image search results using a graph algorithm," in *Proc. Pattern Recognit.*, Tampa, FL, USA, 2008, pp. 1–4.

[30] Y. Ren, M. Yu, X.-J. Wang, L. Zhang, and W.-Y. Ma, "Diversifying landmark image search results by learning interested views from community photos," in *Proc. 19th Int. Conf. World Wide Web*, Raleigh, NC, USA, 2010, pp. 1289–1292.

[31] W. Min, C. Xu, M. Xu, X. Xiao, and B.-K. Bao, "Mobile landmark search with 3D models," *IEEE Trans. Multimedia*, vol. 16, no. 3, pp. 623–636, Apr. 2014.

[32] J. Rees and G. J. Koehler, "Modeling search in group decision support systems," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 34, no. 3, pp. 237–244, Aug. 2004.

[33] V. I. Yukalov and D. Sornette, "Manipulating decision making of typical agents," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 44, no. 9, pp. 1155–1168, Sep. 2014.

[34] H. Deng, M. R. Lyu, and I. King, "Effective latent space graph-based re-ranking model with global consistency," in *Proc. 2nd ACM Int. Conf. Web Search Data Mining*, Barcelona, Spain, 2009, pp. 212–221.

[35] S. Zhu, K. Yu, Y. Chi, and Y. Gong, "Combining content and link for classification using matrix factorization," in *Proc. 30th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, Amsterdam, The Netherlands, 2007, pp. 487–494.

[36] F.-F. Li and P. Perona, "A Bayesian hierarchical model for learning natural scene categories," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, San Diego, CA, USA, 2005, pp. 524–531.

[37] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[38] (Jun. 25, 2012). *Wordnet*. [Online]. Available: http://wordnet.princeton.edu/

[39] D. Scott, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by latent semantic analysis," *J. Amer. Soc. Inf. Sci. (JASIS)*, vol. 41, no. 6, pp. 391–407, 1990.

[40] X. Zhu, J. Lafferty, and R. Rosenfeld, "Semi-supervised learning with graphs," Ph.D. dissertation, School Comput. Sci., Language Technol. Inst., Carnegie Mellon Univ., Pittsburgh, PA, USA, 2005.

[41] W. J. Wilbur and K. Sirotkin, "The automatic identification of stop words," *J. Inf. Sci.*, vol. 18, no. 1, pp. 45–55, 1992.

[42] M. D. Kernighan, K. W. Church, and W. A. Gale, "A spelling correction program based on noisy channel model," in *Proc. 13th Int. Conf. Comput. Linguist. (COLING)*, vol. 2. Helsinki, Finland, 1990, pp. 205–210.

[43] J. M. Ponte and W. B. Croft, "A language modeling approach to information retrieval," in *Proc. 21st Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, Melbourne, VIC, Australia, 1998, pp. 275–281.

[44] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schölkopf, "Ranking on data manifolds," in *Proc. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2004, pp. 169–176.

[45] W. B. Croft, D. Metzler, and T. Strohman, *Search Engines: Information Retrieval in Practice*. Reading, U.K.: Addison-Wesley, 2010, pp. 351–358.

[46] Z. Ji, Y. Pang, and X. Li, "Relevance preserving projection and ranking for Web image search reranking," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4137–4147, Nov. 2015.