

Automatic Image Semantic Annotation Based on the Tourism Domain Ontological Knowledge Base

Pengfei Zhang¹, Junping Du^{1(✉)}, Dan Fan¹, and Yipeng Zhou²

¹ Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia, School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China
{fly2015_zhang, komaconss}@163.com, junpingdu@126.com

² School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China
yipengzhou@163.com

Abstract. In this paper, we proposed a method of automatic image semantic annotation based on the tourism domain's ontological knowledge base. We need to do other things based on the traditional semantic annotation method. Firstly, we need to acquire the names of the scenic spots through image classification. Then we have to build ontological knowledge base on tourism domain and consider the annotation words and the names of the scenic spots as reasoning conditions. At last, we can use the ontological knowledge base to ratiocinate so as to enhance the accuracy of image annotation, and what's more, to associate annotation words with the name of scenic spot so that we can make annotation words more specific.

Keywords: Image annotation · Knowledge base · Tourism ontology · Image classification

1 Introduction

In recent years, with the development of cross media technology, the travel information we got from the internet is not only text information but also contains different types of data. It greatly enriched the source of knowledge and expands the perspective we understand about the tourism information. However, the content of the tourism images is so complicated that the primary problem of image semantic understanding is image semantic analysis. Since the 1980s, the study on the semantic of cross media data has began. Although the technology of text mining based on natural language understanding has made a great achievement, we still face unprecedented difficulties about text mining technology because of the limited feature we can mine [1]. Similarly, semantic learning and recognition of image is currently faced with the problem that how to cross the semantic gap [2-3]. Following is the basic methods of image semantic analysis and automatic annotation [4-5]. The first method is based on the content of cross media data [6]. The second method makes full use of the text information associated with visual data and transforms the problem of visual data into the

problem of text. For example, the Plsa-Words algorithm proposed by Monay belongs to the second method. The third method is automatic image annotation by fusing semantic topics [7]. All these methods depend only on visual data or text data or only can obtain basic elements of the picture and cannot obtain the content what we want.

However, due to the complexity of the tourism image data and there are rich semantic contents contained in images that the traditional annotation method cannot analyze the specific content in the image. So in this paper we propose the method for image annotation tourism based on the tourism domain's ontological knowledge base and do further keywords filtering on this basis. First, SVM-based image classification method was used to obtain the names of scenic spots. Then, build ontological knowledge base according to the information we got from scenic spots [8]. Finally, in order to obtained the specific content keywords and complete the secondary image annotation ,we consider the scenic name and label words as reasoning conditions and reasoning based on the prior knowledge base on the basis of image annotation. This method can be used to extract the specific content contained in the image and has very good effect on analyzing the semantic content of cross media data.

2 Frame Design of Automatic Image Annotation

In this paper, we propose a method of automatic image semantic annotation based on the tourism domain's ontology knowledge base to realize tourism image annotation. The method mainly consists of three parts: SVM-based classification [9], automatic image annotation by fusing semantic topics, and construction and reasoning of tourism domain's prior knowledge base. We consider the result of image classification as inference conditions and make up for the deficiency of the method that uses automatic image annotation by fusing semantic topics. That method can only analyze the obvious content contained in the images but cannot relate to the scenic spot. Good results can be obtained by knowledge base inference and the results we obtained will relate to scenic spot.

As shown in Fig.1, it's the frame of the method that uses automatic image semantic annotation based on the tourism domain's ontological knowledge base.

Following is the basic process of the method:

(1) Annotation for each image with the method that uses automatic image annotation by fusing semantic topics and we can get the keywords correctly represent the basic content in the image.

(2) Classified images according to the names of scenic spots by SVM-based image classification and consider the result we got as one of the inference conditions.

(3) Build ontological knowledge base on travel which contains the name of the scenic spot, location of the scenic spot, features of the scenic spot, entertainments activities in the spot and other information.

(4) Reasoning according to the names of scenic spots we obtained by image classification and the keywords we got by image annotation. By using knowledge base to ratiocinate, we can obtain the keywords which are related to the scenic spot and can express the specific content of the image.

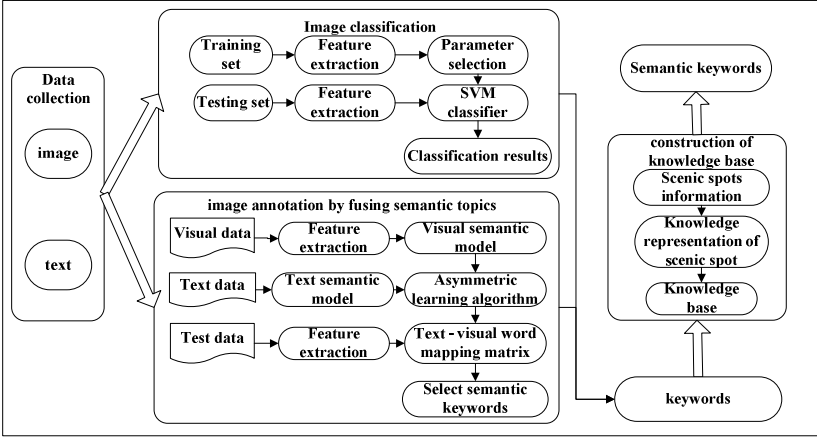


Fig. 1. Frame design of automatic image annotation

3 The Image Annotation Algorithm Based on Reasoning According to Ontological Knowledge

3.1 Automatic Image Annotations by Fusing Semantic Topics

In this paper we model the visual data and text data by probability latent semantic analysis (PLSA) and then we fuse the results of PLSA in which the visual data and text data share the same potential space. The way we fuse the model of visual data and text data is that fusing different distributions of themes we obtained by PLSA with a weight for each image and get a new kind of distribution of themes. The fusion weight of each model is determined by the contribution of the image content which is determined by the entropy of the distribution of visual words.

Suppose that the topic number of visual data is m and the topic number of text data is n , then the model after fusion contains k topics, and $k=m+n$. Using s and t to express the two topics of PLSA model, and then the topics distribution of visual data and text data could be expressed as $P_v(s|d)$ and $P_w(t|d)$. We can get two topics distribution $P_v(s|d_i)$ and $P_w(t|d_i)$ of every image by fusing the two PLSA models. And the topics distribution $P(z|d_i)$ after fusion was determined by the following formula:

$$P(z_k | d_i) = \begin{cases} \alpha_{vi} \cdot P_v(s_k | d_i), & k=1, 2, \dots, m \\ \alpha_{wi} \cdot P_w(t_{k-m} | d_i), & k=m+1, m+2, \dots, m+n \end{cases} \quad (1)$$

Here, α_{vi} and α_{wi} represent the weights of visual data and text data respectively in the image d_i , and the weights can be calculated by the following empirical formula.

The experimental results show good annotation effect when the entropy of the visual words distribution is less than 3 or more than 6, however the effect is not always good when the entropy is between 3 and 6. This is because the images usually contains complex contents, and we cannot fully learn its complexity just rely on the entropy and empirical formula, that is to say, we are unable to determine the most reasonable weight of visual and textual modal data.

Description of the algorithm:

Suppose that there is a training set named $D=\{(d_1,c_1),\dots,(d_N,c_N)\}$ that contains the images and texts, and let $T_D=\{d_1,\dots,d_N\}$ denote the training set of images, and let $L=\{w_1,\dots,w_L\}$ denote the vocabulary list. So the images were included in the training set like $d_i \in T_D$ and the texts were included in vocabulary list like $c_i \in L(i \in 1,\dots,N)$. In addition, we have to suppose that there is a testing set named T_T and $T_T \cap T_D = \emptyset$, $d_{new} \in T_T$.

Following is the description of the training algorithm which is used to model the data included the training set D and learns the association between image and text.

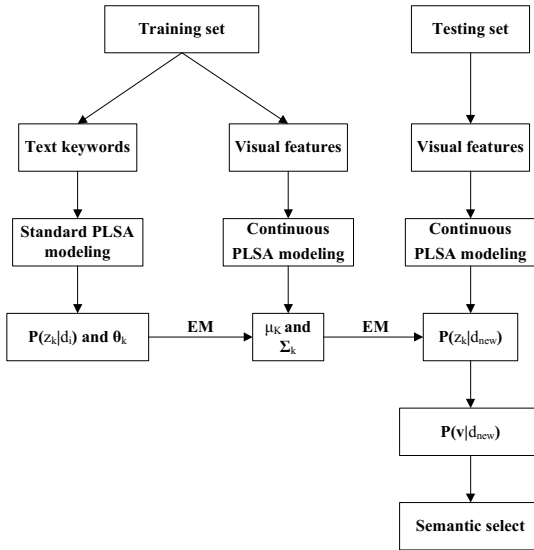


Fig. 2. Automatic image annotation algorithms by fusing semantic topics

(1) Extract the visual features from each image $d_i \in T_D$ and $d_{new} \in T_T$, quantized the features in order to denote the visual words with $v(d_i)$. Similarly, process the text c_d associated with the image d_i and denote the text words with $w(d_i)$.

(2) Fuse the results of the two PLSA models which are respectively based on the denotation of visual words $v(d_i)$ and the denotation of text words $w(d_i)$ and we can get the following results: $P_v(v|s)$, $P_v(s|d)$ and $P_w(w|t)$, $P_w(t|d)$.

(3) In order to measure the importance of the data in visual modality and textual modality, we introduce the fusion parameters α_{v_i} and α_{w_i} . Calculate the fusion parameters by empirical formula and we can get the result $P(z|d_i)$ after fusing the topic distribution $P_v(s|d_i)$ and $P_w(t|d_i)$ by the formula (1).

(4) According to the fused topic distribution $P(z|d_i)$, calculate the final training results $P(v|z)$ and $P(w|z)$ by using EM algorithm.

(5) Calculate the topic distributions $P(z|d_{new})$ of images with EM algorithm by using the denote of visual words $v(d_{new})$ of image d_{ne} and the parameter $P(v|z)$ which is the result of training algorithm.

(6) Calculate the posterior probability of every keyword in the vocabulary list L by the following formula.

$$P(w|d_{new}) = \sum_{k=1}^k P(w|z_k)P(z_k|d_{new}) \quad (2)$$

(7) Select several keywords with maximum posteriori probability to annotate the image d_{new} .

3.2 The Construction and Reasoning of Tourism Ontological Knowledge Base

In this paper, we achieve the further reasoning annotation of image annotation results by constructing tourism repository and make the results more associated with scenery spots and more accurate as well.

At first, we need to define the class structure of the ontology of tourism. In order to meet the requirements of this article, we only need to consider the traveling and entertainment in the tourism which is usually include eating, accommodation, transportation, traveling, shopping, and entertainment. And we also need to know the human and geography knowledge about the scenic spots. So we build four concept classes including scenic spots, scenic characteristics, geographic location, and characteristic activity. Among them, scenic spots usually include the Summer Palace, the Palace Museum, the Great Wall, the Temple of Heaven, the Olympic park and JiuZhaiGou. Scenic spot features include mountain, water, tree, sky, palace, construction, bridges, ships, tourists. Different scenic spots have certain difference. Characteristic activity includes some festival celebration activities and sports activities hold by some scenic spots, etc. The activities may also have obvious difference in different scenic spots.

Secondly, we need define the class attribute of tourism ontology. In order to complete the work in this paper, we need to define all the attributes of different ontology classes. The affiliation among scenic spots: for example “Kunming Lake” is sub attraction of “the Summer Palace”, the relationship between “Kunming Lake” and “the Summer Palace” is affiliation. The “locate in” relationship between scenic spots and geographical location: “the Summer Palace” is located in the “Haidian District of Beijing city”. The containment relationship between scenic spots and their features: for example, the features of the Summer Palace are mountains and water, “the Summer Palace” and “mountains” have the containment relationship. With the above relationships, we labeled the tourism ontological knowledge base manually and then we can complete the subsequent reasoning work based on these relationships.

There are two inputs of the algorithm: one is the test image, and the other one is annotation words list which contains 10 keywords that have larger posterior probability.

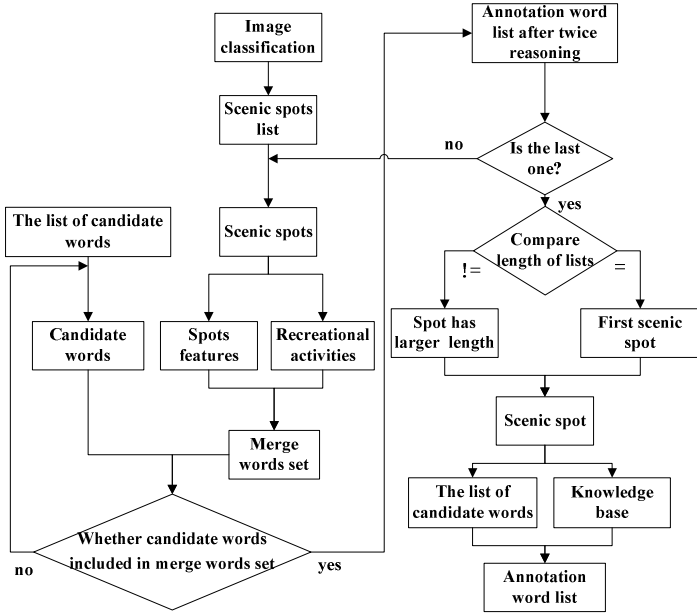


Fig. 3. The reasoning process

The algorithm process is as follows:

(1) Use image classification according to the scenic spots by SVM-based method and select two categories with larger membership degree of class label. Consider these two categories as a scenic spot list.

(2) Choose one spot sequentially from the scenic spot list and query spot features and recreational activities from the knowledge base according to the spot name.

(3) Traverse the candidate annotation word list and judge if a candidate annotation word is included in the spot features or recreational activities, if it is, add the word to the new word list, and if not, read the next word.

(4) Judge whether the spot is the last one of the scenic spots list or not, if it is, execute step (5), and if not, return to step (2).

(5) Compare the length of the annotated word lists which were obtained according to different spots. If the length of all annotated word lists are equal, the first spot in the scenic spots list is the result, or we should choose the spot has longer annotated word list as the result.


(6) Choose the spots corresponding to the candidate annotation word list and combine with the relation between different spots, the relation between spots and the features of all the spots and the relation between spots and recreational activities. Then ratiocinate to acquire the final annotation results.

4 Experimental Result Analysis

The experimental data is collected from Baidu tourism and Chanyouji blogs which include the travel data of the Summer Palace, the Great Wall, the Imperial Palace, the Temple of Heaven, the Olympic Park and so on, as well include the texts and pictures of their sub-attractions. The data is divided into a training set and a testing set. The former set data contains 3500 pictures of the Summer Palace, the Imperial Palace and their sub-attractions, while the latter data set contains 700 images which try to cover all the data included in the former data as much as possible.

We need to set the number of latent topics in the process of using PLSA to establish the main model, which determines the capacity of a PLSA model. If the topic number is too small, the PLSA model we obtained can't fully express the internal training of the data; on the contrary, if the topic number is too large, the efficiency of the system will be greatly reduced. What's more, with the increasing number of the unknown parameters of the PLSA model, the possibility of over fitting will increase. After the validation of experiments, this experiment used 100 latent topics to learn text modal data information and 120 potential subjects to learn the visual modal data. After the fusion of information, there are 220 potential themes obtained.

Table 1. Experimental Results of the two times image annotation algorithm

Figures				
Once annotation	bridge, boat, tree, water	building, sky, palace, mountain	Great Wall, tourists, mountain, water	mountain, Tower of Buddhist Incense, sky, water
Twice annotation	Seventeen Arches Bridge, Summer Palace, Kunming Lake, sky	sky, palace, Imperial Palace, Piled Elegance Hill	Great Wall, tree, sky, tourists	Tower of Buddhist Incense, Summer Palace, Longevity Hill, sky

From table 1 we concluded that after the second mark, the annotation results that we got seems to be more concrete, and it is associated with specific features in scenic spots and easier to understand. In terms of performance of the annotation, it can be measured with the accuracy of the annotation. For a given semantic keywords w_q , the accuracy was denoted as $\text{precision} = B/A$, that A means the number of all figures marked with w_q automatically; B indicates the number of images tagged with w_q correctly. In order to know whether the method we proposed is better than other methods, we annotate the test dataset with Plsa-Words algorithm. The accuracy comparison of the result for images annotation is showed in Fig. 4, which is fused with semantic theme and obtained from the tourism ontology knowledge inference. Table 2 shows the accuracy comparison of three annotation algorithms, where the one-time annotation represents automatic image annotation algorithm by fusing semantic topics and the two-time annotation represents the reasoning process based on knowledge after the one-time annotation.

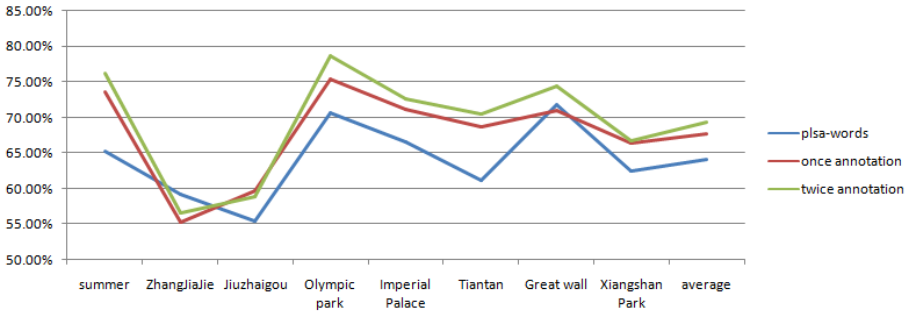


Fig. 4. The accuracy rate comparison of three annotation algorithms

Table 2. Contrast between the accuracy of three annotation algorithms

Scenic spots	Summer Palace	Zhang-JiaJie	JiuZhai-Gou	Olympic Park	Imperial Palace	Tian-Tan	Great Wall	Xiang-shan Park
Plsa-words	0.653	0.591	0.554	0.706	0.665	0.612	0.717	0.625
one-time	0.735	0.553	0.597	0.754	0.711	0.687	0.709	0.664
Two-time	0.762	0.566	0.589	0.787	0.726	0.705	0.744	0.667

With the inference method based on knowledge base, the average accuracy of image annotation is improved from 67.63% for one-time annotation to 69.33% for two-time annotation. Obviously, it's better than 64.04% for Plsa-Words algorithm. However, from the graph we can see that the accuracy of all scenic spots is improved except ZhangJiaJie and JiuZhaiGou. Through analysis, the reason may relates to two points: on the one hand, these two scenic spots are both Natural scenic spot and the scenery elements are complex and contain many kinds of scenery which may lead to multiple results, thus affect the annotation results; on the other hand, the style of scenic spots has a great influence on the annotation results. For example, the different sub attractions have similar style in JiuZhaiGou and it is also very difficult to distinguish even for human. So it will inevitably have a certain impact on the annotation results.

5 Conclusions

Image annotation techniques are the key technology of image semantic analysis. And it plays an important role at the time when our mobile Internet has a high speed of development. Though there is a great progress about image annotation techniques in the world, there are still some limitations. In this study, we put forward an image annotation method relies on reasoning mechanism based on the tourism ontology repository which is on the basis of the image annotation method that fuses the semantic theme. This method combines the characteristics of tourism data, and has some

innovations about the image annotation method the predecessor put up with. At last, we improve the accuracy of image annotation by using this method. And at the same time, by using the tourism ontology repository, the tagging results can combine with specific scenery spots instead of independent general content elements of image. However the accuracy of the results by using this method still has some room to improve. We can obtain more efficient experimental results by building a more complete training set, detailing the classification of the training set, expanding the tourism repository and improving the reasoning model.

Acknowledgments. This work was supported by the National Basic Research Program of China (973 Program) 2012CB821200 (2012CB821206), the National Natural Science Foundation of China (No. 61320106006), Beijing Excellent Talent Founding Project (2013D005003000009).

References

1. Liu, J.: Semantic Analysis and Classification of Food Safety Emergencies Cross Media Information. Beijing University of Posts and Telecommunications (2013)
2. Tang, J., Zha, Z., Tao, D., Chua, T.S.: Semantic-Gap-Oriented Active Learning for Multilabel Image Annotation. *IEEE Transactions on Image Processing* 21(4) (2012)
3. Bahmanyar, R., Datcu, M.: Measuring the semantic gap based on a communication channel model. In: *Image Processing (ICIP)* (2013)
4. Bakalem, M., Benblidia, N., Ait-Aoudia, S.A.: Comparative image auto-annotation. *Signal Processing and Information Technology (ISSPIT)* (2013)
5. Bao, H., Xu, G., Feng, S., Xu, D.: Advances in the technology of automatic image annotation. *Computer Science* (2011)
6. Li, X.: Technology research and analysis of massive image semantic retrieval. School of computer science and technology Zhejiang University (2009)
7. Li, Z., Si, Z., Liu, X.: Image semantic annotation method of modeling continuous visual features. *Journal of Computer Aided Design & Computer Graphics* (2010)
8. Heiyanthuduwege, S.R., Schwitter, R., Orgun, M.A.: Towards an OWL 2 profile for defining learning ontologies. In: *Advanced Learning Technologies (ICALT)* (2014)
9. Saxena, U., Goyal, A.: Content-based image classification using PSO-SVM in fuzzy topological space. In: *Computer and Communication Technology (ICCCT)* (2013)