Automatic Web Image Annotation via Web-Scale Image Semantic Space Learning

Hongtao Xu[1], Xiangdong Zhou[1], Lan Lin[2], Yu Xiang[1], and Baile Shi[1]

[1] Fudan University, Shanghai, China
{061021054,xdzhou,072021109,bshi}@fudan.edu.cn
[2] Tongji University, Shanghai, China
linlan@mail.tongji.edu.cn

Abstract. The correlation between keywords has been exploited to improve Automatic Image Annotation(AIA). Differing from the traditional lexicon or training data based keyword correlation estimation, we propose using Web-scale image semantic space learning to explore the keyword correlation for automatic Web image annotation. Specifically, we use the Social Media Web site: Flickr as Web scale image semantic space to determine the annotation keyword correlation graph to smooth the annotation probability estimation. To further improve Web image annotation performance, we present a novel constraint piecewise penalty weighted regression model to estimate the semantics of the Web image from the corresponding associated text. We integrate the proposed approaches into our Web image annotation framework and conduct experiments on a real Web image data set. The experimental results show that both of our approaches can improve the annotation performance significantly.

1 Introduction

Automatic Image Annotation(AIA) has attracted a great deal of research interests [11,7,6,10,13], due to its critical role in keyword based image retrieval and browsing. However, the long lasting *Semantic Gap* problem still challenges the effectiveness of AIA. It is urgent to improve the annotation performance to meet the increasing requirement of practical applications.

Recently, the correlation between annotated keywords was explored to improve the performance of image annotation. For instance, keyword set {sky, grass} usually has a larger probability to be an image caption than {ocean, grass}. Only a few work had been done to investigate the keyword correlation on AIA, such as CLM [8] and WordNet-based approaches [9,16]. The former employed the co-occurrence of keywords indirectly by using EM algorithm to fit a language model for generating annotations, while the latter made use of WordNet to exploit the hierarchy of the keywords. Zhou [22] proposed an iterative image annotation approach which learn the keywords correlation by "Automatic Local Analysis". Rui et al. [13] presented a bipartite graph reinforcement model (BGRM) for image annotation, which exploit the keywords semantic correlation

Q. Li et al. (Eds.): APWeb/WAIM 2009, LNCS 5446, pp. 211–222, 2009.

based on a large-scale image database maintained by themselves. In general, most of the previous work infer the correlations between keywords according to the co-occurrence of keywords in the training set or the hierarchy of lexicon. However, for the scenario of annotating Web images, the keywords correlation estimation is more subtle and complicated in some extent, due to the problems of the unlimited number of keywords and the intrinsic diversity of Web data space.

To improve the performance of the Web image annotation, we exploit Web Social Media, that is, we use the popular Web Photo Community site Flickr [1] as a Web-Scale Image Semantic Space to learn the keywords correlation graph. Then we propose a novel Web image annotation approach, which incorporates the keywords correlations and the semantic contributions of visual features and associated texts of Web image. Our method conducts the probability estimation using not only Web image textual and visual features, but also the semantic correlations between the keywords and annotated keyword subset assigned previously. In particular, for the Web scale semantic space learning, we submit semantic keywords of the annotation vocabulary to Flickr to obtain the Relative Tag (RT)[1] set as the neighborhood for keyword graph generation. We estimate the contribution of the textual features in deriving the semantics of Web image by a new constraint piecewise penalty weighted regression model. The keyword which brings the maximum annotation conditional probability is selected to be added into the annotation set. Experiments on 4,000 real Web images data set demonstrate the effectiveness of the proposed Web AIA approach.

Our contributions are as follows:

1. We exploit the popular Web Photo Community site Flickr as a Web-Scale Image Semantic Space to analyze the correlations between keywords and incorporate it into our annotation framework.
2. We propose a new constraint piecewise penalty weighted regression model to combine the adaptive estimation of the weight distribution of associated texts and the prior knowledge together for estimating the semantic contributions of the textual features of Web image.

The rest of the paper is organized as follows. Section 2 introduces the related work. Section 3 presents our Web image annotation framework. We discuss the experiment results in Section 4. Section 5 concludes this paper.

2 Related Work

In recent years, much attention has been paid to the research on AIA. Various of machine learning techniques or statistical models have been employed to develop a variety of AIA models, which can mainly be divided into two

[1] RT can be obtained by using Flickr's APIs: flickr.tags.getRelated. It returns "a list of tags 'related' to the given tag, based on clustered usage analysis "–refer to: http://www.flickr.net/services/api/flickr.tags.getRelated.html

categories–probabilistic model based methods and classification based methods. The first category focuses on inferring the correlations or joint probabilities between images and annotation keywords. The representative work include Translation Model(TM) [4], CMRM [7], CRM [10], MBRM [6], etc. The classification based methods try to associate keywords or concepts with images by learning classifiers. Methods like SVM-based approach [3] and Multi-instanced learning [21] fall into this category. However, these approaches do not focus on annotating Web images and neglect the available textual information of Web images, so they cannot be applied directly to annotate Web images.

Sanderson and Dunlop [14] were among the first to model image contents using a combination of texts from associated Web pages, however, they modeled the contents as a bag of keywords without any structure information. Wang et al. [20] proposed a search-based annotation system–AnnoSearch. This system requires an initial keyword as a seed to speed up the search by leveraging on text-based search technologies. Li et al. [11] proposed a search and mining framework to tackle the AIA problem. Given an unlabeled image, content-based image retrieval(CBIR) was firstly performed to find a set of visually similar images from a large-scale image database. Then clustering was performed to find the most representative keywords from the annotations of the retrieved image subset. These keywords, after saliency ranking, were used to annotate the unlabeled images eventually. Its annotation performance was highly dependent on the result of CBIR. Feng et al. [5] described a bootstrapping framework by adopting a co-training approach involving classifiers based on two orthogonal set of features–visual and textual. Tseng et al. [18] built two models based on image visual and textual features, and weighted them to annotate the unlabeled Web images. Xu et al. [19] presents a Web Image Semantic Analysis (WISA) system to adaptively model the distributions of the semantic labels of the web image on its surrounding text.

Some previous work demonstrated that keyword correlations can be utilized to improve the performance of image annotation. Jin et al. [8] address the problem by using EM algorithm to fit a language model to generate an annotation keyword subset. However, the annotation speed is lower due to the EM algorithm. Munirathnam et al. [16] propose a hierarchical classification approach for image annotation. They use a hierarchy induced on the annotation words derived from WordNet. Jin et al. [9] make use of the knowledge-based WordNet and multiple evidence combination to prune irrelevant keywords. Zhou et al. [22] proposed an iterative image annotation approach which learn the keywords correlation by "Automatic Local Analysis". Wang et al. [2] annotate image in the progressive way which explore the keywords correlation by the co-occurrence of keywords in the training images. Tang et al. [17] propose a graph-based learning approach SSMR to measure the pairwise concept similarity. However, these approaches usually learn the keywords correlations according to the appearance of keywords in the training set or lexicon, and the correlation may not reflect the real correlation for annotating Web images. Recent years, with the rapid development of Web social knowledge network, the applications which exploit the manually

tagging resources of Web image, such as Flickr, have attracted researcher's great interests [15]. In this paper, we propose to learn the keywords correlation graph by exploiting Web social knowledge network Flickr.

3 The Web Image Annotation Framework

3.1 The Overview of Web Image Annotation Framework

For a given training set L_{train}, each labeled image $J \in L_{train}$ is demoted by $J = \{W, V, T\}$, where the annotation keywords W is a binary annotation keyword vector indicating whether a keyword is the annotation of J; V is a set of region-based visual features of J; and $T = \{T_1, T_2, \ldots, T_n\}$ is a set of the types of associated texts.

Our annotation framework annotates Web images in an iterative way [22,2]. That is, given a new image I, the annotation keywords set after $i-1$ iterative is denoted as $AN_{i-1}(i = 1, \ldots, k, AN_0 = null$, where k is the size of the annotation keywords set). In the i^{th} iterative, the probability of keyword w to be annotated for I is:

$$P(w|I, AN_{i-1}) = \frac{P(w|I)P(w|AN_{i-1})}{P(w)} = \frac{P(w|I_V, I_T)P(w|AN_{i-1})}{P(w)}, \quad (1)$$

where I_V and I_T is the visual and textual feature of image I respectively. Assuming that $P(w)$ is uniformly distributed, and I_V and I_T are independent, we have:

$$\begin{aligned} w_i^* &= argmax_w P(w|I)P(w|AN_{i-1}) \\ &= argmax_w P(w|I_V)P(w|I_T)P(w|AN_{i-1}). \end{aligned} \quad (2)$$

Then

$$AN_i = AN_{i-1} \cup w_i^* \quad (3)$$

Note that the maximum likelihood estimation for $P(w|AN_i)$ is:

$$P_M(w|AN_i) = \frac{\#\{J|w, AN_i \in J\}}{\#\{J|AN_i \in J\}}, \quad (4)$$

where $\#\{J|w, AN_i \in J\}$ denotes the number of images in which keyword w and keywords subset AN_i appear together. For a limited training set, when $|AN_i|$ is large, the co-occurrence of w and AN_i is rare, which means there will be many zero values in the probability estimation. However, a zero probability event in the limited training set does not mean it never happen in the future, thus smoothing is necessary.

In the text information retrieval, smoothing is usually performed by making use of a large background collection to assign a non-zero probability to the unhappened event in current model. For instance, we can choose a larger training image set for smoothing. However, it is hard to obtain sufficient training images

for the Web-scale image annotation task. Therefore, we propose to explore the Web Social Multimedia to infer semantic correlation, rather than maintaining a large scale image database by ourselves, or using the limited training image set to generate the keywords correlation graph. It is expected that Web-scale semantic space learning is more flexible to deal with the scalability problem of Web images annotation.

Denoting the keywords correlation graph as Sim, then we can smooth the maximum likelihood estimation for $P(w|AN_i)$ by the keyword semantic similarity graph [12] as follows:

$$P(w|AN_i) = (1-\gamma)P_M(w|AN_i) + \gamma \sum_{v \in V} \frac{Sim(w,v)}{Degree_+(v)} P(v|AN_i), \tag{5}$$

where γ is the smoothing factor. V is the vertex set of the graph Sim. $Degree_+(v)$ is the outside degree of vertex v in Sim, that is:

$$Degree_+(v) = \sum_{u \in V} Sim(v,u) \tag{6}$$

Different keywords have different importance for smoothing, here $Degree_+(v)$ captures the importance of keyword v, that is, if keyword v only associates to few keywords, then v is more important for smoothing than those associating to more keywords. Eqn.5 shows that the more similar between keyword v and w, the more important of keyword v for smoothing the probability of w.

The visual generation probability $P(w|I_V)$ is computed as the expectation over the images in the training set, that is:

$$P(w|I_V) \propto P(w, I_V) = \sum_{i=1}^{|T|} P(w, I_V|J_i)P(J_i) = \sum_{i=1}^{|T|} P_V(I|J_i)P(w|J_i)P(J_i), \tag{7}$$

where $P_V(I|J_i)$ is the probability of I being generated from J_i based on their visual features. $P(w|J_i)$ denotes the probability of word w generated from J_i, which can be estimated by maximum likelihood estimation. And we assume $P(J)$ is uniformly distributed.

Based on the assumption that the regions of image are independent each other, $P_V(I|J_i)$ equals to the product of the regional generation probabilities. The regional generation probability $P_V(f_j|J_i)$ can be estimated by non-parameter kernel-based density estimation [10].

3.2 The Keyword Correlation Graph Generation by Web Semantic Space Learning

The directed keyword correlation graph is denoted by $Sim =< V, E >$, where the vertex set V consists of all the annotation keywords, and a directed edge from keyword w to keyword w' is denoted by $e_{ww'} \in E$ which is established if and only if $Sim(w, w') > 0$, where $Sim(w, w')$ is the similarity between w and

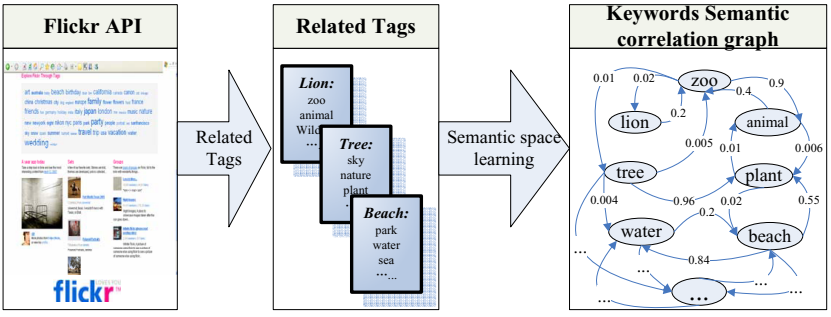

Fig. 1. The generation process of keywords correlation graph. Only part of the vertices and edges are given.

w'. Note that in our keyword correlation graph, $Sim(w, w')$ may not equal to $Sim(w', w)$.

Figure 1 gives the generation process of keywords correlation graph. Firstly, we submit the annotation vocabulary to Filckr to retrieve a neighborhood Image Semantic Subspace (ISS), which is composed of the returned Related Tags (RT)(or concepts). In order to further explore the correlations between the pairs of concepts(keywords) that cannot be directly obtained from Flickr, we build a directed graph $G =< V', E' >$ to represent the semantic correlations obtained from Flickr directly, where the vertex set V' consists of the keywords in ISS, and a directed edge from keyword w to keyword w' is denoted by $e_{ww'} \in E'$ which is established if and only if $w' \in RT(w)$, where $RT(w)$ is the set of Related Tags (RT) of w.

The definition of the directed graph G shows that, if concept (keyword) w' is similar to w, then there exists a path from vertex w to w'. If concept(keyword) w is accessible from w_1 and w_2 in G, and the number of the accessible concept(keyword) of w_1 is larger than w_2, then we can conclude that the similarity between w_1 and w is smaller than the one between w_2 and w. Therefore we can estimate the semantic similarity between w and w' as follows:

$$Sim(w, w') = e^{-dis(w,w') \times \frac{|acess(w)|}{|ISS|}}, \tag{8}$$

where $dis(w, w')$ refers to the distance from w to w' in graph G, which is measured by the length of the shortest path from w to w' in graph G, $acess(w)$ is the set of the accessible concepts(keywords) from w, and $|acess(w)|$ and $|ISS|$ denotes the number of keywords in $acess(w)$ and ISS respectively.

3.3 The Estimation of Textual Generation Probability $P(w|I_T)$

The Object Function. We adopt the linear basic expansion model for estimating the textual generation probability $P(w|I_T)$. Let $H(T)$ denote the set of expansion functions, which represents the associated texts T and their interaction structures; $\omega = \{\omega_1, \ldots, \omega_N\}$ represents the weights of semantic contributions of

$H(T)$ to I. Then the probability $P(w|I_T)$ can be estimated by a linear model as follows:

$$P(w|I_T) = \sum_{j=1}^{N} \omega_j(w) p(w|h_j(T)), \tag{9}$$

where N is the number of original and extended parts of the associated texts.

The Estimation of Probability $p(w|h_j(T))$. We extend the structure of the associated texts to further explore the relationship between image semantic labels and the different types of associated texts. That is, we consider the higher order pairwise structures of the different types of the associated texts by estimating their pairwise joint generation probability $p(w|T_k T_l) = p(w|T_k)p(w|T_l)$, where $(k \neq l) \leq n$ and $p(w|T_i)$ can be estimated by the textual multinomial distribution estimation [22]. Here we define the expansion function set $H(T)$ to represent the associated texts and their higher-order interaction structures identically. For simplicity, we just consider the semantic contributions of the textual data T and their order 2 interaction structures. The probability of keyword w being generated by $h_j(T) \in H(T)$ is estimated as follows:

$$\hat{p}(w|h_j(T)) = \begin{cases} p(w|T_j) & j = 1, \ldots, n \\ p(w|T_i)p(w|T_l) & (i \neq l) \leq n, n < j \leq N \end{cases} \tag{10}$$

The Estimation of Weights $\omega(w, I)$. According to Eqn.9, the weights distribution are crucial to estimate the textual generation probability $P(w|I_T)$. Thus we propose the following constraint piecewise penalized weighted regression model to learn the weight distribution $\omega(w, I)$:

1. For a given unlabeled image I and the corresponding associated texts T, a neighborhood in the training Web image set (denoted as $neighbor(I)$) is first generated under the visual and textual features similarity measurement.

2. Since the textual structures have higher order, we impose different penalty to the associated texts and their pairwise higher order structure. We partition the weight coefficients into k subsets corresponding to T and their $i^{th}(i = 2, \ldots, k)$ order interaction structures. Our aim is to shrinkage the regression coefficients by imposing a L_2 penalty to each part, where the penalty parameters are $\gamma = \{\gamma_1, \ldots, \gamma_k\}(\gamma_1 \geq \ldots \geq \gamma_k)$. Especially, to rectify the statistic error brought by the limitation of the training data, we add the additional prior knowledge into our regression model. Denote the prior knowledge as D_{pre}, which is the set of important dimensions, and the corresponding penalty parameter is γ_{pre}. Then the constraint piecewise penalty weighted regression estimation is defined as follows:

$$\hat{\omega}(w) = arg \min_{\omega(w)} \{ \sum_{i=1}^{K} \mu_i (y_i - \omega_0 - \sum_{j=1}^{N} X_{ij} \omega_j(w))^2 \}$$
$$subjectby : \sum_{\omega_j \in D_s} \omega_j(w)^2 \leq t_s, (s = 1, \ldots, k), \sum_{\omega_j \in D_{pre}} \omega_j(w)^2 \geq t_{pre}. \tag{11}$$

Eqn.11 equals to the following constraint piecewise penalty weighted residual sum of squares:

$$\hat{\omega}(w) = arg\min_{\omega(w)}\{\sum_{i=1}^{K}\mu_i(y_i - \omega_0 - \sum_{j=1}^{N}X_{ij}\omega_j(w))^2 \tag{12}$$
$$+\sum_{s=1}^{k}\gamma_s\sum_{\omega_j\in D_s}\omega_j(w)^2 - \sum_{\omega_j\in D_{pre}}\gamma_{pre}\omega_j(w)^2\},$$

where K is the number of images in $neighbor(I)$, $X_{ij} = p(w|h_j(T))$, y_i refer to the likelihood of the semantic concept w as the label of image J_i (i^{th} Web image in $neighbor(I)$), and T_i denotes the textual features of image J_i, μ_i denotes the similarity between image I and J_i.

4 Experiments

All the data used in our experiments are crawled from Internet. The image data set is obtained by HTML parsing and small icons are filtered out. The size of the image data set L is 4000. We use a heuristic method to generate training set automatically from the download Web pages. The idea is similar to tf/idf heuristic, here we consider two kinds of term frequency, that is, the frequency of the keyword w appears in one type associated text, and the frequency that accounts for the number of the types of associated texts that w appears in. The heuristic rule is that the keyword with higher frequency is more important for the corresponding Web image. At last, we obtain 640 training images, and the rest is used as test set. Each test image is manually labeled with 1-7 keywords as ground truth. The vocabulary of manual annotations consists of about 137 keywords. Each image of L is segmented into 36 blobs based on fixed size grid, and 528 dimensional visual feature for each blob is extracted according to $MPEG7$ standards. Each image associates 5 types of associated texts: image file name, ALT text (ALT tag), caption text (Heading tag), associated text and page title.

We partition half of the training set as validation set to determine the model parameters, such as the smoothing parameter λ, the regularization parameter γ_1, γ_2 and γ_{pre}. Their values are set to 0.6, 0.7, 0.4 and 0.25 respectively in our experiments. The *recall*, *precision* and *F1* measures are adopted to evaluate the annotation performance. That is, given a query keyword w, let $|W_G|$ denote the number of human annotated images with label w in the test set, $|W_M|$ denote the number of images annotated with the same label by our algorithm. The *recall*, *precision* and *F1* are respectively defined as: $Recall = \frac{|W_M \cap W_G|}{|W_G|}$, $Precision = \frac{|W_M \cap W_G|}{|W_M|}$, $F1 = \frac{2(Precision \times Recall)}{Precision + Recall}$. The number of annotation keywords is set to 5, and the average *recall*, *precision* and *F1* over all keywords are calculated as the evaluations of the overall performance.

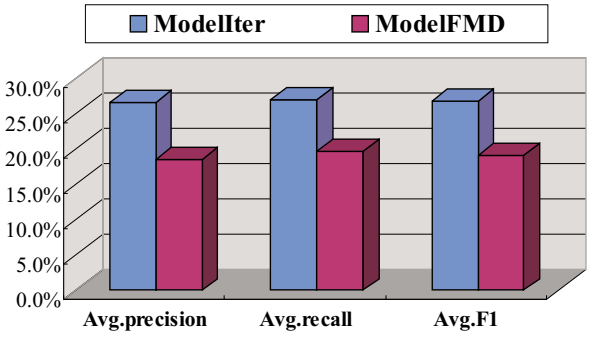

Fig. 2. The overall performance of our Web image annotation approach

4.1 The Overall Performance of Our Annotation Approach

Figure 2 compares the ModelIter approach proposed in this paper and the ModelFMD approach [18]. The ModelFMD model don't use the keywords correlation and learn a fix textual model for estimating the semantic from the associated texts of Web image in the training stage. According to the figure, the performance of our approach is superior to the ModelFMD method significantly. Our annotation framework incorporates two approaches: the keywords correlation and the constraint piecewise penalty weighted regression, we need to test their effectiveness for improving the performance of Web image annotation respectively.

4.2 The Effectiveness of the Web Semantic Space Learning Based Keywords Correlation

To test the effectiveness of our Web semantic space learning based keywords correlation, we compare the performance of our annotation approach which uses the Web concept space learning based keywords correlation(Flickr) and the WordNet-based keywords correlation(WordNet), against the annotation approach without using keywords semantic correlation(noKWCor), that is the probability estimation(Eqn.1) doesn't consider the keywords correlation. Both methods use the constraint piecewise penalty weighted regression model to estimation the semantic contribution of the associated texts, and consider the contribution of visual features. Figure 3 gives the comparison result.

According to Figure 3, the performance of Flickr and WordNet both be superior to noKWCor, which show that the keywords correlation could be an effective way to smooth the maximum likelihood estimation to exploit the keywords correlation in the process of Web image annotation. Meanwhile, we found that the performance of Flickr is better than WordNet significantly, this demonstrates our Web semantic space learning method is more effective than WordNet for measuring the keyword correlation to improve the performance of Web image annotation.

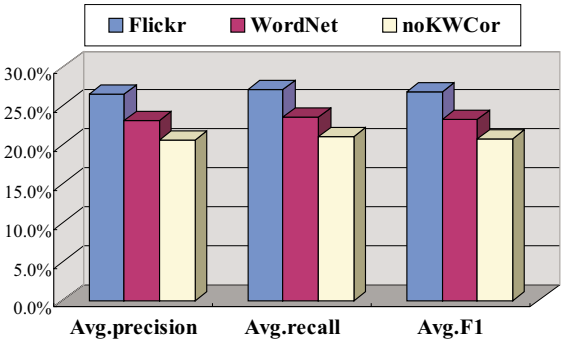

Fig. 3. The effectiveness of keywords semantic correlation

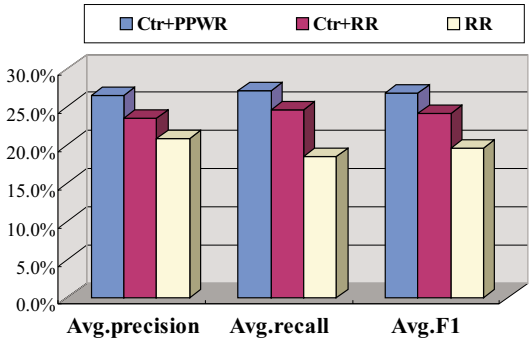

Fig. 4. The effectiveness of prior constraint and Piecewise Penalty Weighted Regression

4.3 The Effectiveness of Prior Constraint and Piecewise Penalty Weighted Regression

To test the effectiveness of prior knowledge constraint(Ctr) and Piecewise Penalty Weighted Regression(PPWR) in the process of associated texts based generation probability estimation, we compare the performance of our approach(Ctr+PPWR) and the approaches which only consider the contribution of Ctr or PPWR. The baseline approach does not consider the prior knowledge constraint and applies Ridge Regression(RR) to learn the weights distribution. All approaches consider the contributions of the keywords correlation and the visual features. Figure 4 gives the comparison result.

The results in Fig.4 show that: (a) The "Ctr+RR" approach is superior to RR approach. This demonstrates that it is effective to impose the prior knowledge constraints in the regress model for Web image semantic annotation. (b) The "Ctr+PPWR" approach is superior to "Ctr+RR" approach. It demonstrates that PPWR algorithm is more effective than ridge regression in learning the weight distribution of the associated texts and their higher order structures when annotating Web images.

5 Conclusions

Ubiquitous image resources on the Web have long been attractive to research community. Web-based AIA is a promising way to manage and retrieve the fast growing Web images. However, its effectiveness still needs to be improved. In this paper, we developed and evaluated a novel automatic Web image annotation approach, which incorporates the image semantic keywords correlations and the semantic contributions of visual features and associated texts of Web image. In particular, we estimate the keywords semantic correlation by using Web image semantic space learning, as well as adaptively model the distribution of semantic labels of Web images on their associated texts by using the proposed constraint piecewise penalty weighted regression. The experimental results demonstrate that both the Web semantic space learning based keywords correlation and the constraint piecewise penalty weighted regression model improve the performance of Web image annotation significantly.

Acknowledgment

This work was partially supported by the Natural Science Foundation of China under Grant No.60403018 and No.60773077.

References

1. http://www.flickr.com
2. Wang, B., Li, Z., Yu, N., Li, M.: Image annotation in a progressive way. In: ICME, pp. 1483–1490 (2007)
3. Chang, E., et al.: Cbsa: Content-based soft annotation for multimodal image retrieval using bayes point machines. CirSysVideo 13(1), 26–38 (2003)
4. Duygulu, P., Barnard, K., de Freitas, J., Forsyth, D.: Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2353, pp. 97–112. Springer, Heidelberg (2002)
5. Feng, H., Shi, R., Chua, T.: A bootstrapping framework for annotating and retrieving www images. In: ACM Multimedia, pp. 960–967 (2004)
6. Feng, S., Manmatha, R., Lavrenko, V.: Multiple bernoulli relevance models for image and video annotation. In: CVPR, pp. 1002–1009 (2004)
7. Jeon, J., Lavrenko, V., Manmatha, R.: Automatic image annotation and retrieval using cross-media relevance models. In: SIGIR, pp. 119–126 (2003)
8. Jin, R., Chai, J., Si, L.: Effective automatic image annotation via a coherent language model and active learning. In: ACM Multimedia (2004)
9. Jin, Y., Khan, L., Wang, L., Awad, M.: Image annotations by combining multiple evidence & wordnet. In: ACM Multimedia, pp. 706–715 (2005)
10. Lavrenko, V., Manmatha, R., Jeon, J.: A model for learning the semantics of pictures. In: NIPS (2003)
11. Li, X., Chen, L., Zhang, L., Lin, F., Ma, W.: Image annotation by large-scale content-based image retrieval. In: ACM Multimedia, pp. 607–610 (2006)

12. Mei, Q., Zhang, D., Zhai, C.: A general optimization framework for smoothing language models on graph structures. In: SIGIR, pp. 611–618 (2008)
13. Rui, X., Li, M., Li, Z., Ma, W., Yu, N.: Bipartite graph reinforcement model for web image annotation. In: ACM Multimedia, pp. 585–594 (2007)
14. Sanderson, H., Dunlop, M.: Image retrieval by hypertext links. In: SIGIR (1997)
15. Song, Y., Zhuang, Z., Li, H., Zhao, Q.: Real-time automatic tag recommendation. In: SIGIR, pp. 515–522 (2008)
16. Srikanth, M., Varner, J., Bowden, M., Moldovan, D.: Exploiting ontologies for automatic image annotation. In: SIGIR, pp. 552–558 (2005)
17. Tang, J., Hua, X.-S., Qi, G.-J., Wang, M., Mei, T., Wu, X.: Structure-sensitive manifold ranking for video concept detection. In: ACM Multimedia (2007)
18. Tseng, V., Su, J., Wang, B., Lin, Y.: Web image annotation by fusing visual features and textual information. In: SAC, pp. 1056–1060 (2007)
19. Xu, H., Zhou, X., Lin, L.: Wisa: A novel web image semantic analysis system. In: SIGIR (2008)
20. Wang, X., Zhang, L., et al.: Annosearch: Image auto-annotation by search. In: CVPR, pp. 1483–1490 (2006)
21. Yang, C., Dong, M.: Region-based image annotation using asymmetrical support vector machine-based multiple-instance learning. In: CVPR, pp. 2057–2063 (2006)
22. Zhou, X., Wang, M., Zhang, Q., Zhang, J., Shi, B.: Automatic image annotation by an iterative approach:incorporating keyword correlations and region matching. In: CIVR, pp. 25–32 (2007)