SPREADING ACTIVATION THEORY BASED IMAGE ANNOTATION

Songhao Zhu[1], Baoyun Wang[1], Yuncai Liu[2]

[1]School of Automation, Nanjing University of Posts and Telecommunications, Nanjing, 210046, China
[2] School of Electronic Information and Electric Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China
{zhush, bywang}@njupt.edu.cn, {homewliu}@sjtt.edu.cn

ABSTRACT

The overwhelming amounts of digital images on the Web and personal computers have triggered the requirement of an effective tool to retrieve images of interest using semantic concepts. Due to the semantic gap between low level content features and its high level semantic features of an image, however, the performances of many existing automatic image annotation algorithms are not so satisfactory. In this paper, a novel approach based on the cognitive science is proposed to improve the quality of annotations. The main idea is that the tags of an image are considered as nodes in a semantic network, and the relevance between the tags and image contents is regulated using the spreading activation theory. After the spreading activation process finishes, each tag will be assigned an appropriate value with respect to its relation to other tags. Experimental results conducted on the 50,000 Flickr images demonstrate that the proposed scheme can effectively improve the performance in automatic image annotation.

Index Terms—Image annotation, cognitive science, spreading activation theory, Flickr images

1. INTRODUCTION

Nowadays, more and more digital images are distributed on the Web and personal computers with the proliferation of digital devices, which demands content-based analysis technologies to effectively organize, manage and utilize such huge amount of information resources. Image annotation, which aims to build an exact correspondence between visual information at the perceptual level and linguistic descriptions at the semantic level, is an elementary step and a promising step for content-based image indexing, retrieval and other related applications.

From the pattern recognition point of view, the task of image annotation can be formulated as a classification problem and completed by machine learning techniques and probabilistic modeling approaches, such as Translation Model [1], Continuous Relevance Model [2], Integrated Multiple/Single Instance Representations [3], Probabilistic Model [4], Dynamic Multi-Scale Cluster Labeling Strategy [5], Probabilistic Collaborative Multi-Label Model [6], Visual Contexts Model [7], etc. However, the annotation qualities of these methods are usually far from satisfactory, due to the well-known semantic gap problem.

To reduce the imprecise and incomplete semantic annotations and better describe the visual content of images, many efforts have been made to complete the issue of annotation refinement. Jin *et al.* [8] estimate the semantic correlation between visual concepts using the word similarity defined by WordNet so as to identify the highly correlated concepts and filter out the weakly correlated concepts. However, this method has not taken into account the visual information clue, and therefore it achieves only partial or limited success. To exploit the visual information clue, a content-based annotation refinement scheme is proposed by Wang *et al.* [9]. This scheme leverages the clues of both visual information and textual information to improve the annotation performance. A regularized latent dirichlet allocation model is presented in [10] to exploit both the statistics of annotations and visual affinities of images. Recently, Liu *et al.* [11] propose to refine candidate annotations by combining visual similarity and semantic similarity in an optimization model. Zhu *et al.* [12] propose to re-rank the annotations for each image by decomposing the user-provided tag matrix into a low-rank refined matrix and a sparse error matrix in a constrained model, where the basic idea is to compute the relevance of an annotation in a constrained yet convex optimization way.

However, most of the current state-of-the-art image annotation methods rarely build and define the correlation between annotations with respect to the brain models in the research field of cognitive science. Motivated by the work on web image retrieval [13] which demonstrates that the spreading activation theory [14] is efficient for estimating and inferring the relevance values of nodes in a semantic network, we propose a novel approach to refine image annotations using the spreading activation algorithm. The candidate annotations of an image are first chosen by a progressive relevance model, which can be here considered as an approximation to the joint probability of multiple words. Then, the significance of each candidate annotation is refined by the propagation procedure based on the spreading activation theory, which can be here considered as a natural choice of computing the relevance of each candidate annotation in a semantic network. The final annotations for each image are the only top ones with the

highest relevance value.

The rest of this paper is organized as follows. Section 2 presents the framework of image annotation. Experimental setup and initial results are discussed in Section 3. Finally, Section 4 offers concluding remarks and some ideas for future research.

2. PROPOSED ANNOTATION SCHEME

It can be seen from figure 1 that the basic process of the proposed automatic image annotation consists of the following two steps: (1) Identifying candidate annotations of an image using a progressive model. (2) Refining each candidate annotation using the propagation procedure based on the spreading activation theory. Final achieved image annotations are these with the highest relevance value with respect to the visual content.

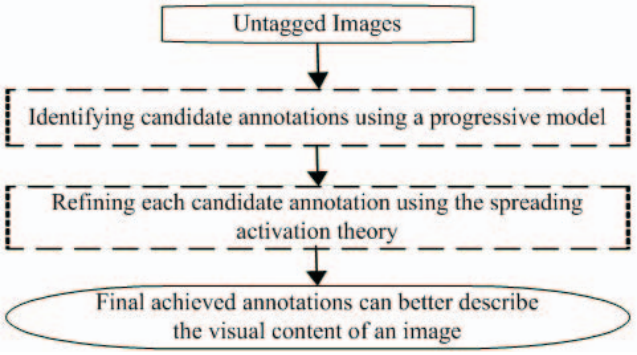

Figure 1: The basic idea of image annotation.

2.1. Identifying Candidate Annotations

Here, identifying candidate annotations is a process of assigning appropriate annotations to unlabeled images by taking into account of both visual and textual information. Supposing there is a set of training images labeled with some descriptive annotations, the annotations are then propagated to the rest of the image database by a relevance model. Supposing that T is a set of training images with one instance S, and Q is a test image. In addition, annotations of S are denoted as:

$$A_s = (a_1, a_2, \ldots\ldots, a_n) \tag{1}$$

where n is the number of annotations.

Inspired by the idea of Wang [15], the process of computing the joint distribution probability of annotations A_Q for Q is describes as below.

Step 1 Picking a training image S from the training set T with a uniform probability:

$$P(S) = \frac{1}{|T|} \tag{2}$$

where $|T|$ is the size of the training set.

Step 2 Sampling the test image Q from S with the probability $P(Q|S)$:

$$P(Q|S) = \exp\left(-\frac{\|Q_i - S_j\|}{\sigma}\right) = \exp\left(-\sum_{r=1}^{d}\frac{\|Q_i^r - S_j^r\|}{\sigma^r}\right) \tag{3}$$

where Q^r and S^r are the r^{th} dimension feature of the query image Q and a training image S respectively, σ^r is a positive parameter reflecting the scope of the r^{th} dimension, and d is dimensionality of the feature space.

Step 3 Sampling one annotation w_i of Q based on the Multiple Bernoulli Model:

$$P(w_i|S) = \frac{\mu * \delta_{w_i,S} + N_{w_i}}{\mu + |T|} \tag{4}$$

where μ is a smoothing parameter, $\delta_{wi,S}$ is 1 when w_i is contained in A_S and 0 otherwise, and $N(w_i)$ represents the frequency that w_i occurs in the training set T.

Step 4 Rewriting the formulation of the annotation w_1 with the largest likelihood:

$$\begin{aligned}\{w_1\} &= argmax\left[P(w_1|Q)\right] \\ &= argmax\left[P(w_1,Q)\right](when\ Q\ is\ given) \\ &= argmax\sum_{S\in|T|}\left[P(w_1|Q) * P(Q|S) * P(S)\right]\end{aligned} \tag{5}$$

Step 5 Inferring other candidate annotations using a progressive way:

$$\begin{aligned}&\{w_1, w_2, ..., w_n\} = argmax\left[P(w_1, w_2, ..., w_n|Q)\right] \\ &= argmax\left[P(w_1, w_2, ..., w_n, Q)\right](when\ Q\ is\ given) \\ &= argmax\sum_{S\in|T|}\left[P(w_1, w_2, ..., w_n|Q) * P(Q|S) * P(S)\right] \\ &= argmax\sum_{S\in|T|}\left[P(w_1,Q) * \cdots * P(w_n,Q) * P(Q|S) * P(S)\right] \\ &\approx argmax\sum_{S\in|T|}\left[P(w_1, w_2, ..., w_{n-1}|Q) * P(w_n,Q) * P(Q|S) * P(S)\right]\end{aligned} \tag{6}$$

Finally, a set of linguistic terms Γ obtained using the progressive method is adopted as the candidate annotations for further refinement process based on the spreading activation algorithm in the field of cognitive science.

2.2 Refining Candidate Annotations

In order to fully utilize the original information of candidate annotations and better reflect the correlation between textual annotations and image features, the image annotation refinement process is reformulated as a graph-based ranking problem dealt with by the spreading activation algorithm.

The main idea of the spreading activation algorithm can be defined formally as follows. For a node x and its neighbor node y, the process of activation propagation can be formulated as:

$$I_y(t+1) = O_x(t) * \Lambda_{xy} * (1-\delta) \tag{7}$$

where $I_y(t+1)$ represents the input value of node y at time $t+1$, $O_x(t)$ represents the output value of node x at time t, Λ_{xy} represents the link between nodes x and y, and δ is a decay

factor describing the energy loss in the spreading activation process. Here, a simplified spreading activation algorithm is utilized to compute the relevance values of nodes. For the simplified algorithm, the output value of node y at time t+1 is just the input value of node y at time t+1, that is $O_y(t+1)=I_y(t+1)$. Therefore, the entire propagation process based on the spreading activation algorithm can be submitted using the following formula:

$$O = \left[\Gamma - (1-\delta)\Lambda^T \right]^{-1} I \tag{8}$$

where Γ is an $n*n$ identity matrix, δ is the decay factor, Λ is the correlation matrix with element Λ_{ij} representing the link between annotations w_i and w_j, $I=[I_1, I_2,\ldots\ldots, I_n]^T$ is the initial values of annotations input into the semantic network, and $O=[O_1, O_2,\ldots\ldots, O_n]^T$ is the final output vector whose element O_i denotes the value of annotation w_i obtained using the spreading activation process.

In this paper, the computation formula of I_j in the vector $I=[I_1, I_2,\ldots\ldots, I_n]^T$ as the input of the spreading activation process is described as:

$$I_l = \frac{num(w_l)}{\sum_{all\ w_l} num(w_l)} \tag{9}$$

where $num(w_l)$ is the number of images annotated by the annotation w_l.

With the given input vector I, the spreading activation procedure is performed on the semantic network Λ to regulate and optimize the relevance of candidate annotations for an image. In the proposed framework, the matrix Λ is formed using the following procedure. The semantic relation between two annotations w_i and w_j, r_{ij} is first extracted from the training set T based on the co-occurrence relation. Then, the value of element Λ_{ij} in network matrix $\Lambda$is formulated as the frequency of semantic relation r_{ij} :

$$\Lambda_{ij}' = \frac{num(w_i, w_j)}{min\left[num(w_i), num(w_j)\right]} \tag{10}$$

where $num(w_i, w_j)$ represents the number of training images annotated by both the two annotations w_i and w_j. This formula reveals the following an interesting phenomenon: The more frequencies of two annotations co-existing in training images and the less frequencies of each single annotation existing in training images, the more correlated are the two annotations. Finally, the semantic network matrix Λ is normalized to ensure the sum of each column being one based on the following formula:

$$\Lambda_{ij} = \frac{\Lambda_{ij}'}{\sum_{all\ j} \Lambda_{ij}'} \tag{11}$$

With the spreading activation theory based optimization formula (8), the relevance value of each annotation can be calculated. A high value obtained denotes that this annotation is more relevant. Therefore, these annotations can be ranked with respect to their relevance values to the visual content of an image.

3. EXPERIMENTAL RESULTS

In this section, we will discuss the issue of experimental design and evaluate the performance of the proposed framework of image annotation by comparing with other two methods: content-based annotation refinement approach [9] (CBAR for short) and regularized latent dirichlet allocation model [10] (RLDA for short).

3.1 Experimental Design

All the experiments in this work are conducted on a dataset consisting of 50, 000 images, crawled from the image share website Flickr. The query keywords selected to implement the annotation-based retrieval on Flickr are the most popular annotations, such as: automobile, cat, bird, flower, mountain, tree, sea, sky, sunset, and water, *etc*. The retrieved images with their associated annotation information are ranked according to the option of interesting. Figure 2 illustrates some exemplaries with repesct to the query keyword "cat". Since many of the collected annotations are misspelling or meaningless, it is necessary to perform a pre-filtering for these annotations. More specifically, one annotation can be kept only when the annotation is matched with a term in the Wikipedia. In our case, 16,738 annotations are obtained for further refinement experiment.

Figure 2: Some exemplar images from Flickr and their associated annotations using the query keyword "cat".

For quantitative evaluation, 30,000 images randomly selected from the dataset are adopted as the training set and other 20,000 images are selected as the testing set. To get the ground truth, five volunteers are invited to view each image and exhaustively give their own annotations. Then, the ground truth annotations of each image is the intersection of the annotations.

To describe the reprehensive visual content of images, the extracted feature vector is a 428-dimensional vector, including 225-dimensional block-wise color moment, 128-dimensional wavelet texture and 75-dimensional edge distribution histogram. The radius parameter σ^r in equation 3 is set to the median value of the r^{th} dimension feature of all pair-wise L1 distances between images.

To evaluate the performance of the proposed scheme, the normalized discounted cumulative gain (*NDCG*) [16] of an image with N ranked annotations is first computed:

$$NDCG = \tau \sum_{j=1}^{N} \frac{2^{r(j)} - 1}{log\,(1 + j)} \quad (12)$$

where $r(j)$ is the relevance level of the j^{th} annotation, and τ is a normalization constant used to ensure the *NDCG* score of optimal ranking is 1. Here, the relevance is divided into three levels: irrelevant (score 1), partially relevant (score 2), relevant (score 3). Then, *NDCG* of all images are averaged and adopted as an overall performance evaluation measure of the annotation ranking method.

3.2 Performance Evaluation

The experimental results of three methods, including CBAR [9] and RLDA [10], are illustrated in figure 3. It can be seen from this figure that the proposed image annotation scheme using the spreading activation algorithm outperforms the other two approaches, which confirms the proposed approach from the two following aspects: (1) the spreading activation theory can better describe and reflect the correlation among image annotations from the semantic level; (2) the performance of information retrieval system can be improved using the research results in the field of cognitive science.

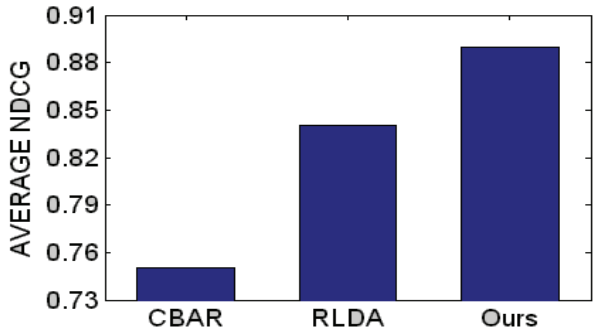

Figure 3: Performance of different annotation ranking methods in terms of average *NDCG*, where CBAR is content-based annotation refinement and RLDA is Regularized Latent Dirichlet Allocation annotation refinement.

4. CONCLUSIONS AND FURTHER WORKS

In this paper, a novel image annotation scheme based on the spreading activation algorithm is proposed, which is inspired by the research results in the field of cognitive science. The spreading activation theory reveals that all the information stored in long-term memory is represented in the form of a semantic network, where linguistic concepts are denoted as nodes and related with others through correlations. Here, the spreading activation algorithm is utilized to describe and measure the relationship among all annotations according to their relevance levels. The experimental results demonstrate the effectiveness of the proposed image annotation ranking approach.

It is worth noting that although the performance of the annotation refinement has been improved to some extent, there are several potential works for future development. Firstly, we will conduct experiments on lager datasets and different datasets using the proposed annotation method. Then, we will also investigate other research results in the field of cognitive science to improve the performance of image annotation. Finally, we will explore annotation quality improvement problem in a more general scenario, such as annotation categorization, to construct the better lexical indexing for social images.

5. ACKNOWLEDGEMENTS

This work is supported by RFDP of 20103223120003, NSFJ of BK2011758 and BK2010077, PAPD, CNSF of 60972045 and 61071089, SKLNST.

6. REFERENCES

[1] P. Duygulu, J. Barnard, and D. Forsyth. Object recognition as machine translation: learning a lexicon for a fixed image vocabulary. In Proc. ECCV, 2002: 97-112.
[2] V. Lavrenko, R. Manmatha and J. Jeon. A Model for Learning the Semantics of Pictures. In Proc. NIPS, 2003: 1-8.
[3] J. Tang, H. Li, and G. Qi. Integrated graph-based semi-supervised multiple/single instance learning framework for image annotation. In Proc. ACM Multimedia, 2008: 631-634.
[4] C. Wang, D. Blei, F. Li. Simultaneous image classification and annotation. In Proc. CVPR, 2009: 1903-1910.
[5] J. Tang, Q. Chen, and S. Yan. One person labels one million images. In Proc. ACM Multimedia, 2010: 1019-1022.
[6] X. Chen, Y. Mu, S. Yan, and T. Chua. Efficient large-scale image annotation by probabilistic collaborative multi-label propagation. In Proc. ACM Multimedia, 2010: 35-44.
[7] A. Ulges, M. Worring, and T. Breuel. Learning Visual Contexts for Image Annotation from Flickr Groups. IEEE Transactions on Multimedia, 2011, 13(2): 330-341.
[8] Y. Jin, L. Khan, L. Wang, and M. Awad. Image annotations by combining multiple evidence & wordNet. In Proc. ACM Multimedia, 2005: 706–715.
[9] C. Wang, F. Jing, L. Zhang, and H. Zhang. Content-based image annotation refinement. In Proc. CVPR, 2007: 1-8.
[10] H. Xu, J. Wang, X. Hua, and S. Li. Tag refinement by regularized LDA. In Proc. ACM Multimedia, 2009: 573-576.
[11] D. Liu, X. Hua, M. Wang, and H. Zhang. Image retagging. In Proc. ACM Multimedia, 2010: 491-500.
[12] G. Zhu, S. Yan, and Y. Ma. Image tag refinement towards low-rank, content-tag prior and error sparsity. In Proc. ACM Multimedia, 2010: 461-470.
[13] X. Jiang and A. Tan. Ontosearch: A full-text search engine for the semantic web. In Proc. IAAI, 2006: 1325-1330.
[14] R. Anderson. A spreading activation theory of memory. Verbal Learning and Verbal Behavior, 1983, 22: 261-295.
[15] B. Wang, Z. Li, N. Yu, and M. Li. Image Annotation in a Progressive Way. In *proc. ICME*, 2007: 811-814.
[16] K. Jarvelin and J. Kekalainen. Cumulated Gain-Based Evaluation of IR Techniques. ACM Transcations on Information System, 2002, 20(4): 422-446.