Automatic semantic annotation of images based on Web data

Guiguang Ding
School of Software
University of Tsinghua
Beijing, China
dinggg@tsinghua.edu.cn

Na Xu
School of Software
University of Tsinghua
Beijing, China
xu-na08@mails.tsinghua.edu.cn

Abstract—**Image annotation is a promising approach to bridging the semantic gap between low-level features and high-level concepts, and it can avoid the heavy manual labor. Most existing automatic image annotation approaches are based on supervised learning. They often encounter several problems, such as insufficiency of training data, lack of ability in dealing with new concept, and a limited number of semantic concepts. Web images are massive, rich information, customized etc. Therefore, Web data is a potential repository to provide a sufficient source for semantic annotation. In this paper, we proposed a novel image annotation method based on Web data, which aims to utilize Web data to perform automatic image annotation. Web data, collected from several image search engine, are first preprocessed, clustered and mined to construct a concept clustering model. And then candidate annotation terms are extracted through the model for query image. Afterwards, a rank algorithm is designed to filter out noise terms. Finally, an update phase is implemented to improve the whole method. Evaluations on benchmark image datasets have indicated the effectiveness of our proposal.**

Keywords-image annotation; Web data; data ming; clustering;

I. Introduction

With the development of network technology, data compression technology and digital photography technology, the number of images and videos has exploded. How to retrieve and manage them presents a significant challenge.

Recent studies reveal that semantic annotation for image or video is a promising approach to bridging the gap [1,2,3]. As presented by Hauptmann [1], automatic semantic annotation splits the semantic gap into two smaller gaps: (1) mapping the low-level features into the intermediate semantic concepts and (b) mapping these concepts into user needs. Annotation is exactly the first mapping. However, manual annotation for a large multimedia database is an expensive, time-consuming, error-prone and subjective process. So, automatic image or video annotation is the subject of much ongoing research, which has attracted a great deal of attention from both academic and industry in recent years.

Existing automatic image and video annotation (also referred to as “high-level feature extraction”, “semantic concept detection”) can be roughly classified into two categories: the model-based and data-driven methods. The model-based methods try to automatically assign concepts onto an image or a video shot by learning the relations between visual features and concepts. The model-based methods can be further divided two directions: generative models and discriminative models. Despite continuous efforts in researching new annotation methods, the annotation performance is usually unsatisfactory. A big problem they encountered is the lack of training data. They can only model a limited number of semantic concepts, which limits the application of them.

In order to solve the problem of large-scale images and videos annotation, many data-driven works have been done in recent year, which automatically annotate images or videos by mining the Internet data. In [4], Antonio et al. utilized a large dataset of 80 million tiny images collected from the Web to perform object recognition. However, Antonio automatically downloaded Web-scale image set usually containing keyword-unrelated images which are too diverse and noisy to be directly used for image annotation. Furthermore, the system is mainly used to object and scene recognition, so it is not satisfied for semantic annotation. Another important work was done by Microsoft Research Asia [5,6,7]. Motivated by Web search technologies in many commercial systems, they developed several search-based image annotation methods, using Web-scale image database and unlimited lexicon.

Web data is a potential repository to provide a sufficient source for semantic annotation. Data-driven methods based on Web data have attracted much research interest due to their effectiveness, practicability and an unlimited lexicon. However, this research area is still in its infancy and it is an under-explored field. We need to further investigate new schemes and frameworks to improve the effectiveness of the data-driven method. In this paper, we proposed a novel image annotation method, which aims to utilize Web data to perform automatic image annotation. The main contributions of this paper can be summarized as follows:

(1) Apply the graph-theoretic clustering to Web data for denoising, clustering and constructing semantic concept model.

(2) Propose a novel image annotation method. Different

978-1-4244-7409-7/10/$26.00 ©2010 IEEE

from the existing data-driven techniques [12], the proposed method offline processes and mines Web data collected from several image search engine to construct a concept clustering model. Moreover, it can be continuously upgraded; the effectiveness of image annotation can be increased gradually with the development of the update phase.

(3) We demonstrate that it can obviously improve the effectiveness of annotation algorithm through offline deleting the noise data and mining the important terms in Internet data.

The remainder of our paper is organized as follows. Section 2 lists some related work. In section 3, we describe the proposed method, and illustrate the key technologies. Some simulation results are presented in Section 4 and Conclusion is made in section 5.

II. Related Work

In this paper, we proposed to apply the graph-theoretic clustering to Web data for denoising, clustering and constructing semantic concept model. In this section, we briefly introduce the graph-theoretical clustering algorithm used in the paper, dominant set clustering, for the convenience of the reader.

Graph-based clustering has recently attracted more and more attention due to its clear intuitiveness, strong theoretical foundations, and successful applications in many fields such as video analysis and image segmentation [12,13,14]. Dominant-set clustering is a newly proposed algorithm that is based on a novel definition of clusters corresponding to dominant sets. It has low computational complexity, and it is flexible enough to allow online clustering [12].

In this paper, we utilize the dominant set clustering method [15,16,17] as the core of the method. We represent the data to be clustered as an undirected edge-weighted graph with no self-loops $G=(V,E,w)$, where V is the vertices set, E is the set of weighted edges that link different vertices. The edge weights w reflect the similarities between samples. Let w_{ij} be the edge weight between samples i and j ($w_{ij} \geq 0$). As customary, the graph G is represented as the corresponding weighted adjacency matrix, which is the n×n nonnegative, symmetric affinity matrix A = (a_{ij}). Where a_{ij} = w_{ij}, if (i, j) ∈ E and a_{ii} = 0, ∀i ∈ V. The definition of dominant sets is as follows.

Let S be a nonempty vertex set, where S ⊆ V . For any vertex i ∈ S, the average weighted degree of i relative to S is defined as

$$aw\deg_S(i) = \frac{1}{|S|}\sum_{j\in S} a_{ij} \qquad (1)$$

where $|S|$ is the number of vertices in S. For a vertex $j \notin S$, the similarity $\phi_S(i,j)$ between vertices i and j relative to S is defined as

$$\phi_S(i,j) = a_{ij} - aw\deg_S(i) \qquad (2)$$

Then, the weight $W_s(i)$ of i ∈ S relative to S is defined as:

$$W_s(i) = \begin{cases} 1, & if \quad |S| = 1 \\ \sum \phi_{S\setminus\{i\}}(j,i) W_{S\setminus\{i\}}(j), & otherwize \end{cases} \qquad (3)$$

The total weight of S is defined to be:

$$W(S) = \sum W_S(i) \qquad (4)$$

$W_s(i)$ is calculated simply as a function of the weights on the edges of the subgraph induced by S. Intuitively, equation (3) indicates that, to examine the weight of i relative to S, the influence of set S\{i} on i is examined. The more the influence is, the more the importance of i in S is.

Based on the definition of the dominant set, Pavan and Pelillo gave a method to find dominant set. In their method [20], a dominant set is found by first localizing a solution of program with an appropriate continuous optimization technique, and then picking up the support set of the solution found. Given an edge-weighted graph $G=(V,E,w)$ and its weighted adjacency matrix A, consider the following quadratic program:

$$\max f(x) = \frac{1}{2}X^T AX$$
$$s.t. \quad X \in \{X \in R^n; X \geq 0 \quad and \quad e^T X = 1\} \qquad (5)$$

Let X^* be a local solution of program (5), Let $\sigma(x^*)$ be its support: $\sigma(x^*) = \{i \in X : x_i^* \neq 0)$. It is proved that the vertex support set $\sigma(x^*)$ corresponds to a dominant set in the graph. Then, a dominant set is found by solving (5). Pavan and Pelillo provided a method to indirectly perform combinatorial optimization via continuous optimization. The following dynamical system is used to solve (5):

$$x_i(t+1) = x_i(t)\frac{(MX(t))_i}{X(t)^T MX(t)} \qquad (6)$$

where t represents the number of iterations. It turns out that their stationary points satisfying $x_i(t+1) = x_i(t)$ [14].

III. Automatic Image Annotation Method Based on Web Data

To improve the effectiveness of the image annotation, there are the following characteristics in our method. (1) Like [4], we first select an appropriate lexicon which contains 4 semantic concepts named as main-concepts, and then we use each term of the lexicon as search keyword to download Internet images and their textual descriptions. The lexicon and all textual descriptions can cover with the main semantic words as completely as possible. (2) The downloaded Web images are clustered to remove the worthless noise images by offline mode, and construct a concept clustering model. (3) The textual information in every cluster is analyzed and mined to generate some extended sub-concepts. These extended sub-concepts and all main-concepts are used as the semantic terms for to-be-annotated images. (4) Based on the concept clustering model, we sufficiently utilize both visual features and textual information of Web images to annotate images. (5) An update phase is developed to continuously upgrade the proposed method. We will give the details of every step in the following sub-sections.

A. Crawling Web data based on LSCOM lexicon

To cover visual forms as completely as possible, we use the LSCOM ontology as the search lexicon to download Internet images. The LSCOM (a Large-Scale Concept Ontology for Multimedia) project has been sponsored by the Disruptive Technology Office (DTO) [18], which includes 856 visual concepts jointly defined by researchers, information analysts, and ontology specialists according to the criteria of usefulness, feasibility, and observability. These concepts are related to events, objects, locations, people, and programs that can be found in general multimedia content. We selected 486 popular concepts as the query concepts.

We selected Google image search engine, Picsearch search engine, Flickr search engine, Yahoo image search engine, Bing image search engine to search related Web images. We automatically download the first 250 images and the surrounding textual descriptions returned by each online search engine for each concept (main-concept) defined by LSCOM as search keyword. This method gathered about 400,000 images and corresponding textual descriptions in total.

B. Constructing Concept Clustering Model

Although the search engines are independent, the duplicate images are inescapable. We removed all duplicate images by the content-based copy detection (CBCD) technology. CBCD extracts the unique feature information (content fingerprint) from the image, and then the duplicate images can be detected through its exclusive feature. In this paper, we utilized the ordinal measure to detect duplicate images. The image is partitioned into M×N equal-sized blocks. After sorting the average intensity values of blocks, each block can get an ordinal number. The vector in M×N dimensions with ordinal numbers of blocks is used as the fingerprint of the image. The rate of duplicate images is about 8% in our dataset.

The images gathered by the engines are loosely labeled in that visual content is often unrelated to the query word. In the first 200 images, the accuracy of images returned by search engines is about 50% [9], hence, a postprocessing step is often required. Various methods exit for cleaning up the noise data by removing images visually unrelated to the query word. In this paper, we utilized dominant-set clustering algorithm to clean up the noisy image and construct concept clustering model.

Figure 1 shows the process of constructing clustering model. The downloaded Web images are split into N subsets according to their corresponding search concepts (main-concepts). For each subset C_n, the corresponding affinity matrix W_n is computed, w_{ij} indicates the similarity between i-th and j-th images. In this paper, we use Grid Color Moments (225-dimensional vector) and Wavelet Texture (48-dimensional vector) as the low-level features of image.

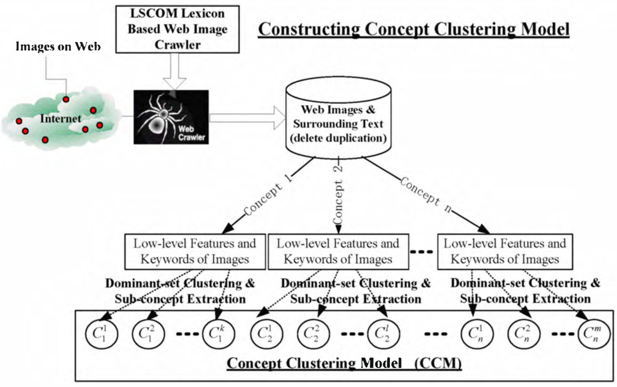
LSCOM Lexicon Based Web Image Crawler
Constructing Concept Clustering Model
Images on Web
Internet
Web Crawler
Web Images & Surrounding Text (delete duplication)
Concept 1
Concept 2
Concept n
Low-level Features and Keywords of Images
Low-level Features and Keywords of Images
Low-level Features and Keywords of Images
Dominant-set Clustering & Sub-concept Extraction
Dominant-set Clustering & Sub-concept Extraction
Dominant-set Clustering & Sub-concept Extraction
C_1^1
C_1^2
C_1^k
C_2^1
C_2^2
C_2^l
C_n^1
C_n^2
C_n^m
Concept Clustering Model (CCM)

Figure 1. The process of constructing clustering model.

And then dominant-set clustering is applied to W_n. Only high-quality clusters are needed for constructing concept clustering model, so dominant-set clustering is terminated when the number of vertex in a dominant set is smaller than a predefined threshold θ_D. The vertexes un-included in any dominant sets are seen as noisy images and cleaned up. An empirical value of θ_D is 80 to ensure the high quality clusters obtained.

For the subset C_n of main-concept n, K_n dominant-set clusters $\{C_n^1, C_n^2, \cdots, C_n^{K_n}\}$ are obtained. All main-concepts' dominant-set clusters are put together to form the final concept clustering model (CCM). After dominant-set clustering, the number of images is reduced to 312,152 in our dataset.

To improve the efficiency of the annotation system, we offline mined the textual descriptions of each dominant-set cluster to obtain its sub-concept set. To obtain these sub-concepts, some keywords are first extracted from the related textual information of each image in the dominant-set cluster. We called the kind of keyword as "image keyword". Some sub-concepts are then extracted from "image keywords" in each cluster. These main-concepts and sub-concepts are the final source of image annotation. The extraction methods of image keywords and sub-concept are described as follows.

(1) Image keywords extraction

Image keywords of each image are obtained from the textual description of it. After stop word removal and stemming, each word is ranked according to the similarity between the main-concept C_i of the cluster and the word W_j. The similarity is measured as follows:

$$Sim(C_i, W_j) = \alpha \times Sim_{closeness} + (1-\alpha) \times Sim_{WordNet} \tag{7}$$

where $Sim_{closeness}$ represents the position similarity which is computed like equation (8), $Sim_{WordNet}$ represents the semantic similarity. In the experiments, we set $\alpha = 0.5$.

$$Sim_{closeness}(C_i, W_j) = \frac{1}{n}\sum_{x=1}^{n} \min(\left|Pos_{C_i}(x) - Pos_{W_j}(1)\right|, \left|Pos_{C_i}(x) - Pos_{W_j}(2)\right|, \cdots) \tag{8}$$

where $Pos_{C_i}(x)$ represents the position of the x^th times occurrence of the main-concept C_i, $Pos_{W_j}(x)$ represents the position of the x^th times occurrence of the word W_j.

Finally, the first 15 words with highest ranks are reserved as image keywords of the image.

(2) Sub-concepts extraction

Sub-concepts should reflect the discriminably semantic concept of the cluster. To extract sub-concepts, we first emerge all image keywords in each cluster into a document. By this way, we will obtain $K_1 + K_2 + \cdots + K_n$ documents. And then, sub-concepts are extracted by mining the $K_1 + K_2 + \cdots + K_n$ documents using, the standard text process technique, tf-idf method. The words with the highest tf-idf scores are selected as the sub-concepts for one single cluster.

(3) Keyword correlation calculating

To increase the efficiency of online annotation, we offline calculated the correlation between sub-concept and its main-concept. There are mainly two categories of calculating word correlation: the lexicon-based and statistics-based methods. The lexicon-based method utilizes a lexicon such as WordNet to measure correlation between words. The statistics-based methods are data-driven and attempt to find word correlation based on term co-occurrence. In [19], a Normalized Google Distance (NGD) is proposed to measure the word correlation. NGD uses the Google search engine to find the words' co-occurrence in the Web pages. In this paper, we utilized NGD to measure the correlation between keyword and its class name. The NGD between sub-concept K_j and its main-concept C_i is calculated as follows:

$$NGD(C_i, K_j) = \frac{num(C_i, K_j)}{\min(num(C_i), num(K_j))} \tag{9}$$

where $num(C_j, K_i)$ represents the number of pages returned using both C_i and K_j submitted as a query by Google search engine, and $num(C_i)$ and $num(K_j)$ are respectively the number of pages returned using C_i and K_j as a query.

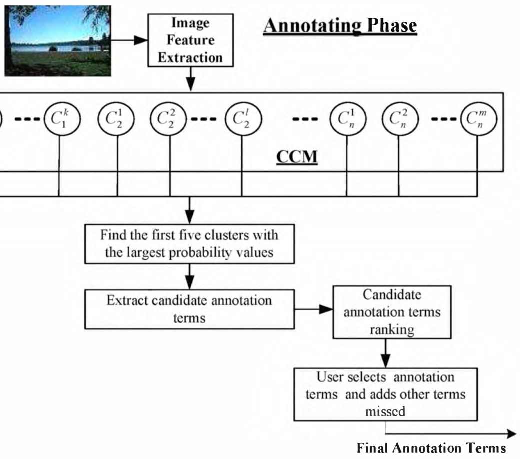
Image
Feature
Extraction
Annotating Phase
C_1^1
C_1^2
C_1^k
C_2^1
C_2^2
C_2^l
C_n^1
C_n^2
C_n^m
CCM
Find the first five clusters with the largest probability values
Extract candidate annotation terms
Candidate annotation terms ranking
User selects annotation terms and adds other terms missed
Final Annotation Terms

Figure 2. The blockgram **of** annotating phase

C. Annotating Phase

Figure 2 gives the blockgram of automatic annotation for one target image or keyframe Q. When the low-level feature of Q is input to CCM, it is decided the probability of the image falls into each dominant-set cluster in CCM. Referring to [15,16], the decision algorithm is described as follows.

Let $\vec{v}$ be the affinity vector describing the similarities between Q and the existing images in the cluster $C_n^{K_n}$. Then, the probability $p_n^{K_n}$ of Q falling into cluster $C_n^{K_n}$ is defined as:

$$p_n^{K_n} = \frac{\left|C_n^{K_n}\right| - 1}{\left|C_n^{K_n}\right| + 1}\left(\frac{\vec{v} \cdot u^{K_n}}{f(u^{K_n})} - 1\right) \tag{10}$$

where $\left|C_n^{K_n}\right|$ denotes the number of images in $C_n^{K_n}$, $f(*)$ is the object function of the equation (5), and u^{K_n} is the vector whose vertex support set is $C_n^{K_n}$. We find, from all the clusters in CMM, the first five clusters with the largest probability values. We select all main-concepts and sub-concepts related to the five clusters as candidate annotation terms for Q. Then the procedure of calculating the ranking value of each candidate annotation term is as follows.

- The ranking value of each main-concept is calculated as follows:

  $$Rank(C_i \mid Q) = Num(Q, C_i) \tag{11}$$

  where C_i is main-concept, $Num(Q, C_i)$ is the number of clusters (in above five clusters) included in this main-concept ($1 \le Num(Q, C_i) \le 5$).

- The ranking value of each sub-concept is calculated as follows:

$$Rank(SC_j \mid Q) = Rank(C_i \mid Q) \bullet NGD(C_i, SC_j) \quad (12)$$

where SC_j is sub-concept, $NGD(C_i, SC_j)$ is the Normalized Google Distance between main-concept C_i and its sub-concept SC_j, which has been offline calculated.

- Output the first N annotation terms with highest rank value.
- User selects the true annotation terms given by the method and adds other terms missed by a user interface.
- Give the final annotation terms for Q.

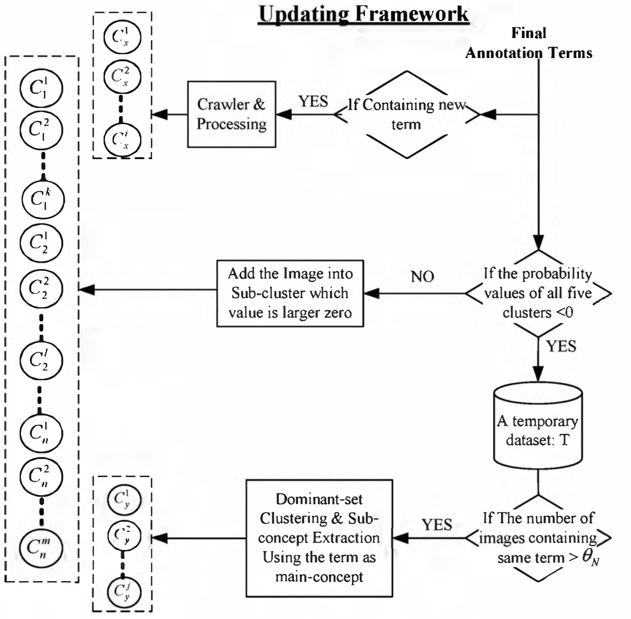
Updating Framework
Final Annotation Terms
Crawler & Processing
YES
If Containing new term
Add the Image into Sub-cluster which value is larger zero
NO
If the probability values of all five clusters <0
YES
A temporary dataset: T
Dominant-set Clustering & Sub-concept Extraction Using the term as main-concept
YES
If The number of images containing same term > θ_N
C_1^1
C_1^2
C_1^k
C_2^1
C_2^2
C_2^l
C_n^1
C_n^2
C_n^m
C_x^1
C_x^2
C_x^i
C_y^1
C_y^2
C_y^j

Figure 3. The Process of updating phase.

D. Updating Algorithm

After annotating an image, the method will automatically conduct an update step to upgrade the framework. Figure 3 shows the process of updating framework, which is summarized as follows.

- In final annotation terms, if there are new terms un-containing in CCM (generally added by user), then use each new term as main-concept to download Internet images and surrounding text and construct the new concept model $C_{n+1}^{K_{n+1}}$ like other concept model.
- In the five clusters, if there are at least one probability value (calculated by equation 10.) of them being larger than zero, then we add the images into the cluster(s) and compute the cluster(s) again.
- Else if the probability values of all five clusters are less than zero, the image is added a temporary dataset T. In the temporary dataset, if the number of images containing the same term is larger than θ_N (θ_N=800), then use the term as main-concept, apply dominant-set clustering to these images and their annotation terms, and construct a new concept model $C_{n+2}^{K_{n+2}}$.

Thus, with the development of the active learning process, the framework can be continuously upgraded.

IV. Experimental Results

To evaluate the effectiveness of the proposed method, we have implemented a prototype system and conducted a series of experiments. We use one benchmark dataset in experiments on image annotation: "Ground Truth Database (GTD)1" provided by the University of Washington. In GTD, there are 1,109 images and each has about five tags on average.

For the performance metric, we adopted N precision and coverage rate to measure the annotation performance of different methods. Top N precision measures the precision of top N ranked annotation terms for one image. Top N coverage rate is defined as the percentage of images that are correctly annotated by at least one word among the first N ranked annotation terms [14].

$$\begin{aligned} \Pr ecision_N &= \frac{1}{M}\sum_{i\in I} Correct\_i(N)/N \\ Coverage_N &= \frac{1}{M}\sum_{i\in I} IsContainCorrect\_i(N) \end{aligned} \quad (13)$$

Where $Correct\_i(N)$ is the number of correct annotation terms in top N ranked annotation terms of image i. I is the test image set, and M is the number of images in test image set. $IsContainCorrect\_i(N)$ judges whether image i contains correct annotation terms in the first N ranked ones. For every annotation result, we manually check these annotation terms.

For performance evaluation, we compare our method with the WordNet-based method (WordNet-based) and Search-based method (Search-based) similar to AnnoSearch [12]. In Search-based method, we first adopted Query-By-Example (QBE) method to retrieve the similar images in our Web data set. And then the Search Result Clustering (SRC) approach was used to mine the common concepts from the descriptions of the retrieval images to obtain the final annotation terms.

For GTD dataset, the precision and coverage rate of the "Top N" results are shown in Figure 4 and Figure 5. The three columns correspond to the proposed method, the Search-based method and the WordNet method, respectively. Our method is better than that of the other two methods. For top 3 precision, the proposed method is about 50% better than the WordNet-based method, and 17% better than the

[1] http://www.cs.washington.edu/research/imagedatabase/groundtruth/

Search-based method. For top 3 coverage, the proposed method is about 40% better than the WordNet-based method, and 18% better than the Search-based method. Each top N precision and coverage rate is also better than that of the other two methods.

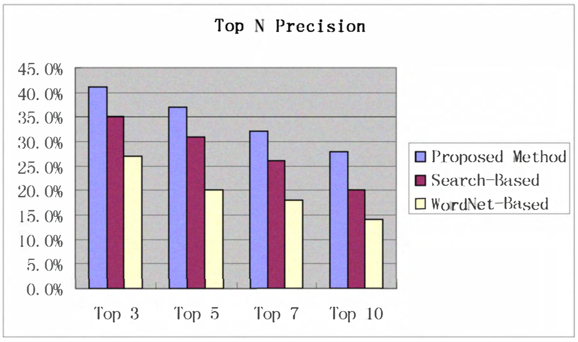

Figure 4. Top N Precision rate comparisons

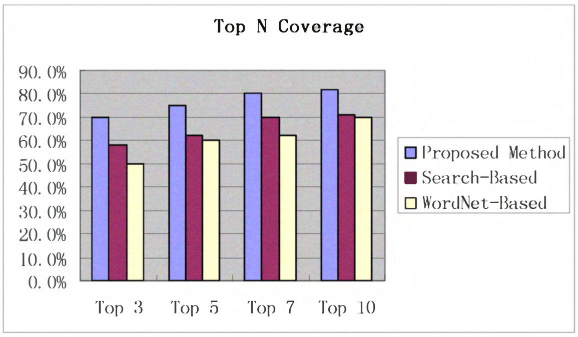

Figure 5. Top N Coverage rate comparisons

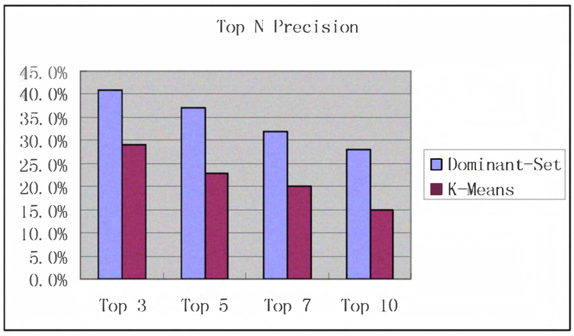

Figure 6. Comparison of effectiveness between the dominant-set clustering and K-Means clustering

To illustrate the effectiveness of the dominant-set clustering method, we also used the K-Means clustering method to construct a concept clustering model. Other parts had the same setup as the ID-MAF framework. Figure 6 gives the comparison of "Top N" precision between the two clustering method for the GTD dataset. As shown in Figure 6, the dominant-set clustering method is obviously more effective than the K-Means clustering method.

ACKNOWLEDGMENT

This work was supported by the National 863 Plans Projects (Grant No. 2009AA01Z410) and the National Natural Science Foundation of China (Grant No. 60972096).

REFERENCES

[1] A. G. Hauptmann. Lessons for the future from a decade of informedia video analysis research. in Proceedings of ACM International Conference on Image and Video Retrieval, 2005.

[2] Meng Wang, Xian-Sheng Hua, Richang Hong, Jinhui Tang, Guo-Jun Qi, and Yan Song. Unified Video Annotation via Multi-Graph Learning. IEEE Transactions on CSVT, 2009.

[3] X. Mu. Content-based video retrieval: Does video's semantic visual feature matter?. in Proceedings of International ACM SIGIR Conference, 2006.

[4] Antonio Torralba, Rob Fergus, William T. Freeman. 80 million tiny images: a large dataset for non-parametric object and scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008,1958-1970.

[5] C. H. Wang, F. Jing, et al. Image Annotation Refinement using Random Walk with Restarts. In proc. ACM MM, 2006, 647-650.

[6] X. Li, L.Chen, L.Zhang, F. Lin, and W. Ma. Image annotation by large-scale content-based image retrieval. In Proceedings of the 14th Annual ACM international Conference on Multimedia, Santa Barbara, CA, USA, October 23 - 27, 2006, pp. 607-610.

[7] X. J. Wang, L. Zhang, F. Jing, and W. Y. Ma. AnnoSearch: Image Auto-Annotation by Search. In Proc. CVPR 2006, 17-22.

[8] Weiming Hu, Wei Hu, Nianhua Xie, Steve Maybank. Unsupervised active learning based on Hierarchical Graph-theoretic clustering. IEEE Tranctions on Systems, Man, and Cybernetics-Part B. Vol.39, No.5, Oct. 2009, 1147-1161.

[9] H. Tong, J. R. He, M. J. Li, C. S. Zhang, and W. Y. Ma. Graph-based multi-modality learning. in Proceedings of ACM Multimedia, 2005.

[10] X. Yuan, X. S. Hua, M. Wang, and X. Wu. Manifold-ranking based video concept detection on large database and feature pool. in Proceedings of ACM Multimedia, 2006.

[11] S. X. Yu and J. Shi. Multiclass spectral clustering. in Proc. IEEE Int. Conf. Comput. Vis., 2003, vol. 1, 313–319.

[12] J. Shi and J. Malik. Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Mach. Intell., Aug. 2000, 888–905.

[13] Z. Wu and R. Leahy. An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation. IEEE Trans. Pattern Anal. Mach. Intell., Nov. 1993, 1101–1113.

[14] M. Pavan and M. Pelillo. A new graph-theoretic approach to clustering and segmentation. in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2003, 18–20.

[15] M. Pavan, M. Pelillo. Dominant sets and hierarchical clustering. The Ninth IEEE International Conference on Computer Vision Proceedings, 2003, 362-369.

[16] M. Pavan, M. Pelillo, D. di Informatica. clustering-Dominant sets and pairwise clustering. IEEE transactions on pattern analysis and machine intelligence, 2007.

[17] M. Naphade, J. R. Smith, J. Tesic, S.-F. Chang, W. Hsu, L. Kennedy, A. Hauptmann, J. Curtis. Large-Scale Concept Ontology for Multimedia. IEEE Multimedia Magazine, 13(3), 2006.

[18] LSCOM Lexicon Definitions and Annotations Version 1.0, DTO Challenge Workshop on Large Scale Concept Ontology for Multimedia. ADVENT Technical Report #217-2006-3 Columbia University, March 2006.

[19] D. M. Blei, M. I. Jordan. Modeling annotated data. In Proc. SIGIR, Toronto, July. 2003.