2013 13th International Conference on Control, Automation and Systems (ICCAS 2013)
Oct. 20-23, 2013 in Kimdaejung Convention Center, Gwangju, Korea

Image Caption Automatic Generation Method Based on Weighted Feature

Su Mei Xi[1,2] and *Young Im Cho[1]

[1] Department of Computer, the University of Suwon, Hwaseong, 445-743, Korea
(Tel : +81-31-229-8214; E-mail: ycho@suwon.ac.kr) *corresponding author
[2] School of Science, Qilu University of Technology, Jinan, 250-353, China
(Tel : +82-10-2433-9750; E-mail: xsm@suwon.ac.kr)

Abstract: For people to use numerous images effectively on the web, technologies must be able to explain image contents and must be capable of searching for data that users need. Moreover, images must be described with natural sentences based not only on the names of objects contained in an image but also on their mutual relations. We propose a novel system which generates sentential annotations for general images. Firstly, a weighted feature clustering algorithm is employed on the semantic concept clusters of the image regions. For a given cluster, we determine relevant features based on their statistical distribution and assign greater weights to relevant features as compared to less relevant features. In this way the computing of clustering algorithm can avoid dominated by trivial relevant or irrelevant features. Then, the relationship between clustering regions and semantic concepts is established according to the labeled images in the training set. Under the condition of the new unlabeled image regions, we calculate the conditional probability of each semantic keyword and annotate the new images with maximal conditional probability. Experiments on the Corel image set show the effectiveness of the new algorithm.

Keywords: Weighted feature; Region Clustering; Image Annotation

1. INTRODUCTION

With the rapid development of multimedia and network technology, sources of image data increase continuously. In order to retrieve the images of the user satisfied quickly and exactly from amounts of different type image data, Content based image retrieval technology is becoming the focus of people study more and more [1,2]. Image retrieval based on semantic is the most ideal way, while automatic image semantic annotation is the key to semantic image retrieval [3].

For the past few years, by using machine learning method and the statistical model, researchers designed a variety of different automatic image annotation models. Literate [4,5] are the methods based on support vector machine (SVM) discriminative model. There are two problems about these methods, and the one is that they are only suitable for the small number of semantic concepts; the other is that there must be some annotated image region as training data. They are typical image annotation methods based on generative model, such as cross-media relevance model (CMRM) [6], multiple Bernoulli relevance models [7], and continuous spatial relevance model [8] and so on. These models can overcome the two problems exist in SVM automatic image annotation methods, and they can implement the annotation of large semantic concepts, in the meantime, it also need not annotate the every image region of training set clearly. Nonetheless, the annotation effect of these methods is more sensitive to the clustering results of image region, therefore, the performance of overall model will be reduce if the clustering result is not ideal. An image annotation method based on weighted feature clustering is put forward in this paper. First, using an improved weighted feature clustering algorithm for image region clustering; then creating the association between image clustering area and semantic keywords, under the condition of unlabeled image region given, calculating the conditional probabilities of all semantic keywords appearance, the corresponding semantic keyword of the maximum conditional probability is assigned to the unlabeled images.

2. IMAGE REGION CLUSTERING BASED ON WEIGHTED FEATURE

2.1 Weight calculation of image feature

In a cluster, objects that are of the same cluster have high similarity, which largely depend on distinctive features, and these features are often influenced by irrelevant noise data, when the image data has a different number of dimensions and scale, the situation will get worse.

When clustering the image region, low-level features of the normal k-means clustering algorithm are assigned the same weight, regardless of the associated characteristics between the feature and its class, that is to say, not distinguish what is the distinguishing feature and what is weakly related or not related feature.

To solve this problem, an image clustering algorithm based on weighted feature is proposed. Intuitively considered, the importance is different about image features relative to the categories, i.e. some features with some category are strongly related, while some features with this category are weakly related, even not related. For example, color feature has the great influence for the class “sun”, while very lower influence of position feature for the class “sun”, almost without considering the impact of this feature for this class. Therefore, for different classes of images, we can assign different weights for its features, raising the feature weight with the strong correlation of class, reducing the weight of weak correlation feature, and discarding the irrelevant features.

In an image class, if the statistical distribution of some

feature is dense and discrete degree is small, the feature plays a dominant role relative to this class, is an "important" feature. Conversely, if the statistical distribution of some feature is scattered, this feature is less important for this class. Computing the "importance" of feature can be key to realize the feature weighted clustering. The standard deviation of data may well reflect the discrete degree of data sets, hence, the standard deviation can be used to describe the feature weight of image.

Due to the inconsistency of the ranges of each feature class, and a big difference, normalized processing for the features is needed.

The specific algorithm of feature weighting is described below.

Assuming total n_j samples of j-th class in the sample set, each sample is represented by L features, of which i-th sample $x_i=\{x_{i1},x_{i2},\ldots,x_{iL}\}$, accordingly, the feature weight of this class is $w_j=\{w_{j1},w_{j2},\ldots,w_{jL}\}$.

Because of each feature weight measured by the standard deviation of the feature distribution, firstly we need to calculate the standard deviation of each feature for each image.

Setting the standard deviation of l-th feature for class j is

$$\sigma_l = \sqrt{\sum_{i=1}^{n_j}(x_{il}-\bar{x_l})^2/(n_j-1)} \quad (1)$$

Where x_{il} is the l-th feature of the i-th sample of the j-th class, $\bar{x_l}$ is the average value of the l-th feature of all samples in this class.

The more intensive distribution of feature values and smaller standard deviation, the higher importance of the feature, its corresponding weight should be greater, that is to say, feature weight is inversely proportional to the standard deviation. Therefore, we introduce a concept of feature importance, instead of directly using standard deviation to describe the image feature weight. Greater the feature importance value is, the more important the feature is.

We define the feature importance is

$$e_l = \frac{1}{1+\sigma_l}, e_l \epsilon [0,1]. \quad (2)$$

It can be seen from eq. (2) that standard deviation σ_l is inversely proportional to feature importance e_l, the smaller the standard deviation σ_l, the greater feature importance e_l, which show that the feature is more important. When the feature values of the image distribute at a point concentrated, it shows that this feature plays a dominant role in the class, this moment σ_l is 0, e_l is 1 to its maximum value. When σ_l is the maximum value, then the e_l is 0, which indicates that the contribution of this feature is minimal in the class.

For the specific image feature weight calculation, getting the feature importance from the feature weight, the weight of the l-th feature of the j-th class is

$$w_{jl} = \frac{e_l}{\sum_{l=1}^{L} e_l}. \quad (3)$$

After the weight of each feature calculated, the weight is applied to an image similarity measure, that is, weighted distance is used to calculate the similarity between the images.

2.2 k-means clustering algorithm based on weighted feature

In this section we cluster the image regions in the training set by using weighted feature clustering algorithm. T is a training image collection, which divides the image regions of the training set into M classes. The algorithm is as follows.

Step 1 randomly select M image regions as the initial cluster centers;

Step 2 calculate the distance between the cluster center and each region in the training samples, and divide the corresponding objects according to the minimum distance;

Step 3 update the cluster centers;

Step 4 according to the description in section 1.1, calculate the weight of each feature in each class by using eq. (1) to eq. (3);

Step 5 calculate the distance between all regions and the cluster centers in training samples by using a weighted distance formula, and divide the corresponding samples according to the minimum distance.

$$\text{If } d=\min d(x_i,c_j)\sqrt{\sum_{l=1}^{L} w_{jl}(x_{il}-c'_{jl})^2}$$
$$1 \ll j \ll M \quad (4)$$
$$1 \ll i \ll N$$

Then divide the i-th image region x_i to the j-th class. Where, w_{jl} is the weight of the l-th feature of the j-th class image; x_{il} is the value of the l-th feature of the i-th region sample in the training set; c'_{jl} is the value of the l-th feature of the j-th class image cluster center after the updating of the cluster center;

Step 6 repeat step 3~ step 5 until the cluster center doesn't change.

It can be seen that, from the description of above algorithm, weighted feature clustering algorithm assigns equal weight to each feature and are 1 when the first sample division, here, the distance formula used with general clustering algorithm used is the same.

3. AUTOMATIC IMAGE ANNOTATION

Image regional clustering of the training set is implemented by weighted feature clustering algorithm. The calculation for the correlation between image region clustering and semantic concept and the realization of automatic image annotation are described below.

3.1 Calculation of keywords and region relational degree

T as training image collection, we cluster the training images by the weighted features and obtain M classes, $b=\{b_1,b_2,\ldots,b_M\}$, defining $W=\{w_1,w_2,\ldots,w_J\}$ as the keywords set of image annotation used. We use training set T to estimate the correlation degree between image regions and semantic keywords, that is, calculating the prior probability $P(w_k|b_i)$, which can be calculated by the following formula (5):

$$P(w_k|b_i) = (1-a)\frac{num(w_k,b_i)}{|B|} + a\frac{num(w_k,T)}{|T|} \quad (5)$$

Where, $num(w_k,b_i)$ is the co-occurrence number of semantic keyword w_k and image region b_i, that is, the

number of containing not only w_k but also b_i in one image; $num(w_k, T)$ is the number of keyword w_k appeared in the total training set T; |B| is the total number of all words and regions among the images of wk and bi co-occurrence; |T| is the number of all images in total training set; a is smoothing coefficient.

In the training set, by calculating the co-occurrence probability of semantic keyword and the image, we can obtain the conditional probability $P(w_k|b_i)$ of keyword w_k appearance under the condition of clustering region b_i given, that is to say, construct the correlation degree model of keyword and the clustering region.

3.2 Automatic annotation

In the image annotation stage, we also need segmentation for the unlabeled images. Assuming the region collection after the unlabeled image segmentation is $B' = \{b'_1, b'_2, \ldots, b'_N\}$, the given candidate keyword collection is the same as in section 2.1, $W=\{w_1, w_2, \ldots, w_J\}$, by calculating the conditional probability $P(w_k|b'_n)$ to determine the semantic description for the unlabeled images. For all $w_k \in W$, we treat the w_k when $P(w_k|b'_n)$ is maximum as the annotation of image region b'_n.

Defining sim(I',I) to denote the similarity of two image region, then given a region b'_n, the conditional probability, of which it is annotated w_k, is calculated as follows:

$$P(w_k|b'_n) = \sum_{k=1}^{N}(sim(b'_n, b_i)P(w_k|b_i)) \quad (6)$$

Where $P(w_k|b_i)$ is the correlation degree of keyword w_k and the annotated region b_i; $sim(b'_n, b_i)$ is the similarity of the unlabeled image region b'_n and the annotated region b_i, the smaller distance between the image regions is, the higher level of similarity of them it indicates.

The calculation formula of measuring the similarity of two image regions is

$$sim(b'_n, b_i) = \exp(-||b_i - b'_n||^2) = exp[-\sum_l (b_{il} - b'_{nl})^2] \quad (7)$$

Where, b_{il} is the l-th feature of region b_i; b'_{nl} is the l-th feature of region b'_n. After given the unlabeled image regions, we can obtain $P(w_k|b'_n)$ by formula (5)~(7), for the value of w_k when the conditional probability is maximum we treat as the annotation of this region.

4. EXPERIMENT

The following two experiments were used to verify the performance of the proposed method, the image dataset of these two experiments came from 500 images of Corel image database.

4.1 Weighted clustering algorithm

We extracted the global feature of the image when Verify the performance of weighted feature clustering algorithm, and each image was represented by one 9-dimension color features and 8-dimension textural features, totally 17-dimension lower feature vector. The figure 1 shows the Experimental comparison results of the general clustering algorithm and our proposed weighted clustering algorithm. For using random initial clustering center selection, the clustering result was unstable and every time the result changed. Therefore figure 1 shows the average result of ten times tests of two algorithms, among which WFC denotes the weighted feature clustering algorithm and C denotes the k-means clustering algorithm.

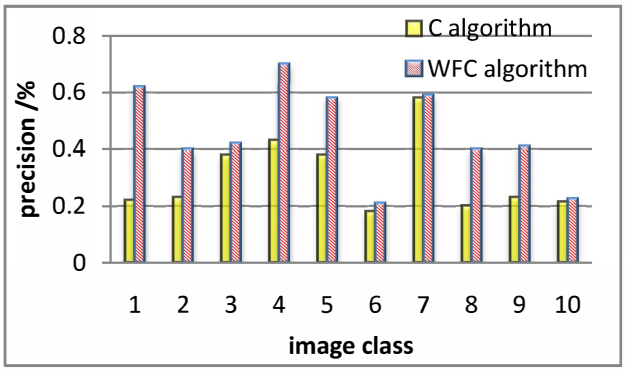

Fig.1 clustering results comparison of WFC algorithm and C algorithm

The application of the weighted features makes the features of highly relevant with the class to be strengthened, to the benefit of within-class image gathering, increasing the distinction between the class images. Figure 1 shows that the clustering algorithm based on weighted feature is superior to the traditional clustering algorithm for the average classification accuracy.

4.2 Automatic annotation of image

The k-means algorithm was used for image segmentation when to demonstrate the effectiveness of image annotation algorithm based on weighted feature. Low-level feature of the image region was a 9-dimension vector, which consisted of 3-dimension wavelet texture features, 3-dimension LUV color features and 3-dimension shape features [9]. The experiment used 200 images for training, and the remaining 300 images for testing. The proposed method compared with CMRM [6], and the experimental results shown in Figure 2.

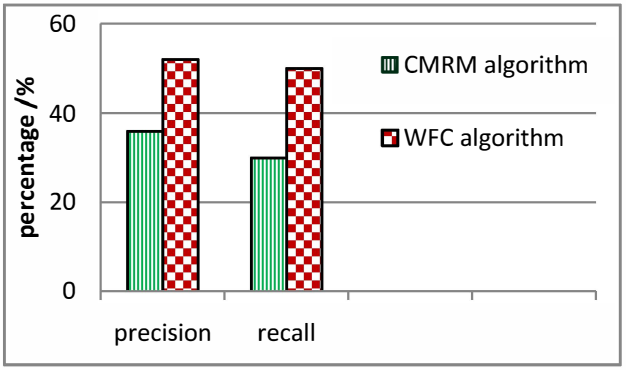

Fig.2 annotation results comparison of CMRM algorithm and WFC algorithm

CMRM algorithm using k-means clustering algorithm

to generate blob, due to the shortcomings of k-means clustering algorithm, it affected the accuracy of the results. While using our weighted feature clustering algorithm, the clustering performance can be improved and the annotation effect also can be improved effectively. It can be seen from figure 2, our proposed automatic annotation method can obtain better experimental results for the recall and precision.

5. CONCLUSION

Firstly we use the weighted feature clustering algorithm to cluster the image region, and then construct the relevance between the image region and semantic keywords on the training set, when the unlabeled image region given, computing the conditional probability of each semantic keyword appearance, the biggest conditional probability of semantic concepts as the image annotation. Since this method is feature weighted according to the statistical distribution of the features when clustering the image region, avoiding the clustering algorithm is dominated by weakly relevant features, it improves the clustering accuracy. Experimental results based on Corel image database show that the proposed method has better precision and recall.

ACKNOWLEDGMENT

This work is supported by three projects of Shandong Province Higher Educational Science and Technology Program (J12LN09), China, Ji'nan Youth Science and Technology Star Project (No.20120104), China, and by natural science fundation Project of Shandong Province, China (No. zr2011fm028).

REFERENCES

[1] Liu Ying, Zhang Dengsheng, Lu Goujun, et al. "A survey of content-based image retrieval with high-level semantics," *Pattern Recognition*, vol.40,no.1,pp.262-282,2007.

[2] Julia Vogel, Bernt Schiele, "Semantic modeling of natural scenes for content-based image retrieval," *International Journal of Computer Vision*, vol.72,no.2, pp.133-157, 2007.

[3] Wang Changhu, Zhang Lei, Zhang Hongjiang, "Learning to reduce the semantic gap in web image retrieval and annotation," *ACM SIGIR 2008. Singapore: ACM Press*, pp.355-362,2008.

[4] Gao Y L, Yin Y X, Uozumt Takashi, "A hierarchical image annotation method based on SVM and semi-supervised EM," *Acta Automatica Sinica*, vol.36,no.7, pp.960-967,2010.

[5] Gao Y L, Fan J P, Xue X Y, et al. "Automatic image annotation by incorporating feature hierarchy and boosting to scale up SVM classifiers," *Proceedings of the 14th ACM International Conference on Multimedia*, Santa Barbara: ACM Press,pp.901-910,2006.

[6] Jeon J, Lavrenko V, Manmatha R, "Automatic image annotation and retrieval using cross-media relevance models," *ACM SIGIR, Toronto: ACM Press*, pp.119-126,2003.

[7] Feng S L, Manmatha R, Lavrenko V, "Multiple Bernoulli relevance models for image and video annotation," *Proceedings of the IEEE Conf Computer Vision and Pattern Recognition*, Washington: IEEE Press, pp.1002-1009, 2004.

[8] Lavrenko V, Manmatha R, Jeon J, "A model for learning the semantics of pictures," *Proceedings of the Neural Information Processing Systems*, Vancouver, Whistler: MIT Press, pp.553-560, 2004.

[9] Chen Yixin, James Z Wang, "Image categorization by learning and reasoning with regions," *Journal of Machine Learning Research*, vol.5,no.8,pp.913-939,2004.