An Improvement Approach Based on the Label Correlation for Automatic Image Annotation

Cong Jin and Jin-An Liu
School of Computer, Central China Normal University, Wuhan, 430079, P. R. China
Email: jincong@mail.ccnu.edu.cn

***Abstract*—Since the semantic gap between low-level visual features and high level-image semantic, the performance of many existing annotation approaches is not satisfactory. In order to bridge the gap and improve the annotation performance, in this paper, an improvement approach based on the measure of label correlation is proposed. According to proposed approach, we can easily measure the correlation between labels. The experimental results confirm that the proposed approach of label set based on the measure of label correlation can improve the efficiency of automatic image annotation systems and achieve better annotation performance than the existing Automatic Image Annotation (AIA) approaches.**

***Index Terms*—automatic image annotation, image retrieval, an improvement algorithm, label correlation, annotation performance**

I. Introduction

As an important part of image retrieval, the annotation accuracy of image semantic directly influences the performance of image retrieval system. Image annotation associating labels or captions to the image, is a key step leading to the semantic label based image retrieval. The early annotation approaches rely on professionals or experts. It suffers from the problems of labor intensity and subjectivity. With the rapid growth of the digital images collections, various Automatic Image Annotation (AIA) approaches based on machine learning and statistical models have been proposed [1]-[5]. However, AIA is a difficult task for two main reasons [6] and [7]:

- It is hard to extract semantically meaningful entities only using low level image features.
- Lack correspondence between the labels and images in the training samples.

In addition, although AIA system can be implemented on many image datasets, the label set obtained by AIA was not accurate enough, and there are some errors, redundant and some illogical labels. For example, words such as "tropical", "snow", and "solar system" cannot simultaneously appear in the label set of an image, it also not appear in real life, however which may appear when using AIA. This situation is caused by three reasons:

- The image is such complicated or vague that cannot be extracted and identified by AIA system.
- Although the extracted visual features of the images and objects of the training set are similar, they does not belong the same class.
- Objects types have not been studied and estimated on the training set.

Therefore, after image annotated, the improvement of label set is necessary, which can greatly improve performance of AIA system.

In this paper, we propose an improvement approach of image label set based on the measure of label correlation, named OMLC. In other words, after a test image is annotated, we need to improve its label set in order to remove some errors, redundant and some illogical labels, or add some missed labels for more accurately describe the image content.

II. Measure of Labels Correlation

There is a correlation between the labels. For example, the probability of "sky" and "grass" appearing in an image is larger than the probability of "ocean" and "grass" appearing in an image, and the probability of "water" and "mountain" appearing in an image is larger than the probability of "horse" and "mountain" appearing in an image. So, {sky, grass} and {water, mountain} is more likely to be an image annotation label set than {ocean, grass} and {horse, mountain} respectively. Information of a label can help to learn information of another related label, especially when the training sample of some labels is not sufficient, the correlation between labels may provide some additional information.

In order to use the existing labels information and establish correspondence relations between the labels and the image, in this section, the measure of label correlation will be discussed. Suppose $I=\{I_1, I_2,..., I_m\}$, $T=\{T_1, T_2,..., T_s\}$ and $L=\{l_1, l_2,\dots, l_n\}$ be the training image set, test image set and the set of all labels of training image set respectively, where l_i is the i-th label, I_j is the j-th image of training set, T_k is the k-th image of test set; m, s and n are the total number of training images, test images and labels of training image set respectively.

Labels correlation is the number of two labels annotated in the same image. In general, the more number of two labels annotating the same image, the stronger correlation is between them. The number of two labels annotating the same image is also called the co-

Manuscript received January 4, 2017; revised May 29, 2017.

doi: 10.18178/ijoee.5.1.96-100

occurrence times. Co-occurrence matrix *C* may be used to describe the correlation between labels, where the element C_{ij} of co-occurrence matrix *C* represents the number of *i*-th and *j*-th labels annotating the same image. In practical applications, after the co-occurrence situation between labels being counted for image training set, the co-occurrence matrix *C* between labels can easily be obtained according to following steps:

- Calculate the adjacency matrix *A* of *L* and *I* as follows: if the *i*-th label l_i is a annotation word of the *j*-th training image I_j, then A_{ij} = 1, otherwise A_{ij} = 0.
- Suppose the *i*-th row vector of the adjacency matrix *A* be A_i, the element C_{ij} of the co-occurrence matrix *C* can be calculated by the inner product $C_{ij} = A_i \cdot A_j$ of the vectors A_i and A_j.

For example, if $L=$ {boat, cloud, ocean, tree, road, snow} and $I = \{image_1, image_2, image_3, image_4, image_5\}$, a simple illustration of the adjacency matrix *A* is shown in Table I.

TABLE I. Illustration of the Adjacency Matrix A

	Image$_1$	*Image*$_2$	*Image*$_3$	*Image*$_4$	*Image*$_5$
boat	1	0	0	0	1
cloud	1	1	1	0	0
ocean	1	0	0	0	0
tree	0	1	1	1	1
road	0	1	0	1	0
snow	0	0	1	1	0

From Table I, the co-occurrence matrix *C* is shown in Table II after calculating the inner product $A_i \cdot A_j$ of A_i and A_j ,.

TABLE II. Illustration of the Co-Occurrence Matrix C

	l_1	l_2	l_3	l_4	l_5	l_6
l_1	2	1	1	1	0	0
l_2	1	3	1	2	1	1
l_3	1	1	1	0	0	0
l_4	1	2	0	4	2	2
l_5	0	1	0	2	1	1
l_6	0	1	0	2	1	2

From Table II, we easily find that, there are the higher correlation between l_2 and l_4, l_4 and l_5, and l_4 and l_6.

However, the above calculation method is not satisfactory such as (1) There are too many elements zero in the adjacency matrix *A*; (2) Many correlation values are the same, which cannot be compared with each other. To avoid this situation, in this paper, we improve the calculation method of labels correlation, and a smoothing method is used in calculation process of labels correlation.

The purpose of introducing smooth method is to assign a minimum value for those zero elements in the adjacency matrix *A*, these values are calculated as follows

$$A_{ij} = \begin{cases} \dfrac{\mu^2}{(m+\mu)^2}, & if\ A_{ij} = 0 \\ 1, & if\ A_{ij} = 1 \end{cases} \quad (1)$$

where $\mu \in (0, 1)$ is a controlling parameter. After matrix *A* smoothed, according to new the adjacency matrix *A*, we recalculate its co-occurrence matrix, and the obtained new co-occurrence matrix is still denoted by *C*.

Let $c_{l_i l_j} = \sum_{k=1}^{n} C_{ik} \cdot C_{kj}$, we have

$$r_{l_i l_j} = c_{l_i l_j} / (c_{l_i l_i} + c_{l_j l_j} - c_{l_i l_j}) \quad (2)$$

Let $R_{l_i} = (r_{l_i l_1}, r_{l_i l_2}, \dots, r_{l_i l_n})$, the correlation $Corr(l_i, l_j)$ between labels l_i and l_j is calculated as follows

$$Corr(l_i, l_j) = \frac{R_{l_i} \cdot R_{l_j}}{|R_{l_i}| \cdot |R_{l_j}|} \quad (3)$$

III. Improvement Approach Based on Label Set

After getting label set of test image set, the obtained label set is impossible to ideally describe the image content. Thus, the improvement based on label set is a necessary task.

Let T_{test} be a test image, its label set obtained by AIA system, denoted by $L_t = \{l'_1, l'_2, \dots, l'_t\} \subset L$, then improvement probability of the annotation set for T_{test} is denoted as follows

$$O^* = \arg\max\{P(L_1, T_{test}), P(L_2, T_{test}), \dots, P(L_n, T_{test})\} \quad (4)$$

Suppose improvement solution of eq.(4) be $P(L_k, T_{test})$, then L_k corresponding to $P(L_k, T_{test})$ is an improvement label set of test image T_{test}.

We notice that, the eq.(4) is not a simple linear programming equation, so the computational cost of solving eq.(4) is very high when the amount of the test images is very large, therefore an approximate solution for improving the performance of AIA system is necessary.

In order to obtain an improved label set, we let

$$l^* = \arg\max_{l \in L \setminus L_{t-1}} P(l \mid L_{t-1}, T_{test}) \quad (5)$$

then obtained word l^* will be added into the set L_{t-1}, the obtained set is denoted by L_t, its improvement algorithm is as follows.

Algorithm 1. Improvement algorithm based on label set

Step 1. Calculate the following eq.(6)

$$l^* = \arg\max\{P(l'_1, T_{test}), P(l'_2, T_{test}), \dots, P(l'_t, T_{test})\} \quad (6)$$

then we obtain the label set L_1 of T_{test} only with a label, where $P(l'_i, T_{test})$ is the probability of T_{test} with label l'_i.

Step 2. According to eq.(5), we may obtain the label set L_k of T_{test} with *k* labels, where $k \in \{2, 3, \dots, n\}$.

Step 3. Calculate eq.(7)

$$L_k^* = \arg\min_{1\le k\le n} \max\{P(L_1, T_{test}), P(L_2, T_{test}), \dots, P(L_n, T_{test})\} \tag{7}$$

then L_k^* is improved label set of test image T_{test}.

In the previous calculations, the several issues should be noticed.

Note 1. In eq.(5), suppose L_{t-1} and T_{test} be independent of each other, then $P(l \mid L_{t-1}, T_{test})$ is as follows

$$P(l \mid L_{t-1}, T_{test}) = \frac{P(l \mid L_{t-1})P(l, T_{test})}{P(l)} \tag{8}$$

By ignoring $P(l)$ to not influence the improvement solution, the eq.(5) may be rewritten as

$$l^* = \arg\max_{l\in L\setminus L_{t-1}} P(l \mid L_{t-1})P(l, T_{test}) \tag{9}$$

According to the maximum likelihood estimation, we have

$$P(l \mid L_t) = \frac{\#\{T_k \mid l, L_t \in T_k)}{\#\{T_k \mid L_t \in T_k)} \tag{10}$$

where $\#\{T_k \mid l, L_t \in T_k)$ represents the number of test image including label *l* and label set L_t simultaneously.

Note 2. For a test image T_{test} and $\forall l \in L$, we know that *l* is a label of T_{test} or not, thus $P(l, T_{test})$ in eq.(6), eq.(8) or (9) can be described by Bernoulli distribution as follows

$$P(l, I_{test}) = \sum_{k=1}^{s} P(T_k)P(l, T_{test} \mid T_k) \tag{11}$$

Note 3. If $P(l \mid L_t)$ is calculated by eq.(10), their many values are zero, so we also give them a small value to avoid appearing 0 according to the Jelinek-Mercer smoothing method [8] as follows

$$P(l \mid L_t) = \omega \cdot P(l \mid L_t) + (1-\omega)\cdot \theta(l, L_t) \tag{12}$$

where ω is a controlling parameter, and $\theta(l, L_t)$ is calculated as follows

$$\theta(l_i, L_t) = \begin{cases} \sum_{l_j\in L_t} Corr(l_i, l_j), & |L_t| > 1 \\ 1, & |L_t| = 1 \end{cases} \tag{13}$$

and $\theta(l, L_t)$ also satisfies the probability characteristic, i.e.,

$$\sum_{l_i\in L, l_j\notin L_t} Corr(l_i, l_j) = 1 \tag{14}$$

IV. Experimental and Results

ESP game

IAPR-TC12

Corel5K

Figure 1. Some image examples.

For evaluating the annotation performance of proposed OMLC approach, we used three standard benchmark datasets for AIA task, namely the ESP Game [9], the IAPR TC-12 [10] and the Corel5K [11]. Recent very works have used these three datasets.

The ESP game dataset is a subset of an image-label database obtained from an online image labeling game called the ESP collaborative image labeling task. In ESP game, many labels are assigned to the same image. Only common labels are accepted. The set contains a wide variety of images annotated by 268 keywords, and is split into 19659 train and 2185 test images.

IAPR-TC12 was originally used in Image CLEF, and it is a collection of 19805 images of natural scenes. Unlike other similar databases, images in IAPR TC-12 are accompanied by free-flowing text captions. This image dataset is typically used for cross-language retrieval.

17825 images were used for training, and the remaining 1980 for testing.

Corel5K set contains 5000 images collected from the larger Corel CD set, split into 4500 training and 500 test examples. Each image is annotated with an average of 3.5 labels, and the dictionary contains 260 labels that appear in both the train and test set.

Some examples of the three image datasets are listed in Fig. 1.

Table III also summarizes the statistics information for each image dataset.

TABLE III. IMAGE STATISTICS FOR THE DATASETS USED IN THE EXPERIMENTS

Dataset	# of images	# of labels	Labels per image	Images per label
ESP Game	21844	268	4.69/15	363/5059
IAPR TC-12	19805	291	5.72/23	386/5534
Corel5K	5000	260	3.5/5	58.6/1004

A. Evaluation Measure

In this section, we use the evaluation measures that have been proposed in the literatures for estimating the performance of AIA approach, and precision is referred as the ratio of the times of correct annotation in relation to all the times of annotation, while recall is referred as the ratio of the times of correct annotation in relation to all the positive samples. The detailed definitions are as follows:

$$Precision = \frac{Q}{P} \tag{15}$$

$$Recall = \frac{Q}{R} \tag{16}$$

where P is the number of images annotated by some labels using proposed OMLC, Q is the number of images annotated correctly, and R is the total number of images annotated by some labels used image set. As a tradeoff between the above indicators, their geometric mean is adopted widely, namely

$$F_1 = \frac{2(\text{Precision}\cdot\text{Recall})}{\text{Precision}+\text{Recall}} \tag{17}$$

The larger values of F_1, the better performance of AIA approach.

B. Results and Discussions

In this paper, AIA system is viewed as a classification problem, which can be solved by using the SVM classifiers, and Gaussian kernel $K(I_i, I_j) = \exp(-\|I_i - I_j\|^2 / \sigma^2)$ is used in SVM, called CAIA. In traditional annotating process, only 5 labels are considered. The selection of parameters in improvement OMLC approach is listed in Table IV.

Table V lists all results averaged over 50 runs, the performances of other state-of-the-art image annotation approaches [12]-[17] are also compared with proposed OMLC.

As shown in Table V, OMLC provides the highest F_1 on three datasets, their values are 0.384, 0.470 and 0.499, and the improvements are at least 0.013, 0.018 and 0.021 respectively.

Therefore, the improvement performance of OMLC is very effective on three datasets, which shows that the performance of proposed improvement approach OMLC is satisfactory.

TABLE IV. THE PARAMETERS IN EXPERIMENTS

Parameter	Value
μ	0.35
ω	0.75

TABLE V. ALL RESULTS F1 HAVE BEEN AVERAGED OVER 50 RUNS

Algorithm	ESP Game			IAPR TC12			Corel5K		
	AP	AR	F_1	AP	AR	F_1	AP	AR	F_1
2PKNN [a]	**0.53**	0.27	0.358	0.54	0.37	0.439	0.44	0.46	0.450
CCD [a]	0.36	0.24	0.350	0.44	0.29	0.284	0.36	0.41	0.383
MBRM [a]	0.18	0.19	0.235	0.24	0.23	0.185	0.24	0.25	0.245
JEC [a]	0.22	0.25	0.285	0.28	0.29	0.234	0.27	0.32	0.293
NSIDML[a]	0.49	0.30	0.371	**0.57**	0.37	0.452	0.44	**0.52**	0.478
TagProp [a]	0.49	0.20	0.329	0.48	0.25	0.284	0.31	0.37	0.337
OMLC	0.48	**0.32**	**0.384**	0.55	**0.41**	**0.470**	**0.56**	0.45	**0.499**

[a] Provided by [12]-[17] respectively, AP and AR are average *Precision* and *Recall* respectively.

TABLE VI. EXPERIMENT RESULTS ON THREE DATASETS

Annotation result	cloud, grass, green, hill, red	man, old, picture, red, wall	band, light, man, music, play
Improvement result	cloud, green, mountain, man, sky, stone	black, man, red, sofa, picture	band, light, man, music, red, wheel

ESP Game

Annotation result	edge, front, glacier, life, tourist	court, player, sky, stadium, tennis	clothes, jean, man, shop, square
Improvement t result	glacier, people, life, rock, sky, water	game, player, sky, stadium, man, tennis, viewers	clothes, jean, man, pavement, shop, letter

IAPR TC12

Annotation result	bear, snow, wood, deer, cat	sky, jet, plane, smoke, river	grass, rocks, sand, valley, canyon
Improvement result	tree, snow, wood, fox	sky, jet, plane, smoke	rocks, sand, lake, valley, canyon, blue

Corel5K

We also show the annotation and its improvement results of some examples for AIA approach using SVM on three datasets, and they are listed in Table VI.

V. Conclusions

In this paper, since there is the semantic gap between low-level visual features and high level-image semantic, an improvement approach based on the label correlation for AIA is proposed. In order to bridge the gap and improve the annotation performance, according to proposed improvement approach, we can easily measure the correlation between labels. We investigated the improvement approach of label set based on the measure of label correlation using the co-occurrence matrix. The main advantages of proposed improvement approach are as follows:

- The proposed measure method of label correlation can effectively reflect the correlation between the labels, whose computation is very simple.
- The proposed approach can improve image annotation performance of AIA system, which shows that the proposed approach is very effective.
- After improving label set, the size of label set is no longer equal, and which shows that the number of labels should be decided by the image content, rather than predetermined by the users.

Acknowledgment

This work was supported by Natural Social Science Foundation of China (Grant No. 13BTQ050).

References

[1] S. H. Amiri and M. Jamzad, “Automatic image annotation using semi-supervised generative modeling,” *Pattern Recognition*, vol. 48, pp. 174-188, 2015.

[2] Y. Yu, W. Pedrycz, and D. Q. Miao, “Neighborhood rough sets based multi-label classification for automatic image annotation,” *International Journal of Approximate Reasoning*, vol. 54, pp. 1373-1387, 2013.

[3] C. Jin and S. W. Jin, “Automatic image annotation using feature selection based on improving quantum particle swarm optimization,” *Signal Processing*, vol. 109, pp. 172-181, 2015.

[4] D. D. Burdescu, C. G. Mihai, L. Stanescu, and M. Brezovan, “Automatic image annotation and semantic based image retrieval for medical domain,” *Neurocomputing*, vol. 109, pp. 33-48, 2013.

[5] A. Siddiqui, N. Mishra, and J. S. Verma, “A survey on automatic image annotation and retrieval,” *International Journal of Computer Applications*, vol. 118, pp. 27-32, 2015.

[6] X. D. Zhou, M. Wang, Q. Zhang, J. Q. Zhang, and B. L. Shi, “Automatic image annotation by an iterative approach: Incorporating keyword correlations and region matching,” in *Proc. 6th ACM International Conference on Image and Video Retrieval*, Amsterdam, The Netherlands, 2007, pp. 22-25.

[7] C. Jin and S. W. Jin, “Image semantic distance metric learning approach for large-scale automatic image annotation,” in *Proc. International Conference on Internet of Things and Big Data*, 2016, pp. 277-283.

[8] C. X. Zhai and J. Lafferty, “A study of smoothing methods for language models applied to information retrieval,” *ACM Transactions on Information Systems*, vol. 22, pp. 179-214, 2004.

[9] L. V. Ahn and L. Dabbis, “Labeling images with a computer game,” in *Proc. SIGCHI Conference on Human Factors in Computing Systems*, 2004, pp. 319-326.

[10] M. Grubinger, “Analysis and evaluation of visual information systems performance,” PhD thesis, Victoria University, Melbourne, Australia, 2007.

[11] P. Duygulu, K. Barnard, J. F. G. de Freitas, and D. A. Forsyth, “Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary,” in *Proc. European Conference on Computer Vision*, 2002, pp. 97-112.

[12] Y. Verma and C. V. Jawahar, “Image annotation using metric learning in semantic neighbourhoods,” in *Proc. Computer Vision*, 2012, pp. 836-849.

[13] H. Nakayama, “Linear distance metric learning for large-scale generic image recognition,” PhD thesis, The University of Tokyo, Japan, 2011.

[14] S. L. Feng, R. Manmatha, and V. Lavrenko, “Multiple Bernoulli relevance models for image and video annotation,” in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004.

[15] A. Makadia, V. Pavlovic, and S. Kumar, “A new baseline for image annotation,” in *Proc. Computer Vision*, 2008, pp. 316-329.

[16] C. Jin and S. W. Jin, “Image distance metric learning based on neighborhood sets for automatic image annotation,” *Journal of Visual Communication and Image Representation*, vol. 34, pp. 167-175, 2016.

[17] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, “Tagprop: Discriminative metric learning in nearest neighbor models for image auto-annotation,” in *Proc. 12th International Conference on Computer Vision*, Kyoto, Japan, 2009, pp. 309-316.

Cong Jin is a full professor of the school of computer, Central China Normal University, China. She has published more than 150 papers on digital image processing, automatic image annotation, and algorithm design and analysis. Her main research interests include digital image processing, artificial intelligence, and software reliability prediction, etc.

Jin-An Liu is a senior experimental technician of the school of computer, Central China Normal University, China. He has published nearly 10 papers on digital image processing. His main research interests include automatic image annotation, and artificial intelligence, etc.