Automatic Image Annotation Based on Semi-supervised Probabilistic CCA

Bo Zhang[1,2](✉), Gang Ma[2,3], Xi Yang[2], Zhongzhi Shi[2], and Jie Hao[4]

[1] China University of Mining and Technology, Xuzhou 221116, China
zhangb@ics.ict.ac.cn

[2] Institute of Computing Technology, Chinese Academy of Sciences,
Beijing 100190, China
{mag,yangx,shizz}@ics.ict.ac.cn

[3] University of Chinese Academy of Sciences, Beijing 100190, China

[4] School of Medicine Information, Xuzhou Medical University, Xuzhou 221000, China
haojie@xzmc.edu.cn

Abstract. We propose a novel semi-supervised method for building a statistical model that represents the relationship between images and text labels (tags) based on a semi-supervised variant of CCA called Semi-PCCA, which extends the probabilistic CCA model to make use of the labelled and unlabelled images together to extract the low-dimensional latent space representing topics of images. Real-world image tagging experiments indicate that our proposed method improves the accuracy even when only a small number of labelled images are available.

Keywords: Probabilistic CCA · Semi-supervised method · Automatic image annotation

1 Introduction

Automatic image annotation has become an important and challenging problem due to the existence of semantic gap. The state-of-the-art techniques of image auto-annotation can be roughly categorized into two different schools of thought. The first one defines auto-annotation as a traditional supervised classification problem, which treats each word (or semantic concept) as an independent class and creates different classifiers for every word. This approach computes similarity at the visual level and annotates a new image by propagating the corresponding words. The second perspective takes a different stand and treats images and texts

This work is supported by the National Program on Key Basic Research Project (973 Program) (No. 2013CB329502), National Natural Science Foundation of China (No. 61035003), National High-tech R&D Program of China (863 Program) (No. 2012AA011003), National Science and Technology Support Program (No. 2012BA107B02), Natural Science Foundation of Jiangsu Province (No. BK20160276).

Z. Shi et al. (Eds.): IIP 2016, IFIP AICT 486, pp. 211–221, 2016.
DOI: 10.1007/978-3-319-48390-0_22

as equivalent data. It attempts to discover the correlation between visual features and textual words on an unsupervised basis, by estimating the joint distribution of features and words. Thus, it poses annotation as statistical inference in a graphical model. Under this perspective, images are treated as bags of words and features, each of which are assumed generated by a hidden variable. Various approaches differ in the definition of the states of the hidden variable: some associate them with images in the database, while others associate them with image clusters or latent aspects (topics).

As latent aspect models, PLSA [8] and latent Dirichlet allocation (LDA) [3] have been successfully applied to annotate and retrieve images. PLSA-WORDS [12] is a representative approach, which achieves the annotation task by constraining the latent space to ensure its consistency in words. However, since standard PLSA can only handle discrete quantity (such as textual words), this approach quantizes feature vectors into discrete visual words for PLSA modeling. Therefore, its annotation performance is sensitive to the clustering granularity. GM-PLSA [11] deals with the data of different modalities in terms of their characteristics, which assumes that feature vectors in an image are governed by a Gaussian distribution under a given latent aspect other than a multinomial one, and employs continuous PLSA and standard PLSA to model visual features and textual words respectively. This model learns the correlation between these two modalities by an asymmetric learning approach and then it can predict semantic annotation precisely for unseen images.

Canonical correlation analysis (CCA) is a data analysis and dimensionality reduction method similar to PCA. While PCA deals with only one data space, CCA is a technique for joint dimensionality reduction across two spaces that provide heterogeneous representations of the same data. CCA is a classical but still powerful method for analyzing these paired multi-view data. Since CCA can be interpreted as an approximation to Gaussian PLSA and also be regarded as an extension of Fisher linear discriminant analysis (FDA) to multi-label classification [1], learning topic models through CCA is not only computationally efficient, but also promising for multi-label image annotation and retrieval.

However, CCA requires the data be rigorously paired or one-to-one correspondence among different views due to its correlation definition. However, such requirement is usually not satisfied in real-world applications due to various reasons. To cope with this problem, several extensions of CCA have been proposed to utilize the meaningful prior information hidden in additional unpaired data. Blaschko et al. [2] proposes semi-supervised Laplacian regularization of kernel canonical correlation (SemiLRKCCA) to find a set of highly correlated directions by exploiting the intrinsic manifold geometry structure of all data (paired and unpaired). SemiCCA [10] resembles the manifold regularization, i.e., using the global structure of the whole training data including both paired and unpaired samples to regularize CCA. Consequently, SemiCCA seamlessly bridges CCA and principal component analysis (PCA), and inherits some characteristics of both PCA and CCA. Gu et al. [6] proposed partially paired locality correlation analysis (PPLCA), which effectively deals with the semi-paired scenario of wireless sensor network localization by virtue of the combination of the neighbourhood structure information in data. Most recently, Chen et al. [4] presents a general dimensionality reduction framework for semi-paired and semi-supervised multi-view data which naturally generalizes existing related works by using different kinds of prior information. Based on the framework, they develop a novel dimensionality reduction method, termed as semi-paired and semi-supervised generalized correlation analysis (S2GCA), which exploits a small amount of paired data to perform CCA.

We propose a semi-supervised variant of CCA named SemiPCCA based on the probabilistic model for CCA. The estimation of SemiPCCA model parameters is affected by the unpaied multi-view data (e.g. unlabelled image) which revealed the global structure within each modality. Then, an automatic image annotation method based on SemiPCCA is presented. Through estimating the relevance between images and words by using the labelled and unlabelled images together, this method is shown to be more accurate than previous publish methods.

This paper is organized as follows. After introducing the framework of the proposed SemiPCCA model briefly in Sect. 2, we formally present our automatic image annotation method based on SemiPCCA in Sect. 3. Finally Sect. 4 illustrates experiments results and Sect. 5 concludes the paper.

2 Framework

In this section, we first review a probabilistic model for CCA. Then armed with this probabilistic reformulation of CCA, we present our semi-supervised variant of CCA named SemiPCCA based on the probabilistic model for CCA. The estimation of SemiPCCA model parameters is affected by the unlabelled multi-view data which revealed the global structure within each modality.

2.1 Probabilistic Canonical Correlation Analysis

In [1], Bach and Jordan propose a probabilistic interpretation of CCA. In this model, two random vectors $x_1 \in \mathbb{R}^{m_1}$ and $x_2 \in \mathbb{R}^{m_2}$ are considered generated by the same latent variable $z \in \mathbb{R}^d (\min(m_1, m_2) \geqslant d \geqslant 1)$ and thus the "correlated" to each other.

In this model, the observations of x_1 and x_2 are generated form the same latent variable z (Gaussian distribution with zero mean and unit variance) with unknown linear transformations W_1 and W_2 by adding Gaussian noise ε_1 and ε_2, i.e.,

$$\begin{gathered} P(z) \sim \mathcal{N}(0, I_d), \\ P(\varepsilon_1) \sim \mathcal{N}(0, \Psi_1), P(\varepsilon_2) \sim \mathcal{N}(0, \Psi_2), \\ x_1 = W_1 z + \mu_1 + \varepsilon_1, W_1 \in \mathbb{R}^{m_1 \times d}, \\ x_2 = W_2 z + \mu_2 + \varepsilon_2, W_2 \in \mathbb{R}^{m_2 \times d}. \end{gathered} \tag{1}$$

From [1], the corresponding maximum-likelihood estimations to the unknown parameters μ_1, μ_2, W_1, W_2, Ψ_1 and Ψ_2 are

$$\begin{aligned} \hat{\mu}_1 &= \frac{1}{N}\sum_{i=1}^{N} x_1^2, \hat{\mu}_2 = \frac{1}{N}\sum_{i=1}^{N} x_2^2, \\ \hat{W}_1 &= \widetilde{\Sigma}_{11} U_{1d} M_1, \hat{W}_2 = \widetilde{\Sigma}_{22} U_{2d} M_2, \\ \hat{\Psi}_1 &= \widetilde{\Sigma}_{11} - \hat{W}_1 \hat{W}_1^T, \hat{\Psi}_2 = \widetilde{\Sigma}_{22} - \hat{W}_2 \hat{W}_2^T, \end{aligned} \tag{2}$$

where $\widetilde{\Sigma}_{11}$, $\widetilde{\Sigma}_{22}$ have the same meaning of standard CCA, the columns of U_{1d} and U_{2d} are the first d canonical directions, P_d is the diagonal matrix with its diagonal elements given by the first d canonical correlations and M_1, $M_2 \in \mathbb{R}^{d\times d}$, with spectral norms smaller the one, satisfying $M_1 M_2^T = P_d$. In our expectations, let $M_1 = M_2 = (P_d)^{1/2}$. The posterior expectations of z given x_1 and x_2 are

$$\begin{aligned} E\left(z|x_1\right) &= M_1^T U_{1d}^T \left(x_1 - \hat{\mu}_1\right), \\ E\left(z|x_2\right) &= M_2^T U_{2d}^T \left(x_2 - \hat{\mu}_2\right). \end{aligned} \tag{3}$$

Thus, $E\left(z|x_1\right)$ and $E\left(z|x_2\right)$ lie in the d dimensional subspace that are identical with those of standard CCA.

2.2 Semi-supervised PCCA

Consider a set of paired samples of size N_p, $X_1^P = \{(x_1^i)\}_{i=1}^{N^p}$ and $X_2^P = \{(x_2^i)\}_{i=1}^{N^p}$, where each sample x_1^i (resp. x_2^i) is represented as a vector with dimension of m_1 (resp. m_2). When the number of paired of samples is small, CCA tends to overfit the given paired samples. Here, let us consider the situation where unpaired samples $X_1^U = \{(x_1^j)\}_{j=N^p+1}^{N^1}$ and/or $X_2^U = \{(x_2^k)\}_{k=N^p+1}^{N^2}$ are additional provided, where X_1^U and X_2^U might be independently generated. Since the original CCA and PCCA cannot directly incorporate such unpaired samples, we proposed a novel method named Semi-supervised PCCA (SemiPCCA) that can avoid overfitting by utilizing the additional unpaired samples. See Fig. 1 for an illustration of the graphical model of the SemiPCCA model.

The whole observation is now $D = \{(x_1^i, x_2^i)\}_{i=1}^{N^p} \cup \{(x_1^j)\}_{j=N^p+1}^{N^1} \cup \{(x_2^k)\}_{k=N^p+1}^{N^2}$. The likelihood, with the independent assumption of all the data points, is calculated as

$$L(\Theta) = \prod_{i=1}^{N^p} P(x_1^i, x_2^i; \Theta) \prod_{j=N^p+1}^{N^1} P(x_1^j; \Theta) \prod_{k=N^p+1}^{N^2} P(x_2^k; \Theta) \tag{4}$$

In SemiPCCA model, for paired samples $\{(x_1^i, x_2^i)\}_{i=1}^{N^p}$, x_1^i and x_2^i are considered generated by the same latent variable z^i and $P(x_1^i, x_2^i)$ is calculated as in PCCA model, i.e.

$$P(x_1^i, x_2^i; \Theta) \sim \mathcal{N}\left(\begin{pmatrix}\mu_1 \\ \mu_2\end{pmatrix}, \begin{pmatrix} W_1 W_1^T + \Psi_1 & W_1 W_2^T \\ W_2 W_1^T & W_2 W_2^T + \Psi_2 \end{pmatrix}\right). \tag{5}$$

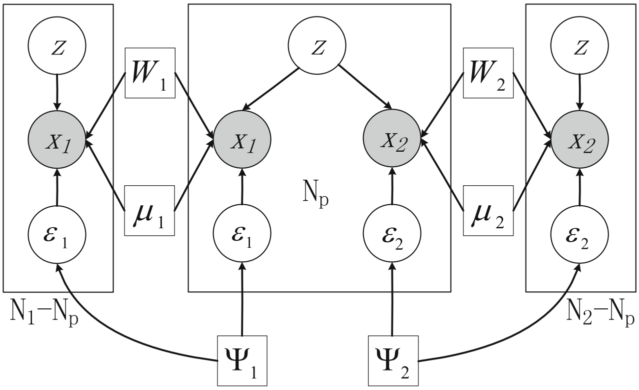

Fig. 1. Graphical model for Semi-supervised PCCA. The box denotes a plate comprising a data set of N_p paired observations, and additional unpaired samples.

Whereas for unpaired observations $X_1^U = \{(x_1^j)\}_{j=N^p+1}^{N^1}$ and/or $X_2^U = \{(x_2^k)\}_{k=N^p+1}^{N^2}$, x_1^j and x_2^k are separately generated from the latent variable ${z_1}^j$ and ${z_2}^k$ with linear transformations W_1 and W_2 by adding Gaussian noise ε_1 and ε_2. From Eq. (1),

$$\begin{aligned} P(x_1^j;\Theta) &\sim \mathcal{N}\left(\mu_1, W_1 W_1^T + \Psi_1\right), \\ P(x_2^k;\Theta) &\sim \mathcal{N}\left(\mu_2, W_2 W_2^T + \Psi_2\right). \end{aligned} \tag{6}$$

For means of x_1 and x_2 we have

$$\hat{\mu_1} = \frac{1}{N_1}\sum_{i=1}^{N_1} x_1^i, \hat{\mu_2} = \frac{1}{N_2}\sum_{i=1}^{N_2} x_1^i, \tag{7}$$

which are just the sample means. Since they are always the same in all EM iterations, we can centre the data $X_1^P \cup X_1^U$, $X_2^P \cup X_2^U$ by subtracting these means in the beginning and ignore these parameters in the learning process. So for simplicity we change the notation x_1^i, x_2^i, x_1^j and x_2^k to be the centred vectors in the following.

For the two mapping matrices, we have the updates

$$\hat{W_1} = (\sum_{i=1}^{N_p} x_1^i \langle z^i\rangle^T + \sum_{j=N_p+1}^{N_1} x_1^j \langle {z_1}^j\rangle^T)(\sum_{i=1}^{N_p}\langle z^i {z^i}^T\rangle + \sum_{j=N_p+1}^{N_1}\langle {z_1}^j {z_1}^{jT}\rangle)^{-1} \tag{8}$$

$$\hat{W_2} = (\sum_{i=1}^{N_p} x_2^i \langle z^i\rangle^T + \sum_{k=N_p+2}^{N_2} x_2^k \langle {z_2}^k\rangle^T)(\sum_{i=1}^{N_p}\langle z^i {z^i}^T\rangle + \sum_{k=N_p+1}^{N_2}\langle {z_2}^k {z_2}^{kT}\rangle)^{-1} \tag{9}$$

Finally the noise levels are updated as

$$\begin{aligned}\hat{\Psi}_1 = \frac{1}{N_1}\{(\sum_{i=1}^{N_p}(x_1^i - \hat{W}_1\langle z^i\rangle)(x_1^i - \hat{W}_1\langle z^i\rangle)^T \\ + \sum_{j=N_p+1}^{N_1}(x_1^j - \hat{W}_1\langle {z_1}^j\rangle)(x_1^j - \hat{W}_1\langle {z_1}^j\rangle)^T)\}\end{aligned} \tag{10}$$

$$\begin{aligned}\hat{\Psi}_2 = \frac{1}{N_2}\{(\sum_{i=1}^{N_p}(x_2^i - \hat{W}_2\langle z^i\rangle)(x_2^i - \hat{W}_2\langle z^i\rangle)^T \\ + \sum_{k=N_p+1}^{N_2}(x_2^k - \hat{W}_2\langle {z_2}^k\rangle)(x_2^k - \hat{W}_2\langle {z_2}^k\rangle)^T)\}\end{aligned} \tag{11}$$

2.3 Projections in SemiPCCA Model

Analogous to the PCCA model, the projection of a labelled image (x_1^i, x_2^i) in SemiPCCA model is directly given by Eq. (3).

Although this result looks similar as that in PCCA model, the learning of W_1 and W_2 are influenced by those unpaired samples. Unpaired samples reveal the global structure of whole the samples in each domain. Note once a basis in one sample space is rectified, the corresponding bases in the other sample space is also rectified so that correlations between two bases are maximized.

3 Annotation on Unlabelled Image

Now, we presents an automatic image annotation method based on the SemiPCCA, which estimating the association between images and words by using the labelled and unlabelled images together.

Let $X_1^P = \{(x_1^i)\}_{i=1}^{N^p}$ and $X_2^P = \{(x_2^i)\}_{i=1}^{N^p}$ be the set of labelled images and its corresponding semantic features with m_1 and m_2 dimensions of size N_p, and $X_1^U = \{(x_1^j)\}_{j=N^p+1}^{N^1}$ be a set of unlabelled images.

The first step is to extracts image features and labels features of training samples, and generates the essential latent space by fitting SemiPCCA.

In the context of automatic image annotation, X_1^U only exists, whereas X_2^U is empty. So, for the mapping matrices W_2 and the noise levels Ψ_2, we have to change the updates as follows,

$$\hat{W}_2 = (\sum_{i=1}^{N_p} x_2^i\langle z^i\rangle^T)(\sum_{i=1}^{N_p}\langle z^i {z^i}^T\rangle)^{-1} \tag{12}$$

$$\hat{\Psi}_2 = \frac{1}{N_p}\{(\sum_{i=1}^{N_p}(x_2^i - \hat{W}_2\langle z^i\rangle)(x_2^i - \hat{W}_2\langle z^i\rangle)^T)\} \tag{13}$$

Using this model, we derive the posterior probability of a sample in the latent space. When only an image feature x_1 is given, the posterior probability $P(z_1|x_1)$ of estimated latent variable z_1 becomes a normal distribution whose mean and variance are,

$$\begin{aligned} \mu_{z_1} &= \hat{W_1}^T (\hat{W_1}\hat{W_1}^T + \hat{\Psi_1})^{-1} (x_1 - \hat{\mu_1}), \\ \Psi_{z_1} &= I - \hat{W_1}^T (\hat{W_1}\hat{W_1}^T + \hat{\Psi_1})^{-1}, \end{aligned} \tag{14}$$

respectively. Also, when both an image feature x_1 and semantic feature x_2 are given, the posterior probability $P(z|x_1, x_2)$ becomes,

$$\begin{aligned} \mu_z &= \hat{W}^T (\hat{W}\hat{W}^T + \hat{\Psi})^{-1} \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \hat{\mu} \right), \\ \Psi_z &= I - \hat{W}^T (\hat{W}\hat{W}^T + \hat{\Psi})^{-1} \hat{W}. \end{aligned} \tag{15}$$

The second step is to map labelled training images $\{T_i^{(P)} = (x_1^i, x_2^i)\}_{i=1}^{N^p}$ and unlabelled images $\{(Q_j^{(U)} = x_1^j\}_{j=N^p+1}^{N^1}$ to the latent space with posterior probability $P(z|x_1, x_2)$ and $P(z|x_1)$ separately, and K-L distance is used for measuring the similarity between two images.

We define the similarity between two samples as follows. When two labelled images $T_i^{(P)} = (x_1^i, x_2^i)$ and $T_j^{(P)} = (x_1^j, x_2^j)$ are available, then similarity is defined as,

$$D\left(T_i^{(P)}, T_j^{(P)}\right) = \left(\mu_z^i - \mu_z^j\right)^T {\Psi_z}^{-1} \left(\mu_z^i - \mu_z^j\right), \tag{16}$$

which measuring essential similarities both in terms of appearance and semantics.

Furthermore, when labels feature of one of two samples is not available, e.g. one labelled image $T_i^{(P)} = (x_1^i, x_2^i)$ and one unlabelled image $Q_j^{(U)} = x_1^j$, which is the usual case in automatic image annotation, our framework also enables measuring similarities with semantic aspects even in the absence of labels features, and their similarity becomes:

$$D\left(T_i^{(P)}, Q_j^{(U)}\right) = \left(\mu_z^i - \mu_{z_1}^j\right)^T \left(\frac{{\Psi_z}^{-1} + {\Psi_{z_1}}^{-1}}{2}\right) \left(\mu_z^i - \mu_{z_1}^j\right) \tag{17}$$

As we described, we can formalize a new image annotation method. Let x_{new} demote a newly input image. To annotate x_{new} with some words, we calculate the posterior probability posterior probability of a word w given by x_{new}, which is represented as

$$P\left(w|x_{new}\right) = \sum_{i=1}^{N_p} P\left(w|T_i^{(P)}\right) P\left(T_i^{(P)}|x_{new}\right). \tag{18}$$

The posterior probability $P\left(T_i^{(P)}|x_{new}\right)$ of each labelled image $T_i^{(P)}$ using the above similarity measurement is defined as follow,

$$P\left(T_i^{(P)}|x_{new}\right) = \frac{exp\left(-D\left(T_i^{(P)}, x_{new}\right)\right)}{\sum_{i=1}^{N_p} exp\left(-D\left(T_i^{(P)}, x_{new}\right)\right)}, \tag{19}$$

where, the denominator is a regularization term so that $\sum_{i=1}^{N_p} P\left(T_i^{(P)}|x_{new}\right) = 1$. $P\left(w|T_i^{(P)}\right)$ corresponds to the sample-to-label model, which is defined as

$$P\left(w|T_i^{(P)}\right) = \mu\delta_{w,T_i^{(P)}} + (1-\mu)\frac{N_w}{NW}, \tag{20}$$

where N_w is the number of the images that contain w in the training data set, $\delta_{w,T_i^{(P)}} = 1$ if word w is annotated in the training sample $T_i^{(P)}$, otherwise $\delta_{w,T_i^{(P)}} = 0$. μ is a parameter between zero and one. NW is the number of the words.

The words are sorted in descending order of the posterior probability $P\left(w|x_{new}\right)$. The highest ranked words are used to annotate the image x_{new}.

4 Experiments

This section describes the results for the automatic image annotation task.

We use Corel5K and Corel30K to evaluate the performance of the proposed method. Corel5K contains 5,000 pairs of the image and the labels. Each image is manually annotated with one to five words. The training data has 371 words. 260 words among them appear in the test data.

Corel30K dataset is an extension of the Corel5K dataset based on a substantially larger database, which tries to correct some of the limitation in Corel5k such as small number of examples and small size of the vocabulary. Corel30K dataset contains 31,695 images and 5,587 words.

We follow the methodology of previous works, 500 images from the Corel5K are the test data. The other 1500, 2250 and 4500 images are selected from the Corel5K as the training data respectively, alone with the remaining training images in Corel5K and 31,695 images in Corel30K which acted as the unlabelled image to estimate the parameters of SemiPCCA together.

4.1 Feature Representation

As the image feature, we use the color higher-order local auto-correlation (Color-HLAC) features. This is a powerful global image feature for color images. Generally, global image features are suitable for realizing scalable systems because they can be extracted quite fast. Also, they are well suited for unconstrained image level annotation.

The Color-HLAC features enumerate all combinations of mask patterns that define autocorrelations of neighboring points and include both color information and texture information simultaneously. In this paper we use at most the 2nd order correlations, whose dimension is 714. The 2nd order Color-HLAC feature is reduced by PCA to preserve the 80 dimensions.

We extract Color-HLAC features from two scales (1/1, 1/2 size) to obtain robustness against scale change. Also, we extract them from edge images obtained by using the Sobel filter as well as the normal images. In all, the final image features are 320 dimensions.

As for labels feature, we use the word histogram. In this work, each image is simply annotated with a few words, so the word histogram becomes a binary feature.

4.2 Evaluation and Results

In this section, the performance of our model (SemiPCCA) is compared with several models. Image annotation performance is evaluated by comparing the captions automatically generated for the test set with the human-produced ground truth. For evaluation of annotation performance of our method, we follow the methodology of previous works. We define the automatic annotation as the five semantic words of largest posterior probability, and compute the recall and precision of every word in the test set. For a given semantic word, recall = B/C and precision = B/A, where A is the number of images automatically annotated with a given word; B is the number of images correctly annotated with that word; C is the number of images having that word in ground truth annotation. The average word precision and word recall values summarize the system performance.

Table 1. Performance comparison of different automatic image annotation models on Corel5k dataset.

Models	CRM	MBRM	PLSA WORDS	GM PLSA	Semi PCCA
#Words with *recall* > 0	107	122	105	125	151
Results on 49 best words,					
MR	0.70	0.78	0.71	0.79	0.94
MP	0.59	0.74	0.56	0.76	0.77
F1	0.64	0.76	0.63	0.77	0.85
Results on all 260 words,					
MR	0.19	0.25	0.20	0.25	0.32
MP	0.24	0.24	0.14	0.26	0.24
F1	0.21	0.24	0.16	0.25	0.27

Table 1 shows the results obtained by the proposed method and various previously proposed methods - - CRM [9], MBRM [5], PLSA-WORDS [12], GM-PLSA [11], using Corel5K. In order to compare with those previous models, we divide this dataset into 2 parts: a training set of 4,500 images and a test set of 500 images. We report the results on two sets of words: the subset of 49 best words and the complete set of all 260 words that occur in the training set. From the table, we can see that our model performs significantly better than all other models. We believe that using SemiPCCA to model visual and textual data by labelled and unlabelled images respectively is the reason for this result.

5 Conclusions

This paper presents an automatic image annotation method based on the SemiPCCA. Through estimating the association between images and words by using the labelled and unlabelled images together, this method is shown to be more accurate than previous publish methods. Experiments on the Corel dataset prove that our approach is promising for semantic image annotation. In comparison to several state-of-the-art annotation models, higher accuracy and superior effectiveness of our approach are reported.

References

1. Bach, F.R., Jordan, M.I.: A probability interpretation of canonical correlation analysis. Technical Report 688, Department of Statistics, Universityof California, Berkeley (2005)
2. Blaschko, M.B., Lampert, C.H., Gretton, A.: Semi-supervised Laplacian regularization of Kernel canonical correlation analysis. In: Daelemans, W., Goethals, B., Morik, K. (eds.) ECML PKDD 2008. LNCS (LNAI), vol. 5211, pp. 133–145. Springer, Heidelberg (2008). doi:10.1007/978-3-540-87479-9_27
3. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**(4–5), 993–1022 (2003)
4. Chen, X., Chen, S., Xue, H., Zhou, X.: A unified dimensionality reduction framework for semi-paired and semi-supervised multiview data. Pattern Recogn. **45**(5), 2005–2018 (2012)
5. Feng, S.L., Manmatha, R., Lavrenko, V.: Multiple bernoulli relevance models for image and video annotation. In: CVPR 2004, Washington, DC, United States, pp. 1002–1009 (2004)
6. Gu, J., Chen, S., Sun, T.: Localization with incompletely paired data in complex wireless sensor network. IEEE Trans. Wirel. Commun. **10**(9), 2841–2849 (2011)
7. Harada, T., Nakayama, H., Kuniyoshi, Y.: Image annotation and retrieval based on efficient learning of contextual latent space. In: ICME 2009, Piscataway, USA, pp. 858–861 (2009)
8. Hofmann, T.: Unsupervised learning by probabilistic latent semantic analysis. Mach. Learn. **42**, 177–196 (2001)
9. Jeon, J., Lavrenko, V., Manmatha, R.: Automatic image annotation and retrieval using crossmedia relevance models. In: SIGIR 2003, Toronto, Canada, pp. 119–126 (2003)

10. Kimura, A., Kameoka, H., Sugiyama, M., Nakano, T.: Semicca: efficient semi-supervised learning of canonical correlations. In: ICPR 2010, Istanbul, Turkey, pp. 2933–2936 (2010)
11. Li, Z., Shi, Z., Liu, X., Shi, Z.: Modeling continuous visual features for semantic image annotation and retrieval. Pattern Recogn. Lett. **32**(3), 516–523 (2011)
12. Monay, F., Gatica-Perez, D.: Modeling semantic aspects for cross-media image indexing. IEEE Trans. Pattern Anal. Mach. Intell. **29**(10), 1802–1817 (2007)