Simultaneous Image Classification and Annotation via Biased Random Walk on Tri-relational Graph

Xiao Cai[1], Hua Wang[2], Heng Huang[1], and Chris Ding[1]

[1] Department of Computer Science and Engineering, The University of Texas at Arlington, Arlington, Texas, 76019-0015, USA
[2] Department of Electrical Engineering and Computer Science, Colorado School of Mines, Golden, Colorado, 80401, USA
xiao.cai@mavs.uta.edu, huawang@mines.edu, {heng,chqding}@uta.edu

Abstract. Image annotation as well as classification are both critical and challenging work in computer vision research. Due to the rapid increasing number of images and inevitable biased annotation or classification by the human curator, it is desired to have an automatic way. Recently, there are lots of methods proposed regarding image classification or image annotation. However, people usually treat the above two tasks independently and tackle them separately. Actually, there is a relationship between the image class label and image annotation terms. As we know, an image with the sport class label rowing is more likely to be annotated with the terms water, boat and oar than the terms wall, net and floor, which are the descriptions of indoor sports.

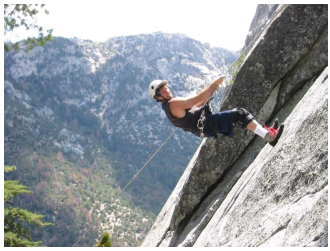
In this paper, we propose a new method for jointly class recognition and terms annotation. We present a novel *Tri-Relational Graph* (TG) model that comprises the data graph, annotation terms graph, class label graph, and connect them by two additional graphs induced from class label as well as annotation assignments. Upon the TG model, we introduce a *Biased Random Walk* (BRW) method to jointly recognize class and annotate terms by utilizing the interrelations between two tasks. We conduct the proposed method on two benchmark data sets and the experimental results demonstrate our joint learning method can achieve superior prediction results on both tasks than the state-of-the-art methods.

1 Introduction

Image classification and image annotation are both important and challenging problems in computer vision research. Automatic approaches are desired due to the rapid increasing number of images and inevitable biased annotation or classification by the human curator. In real life, we often encounter the images that are both labeled with one category as well as annotated with some free text, such as the images shown in Fig. 1, where the category label is the global description of the image from a bigger scope point of view and annotation terms are the local components derived from a smaller scope perspective.

In the previous research, people usually consider image classification and image annotation as two independent tasks. However, because of the existing relationship between the image class label and image annotation terms, these two problems can be

A. Fitzgibbon et al. (Eds.): ECCV 2012, Part VI, LNCS 7577, pp. 823–836, 2012.



Class: rockclimbing
Annotations: climber, hook, mountain, rock, rope, sky, tree

(a) UIUC-sports

Class:coast
Annotations:sky,building, sea, palm tree,sand beach,

(b) LabelMe

Fig. 1. Example images with one class label and several annotation terms from UIUC-sport [4] and LabelMe [5] datasets

tackled together [1] [2]. As we know, an image with the sport class label rowing is more likely to be annotated with the terms water, boat and oar than the terms wall, net and floor, which are the descriptions of indoor sports. In this paper, we propose a novel *Tri-Relational Graph* (TG) model that comprises the data graph, annotation term graph, class label graph, and connect them by two additional graphs induced from label as well as annotation assignments. Upon the TG model, we introduce a Biased Random Walk (BRW) method to simultaneously produce image-to-class, image-to-annotation, image-to-image, class-to-image, class-to-annotation, class-to-class, annotation-to-image, annotation-to-class, and annotation-to-annotation relevances to jointly learn the salient patterns among images that are predictive of their class label and annotation terms, and achieve both superior performances compared with the state-of-the-art methods. We consider each image as a data point and extract the Dense SIFT features [3] as the corresponding descriptors. We summarize our contributions as follows: **1**. This paper proposes a novel solution to questions "What is the image class?" and "What are the image annotations" simultaneously, given an unlabeled and unsegmented image; **2**. Via the new TG model that we constructed, the relationships between class label and annotation terms as well as the correlations among annotation terms can be naturally and explicitly propagated by the graph-based learning method; **3**. Other than using image-to-image relevance only conducted by the existing graph based methods, we propose a new BRW method to exploit the hidden annotation-to-annotation and annotation-to-class relevances.

2 Related Work

There are lots of work proposed recently regarding image classification or image annotation [6]. For image classification, Fei-fei Li [3] firstly used bag-of-word feature with the help of modified LDA model to classify the nature scenes. A.Bosch [7] utilized another generative model pLSA [8] with addition of KNN to do the scene classification. These topic models find a low dimensional representation of data under the assumption that each data point can exhibit multiple "topics". Discriminate model can solve

image classification problem as well, for example, multi-class support vector machine (MSVM). For image annotation, Ji *et al*. proposed Multi-Label Least Square (MLLS) method [9] to extract a common structure (subspace) shared among multiple labels. Yu *et al* extended unsupervised latent semantic indexing(LSI) [10] to make use of supervision information, called Multi-label informed Latent Semantic Index (MLSI) [10] and *etc* [11] [12] [13] [14]. Nevertheless, the above methods can only handle one task only. None of them consider the two problems together.

Recently, Chong Wang [15] takes advantage of the novel generative model JLDA, combining classification model SLDA [16] and annotation model [17] to do the image classification and image annotation simultaneously. Taking advantage of the relationship between class label and annotation terms, annotation can boost the classification performance and vice versa.

3 Method

In this section, we first construct a *Tri-Relational Graph* (TG) to model the intra-relationships as well as inter-relationships among images, class label and annotation assignments, followed by proposing a novel *Biased Random Walk* (BRW) method. Using BRW on TG, we can jointly make class classification and term annotation for the test image. We summarize the notation as follows. Matrices are written as uppercase letters and column vectors are written as boldface lowercase letters. v_i is the i-th element of column vector $\mathbf{v}$ and $M(i,j)$ is the entry located at i-th row and j-th column of matrix M. $\mathbf{e}$ is the column vector with all the elements being 1.

We have n images $X = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$, where each image is abstracted as a data point $\mathbf{x}_i \in \mathbb{R}^p$. Each data point $\mathbf{x}_i$ belongs to one and only one of K_c category classes $\mathcal{C} = \{c_1, \cdots, c_{K_c}\}$ represented by $\mathbf{y}_i^c \in \{0,1\}^{K_c}$, such that $\mathbf{y}_i^c(k) = 1$ if $\mathbf{x}_i$ is classified into class c_k, and 0 otherwise. Meanwhile, each image $\mathbf{x}_i$ is also annotated with a number of annotation terms $\mathcal{A} = \{a_1, \cdots, a_{K_a}\}$ represented by $\mathbf{y}_i^a \in \{0,1\}^{K_a}$, such that $\mathbf{y}_i^a(k) = 1$ if $\mathbf{x}_i$ is annotated with term a_k, and 0 otherwise. For convenience, we write $\mathbf{y}_i = [\mathbf{y}_i^{cT}, \mathbf{y}_i^{aT}]^T \in \{0,1\}^{K_c+K_a}$. Without loss of generality, we assume the first $l < n$ images are already labeled, which are denoted as $T = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^l$. Our task is to learn a function $f : X \to \{0,1\}^{K_c+K_a}$ from T that is able to classify the given test image $\mathbf{x}_i (l+1 \le i \le n)$ into one category class in $\mathcal{C}$ and to annotate it with a number of annotation terms in $\mathcal{A}$ at the same time. For simplicity, we write $Y_c = [\mathbf{y}_1^c, \cdots, \mathbf{y}_n^c]$, $Y_a = [\mathbf{y}_1^a, \cdots, \mathbf{y}_n^a]$ and $Y = [\mathbf{y}_1, \cdots, \mathbf{y}_n]$.

3.1 The Construction of Tri-relational Graph

Given image data set X, the pairwise similarity $W_X \in \mathbb{R}^{n \times n}$ between data points can be computed using the Gaussian kernel function,

$$W_X(i,j) = \begin{cases} \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2, & i \neq j \\ 0, & otherwise \end{cases} \tag{1}$$

where each image $\mathbf{x}$ is represented by the dense SIFT feature [3]. Regarding the parameter σ, we utlize self-tuning method [18]. In addition, we use kNN graph. To be specific,

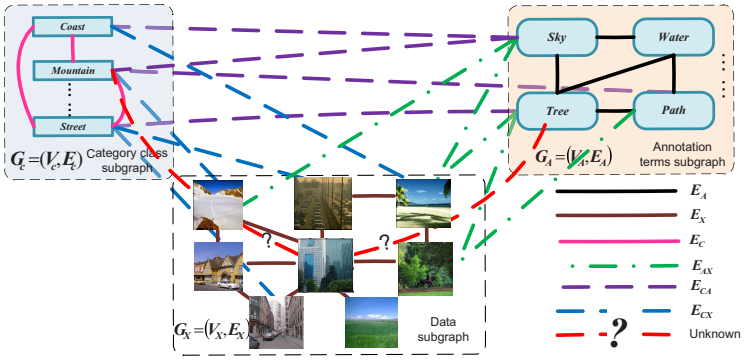


Fig. 2. The Tri-Relational Graph constructed from the LabelMe data. Solid lines indicate affinity between vertices in the same subgraph, dashed lines indicate associations between vertices in two different subgraphs.

we connect $\mathbf{x}_i, \mathbf{x}_j$ if one of them is among the other's k nearest neighbor and determine the value of the edge connecting them by Eq.(1). $W_X(i,j)$ indicates how closely $\mathbf{x}_i$ and $\mathbf{x}_j$ are related. From W_X, a graph $\mathcal{G}_X = (\mathcal{V}_X, \mathcal{E}_X)$ can be induced, where $\mathcal{V}_X = X$ and $\mathcal{E}_X \subseteq \mathcal{V}_X \times \mathcal{V}_X$. Because $\mathcal{G}_X$ characterizes the intra-relationships between data points, it is usually called data graph, such as the middle subgraph in Fig. 2.

Moreover, we exploit the intra-relationship between different category classes by the category class graph shown as the left subgraph in Fig. 2. Specifically, the category class graph $\mathcal{G}_C = (\mathcal{V}_C, \mathcal{E}_C)$ can be constructed based on the category class information, where $\mathcal{V}_C = \mathcal{C}$ and $\mathcal{E}_C \subseteq \mathcal{V}_C \times \mathcal{V}_C$. And we define the value of the edge connecting two category classes by $W_C(i,j) = \|S_{bi} - S_{bj}\|_F$, where $\|\quad\|_F$ means Frobenius norm and S_{bi} denotes the between class scatter matrix of the i-th category.

Different from conventional single-label classification learning problem in which classes are mutual exclusive, the annotation terms are interrelated with one another. We resort to the following cosine similarity to calculate the annotation term affinity matrix

$$W_A(i,j) = \cos(\widetilde{\mathbf{y}}_i^a, \widetilde{\mathbf{y}}_j^a) = < \widetilde{\mathbf{y}}_i^a, \widetilde{\mathbf{y}}_j^a > / (\left\|\widetilde{\mathbf{y}}_i^a\right\| \left\|\widetilde{\mathbf{y}}_j^a\right\|) \tag{2}$$

where $\widetilde{\mathbf{y}}_i^a$ and $\widetilde{\mathbf{y}}_j^a$ are the i-th and j-th rows of Y_a respectively. Thus, a graph $\mathcal{G}_A = (\mathcal{V}_A, \mathcal{E}_A)$ is induced, where $\mathcal{V}_A = \mathcal{A}$ and $\mathcal{E}_A \subseteq \mathcal{V}_A \times \mathcal{V}_A$. We call $\mathcal{G}_A$ as annotation terms subgraph shown as the right subgraph in Fig. 2.

As introduced in sec 1, the category class and annotation terms have some relations. We utilize the following cosine similarity to measure their interrelations,

$$R(i,j) = \cos(\widetilde{\mathbf{y}}_i^c, \widetilde{\mathbf{y}}_j^a) = < \widetilde{\mathbf{y}}_i^c, \widetilde{\mathbf{y}}_j^a > / (\left\|\widetilde{\mathbf{y}}_i^c\right\| \left\|\widetilde{\mathbf{y}}_j^a\right\|) \tag{3}$$

where $\widetilde{\mathbf{y}}_i^c$ is the i-th row of Y_c and $\widetilde{\mathbf{y}}_j^a$ is the j-th row of Y_a. $R \in \mathbb{R}^{K_c \times K_a}$, where $R(i,j)$ indicates how closely the i-th category class and the j-th annotation term are related.

Obviously, the subgraph $\mathcal{G}_{AX} = (\mathcal{V}_X, V_A, \mathcal{E}_{XA})$ connects G_X and G_A, whose adjacency matrix is Y_a^T. Similarly, the adjacency matrix of $\mathcal{G}_{CX} = (\mathcal{V}_X, \mathcal{V}_C, \mathcal{E}_{XC})$ is Y_c^T.

The subgraph $(\mathcal{V}_C, \mathcal{V}_A, \mathcal{E}_{CA})$ characterizes the associations between category classes and annotation terms whose adjacency matrix is R defined in Eq.(3). Connecting $\mathcal{G}_X$ and $\mathcal{G}_A$ by the annotation associations via the green dashed lines, connecting $\mathcal{G}_X$ and $\mathcal{G}_C$ by the class associations via the blue dashed lines and connecting $\mathcal{G}_C$ and $\mathcal{G}_A$ by the class-annotation association via the purple dashed lines, we construct a *Tri-Relational Graph* (TG), $G = (\mathcal{V}_X \cup \mathcal{V}_C \cup \mathcal{V}_A, \mathcal{E}_X \cup \mathcal{E}_A \cup \mathcal{E}_{XC} \cup \mathcal{E}_{XA} \cup \mathcal{E}_{CA})$, which is illustrated in Fig. 2.

In contrast to existing graph-based learning methods that only use information conveyed by $\mathcal{G}_X$ only [19,20,21], on which labeling information is propagated, we aim to simultaneously classify and annotate an unlabeled data point using all the information encoded in $\mathcal{G}$. Because all data points (images), category class and annotation terms are equally regarded as vertices on $\mathcal{G}$, our task is to measure the relevance between a category class/anntotation term vertex and a data point vertex. As each category class/annotation term has a set of associated training data points, which convey the same scene information, we consider both the category class/annotation term vertex and its labeled training image vertices as a group set. As a result, instead of measuring vertex-to-vertex relevance between a class/annotatation vertex and an unlabeled data point vertex, we may measure the set-to-vertex relevance between the group set and the unlabeled data point. Motivated by [22,23], we consider to further develop standard random walk and use its equilibrium probability to measure the relevance between a group set and an unlabeled data point.

3.2 Biased Random Walk

Standard random walk on a graph W can be described as a Markov process with transition probability $M = D^{-1}W$, where $d_i = \sum_j W(i,j)$ is the degree of vertex i and $D = diag(d_1, \cdots, d_n)$. Clearly, $M^T \neq M$ and $\sum_j M(i,j) = 1$. Let $\mathbf{p}^{(t)}$ be the distribution of the random walker at time t, the distribution at $t+1$ is given by $\mathbf{p}^{(t+1)}(j) = \sum_i \mathbf{p}^{(t)}(i)M(i,j)$. Thus the equilibrium (stationary) distribution of the random walk $\mathbf{p}^* = \mathbf{p}^{(t=\infty)}$ is determined by $M^T\mathbf{p}^*$. It is well known that the solution is simply given by $\mathbf{p}^* = W\mathbf{e}/(\sum_i d_i) = \mathbf{d}/(\sum_i d_i)$, where $\mathbf{d} = [d_1, \cdots, d_n]^T$.

It can be seen that the equilibrium distribution of a standard random walk is solely determined by the graph itself, but independent of the location where the random walk is initiated. In order to incorporate label information, we propose the following *Biased Random Walk* (BRW):

$$\mathbf{p}^{(t+1)}(j) = (1-\alpha)\sum\nolimits_i \mathbf{p}^{(t)}(i)M(i,j) + \alpha\mathbf{h}_j, \tag{4}$$

where $0 \leqslant \alpha \leqslant 1$ is a fixed parameter, and $\mathbf{h}$, called *biased distribution*, is a probability distribution such that $\mathbf{h}(i) \geqslant 0$ and $\sum_i \mathbf{h}(i) = 1$. Eq. (4) describes a random walk process in which the random walker hops on the graph W according to the transition matrix M with probability $1-\alpha$, and meanwhile it takes a preference to go to other vertices specified by $\mathbf{h}$ with probability α. The equilibrium distribution of BRW in Eq. (4) is determined by $\mathbf{p}^* = (1-\alpha)M^T\mathbf{p}^* + \alpha\mathbf{h}$, which leads to:

$$\mathbf{p}^* = \alpha[I - (1-\alpha)M^T]^{-1}\mathbf{h}. \tag{5}$$

Due to *Perron-Frobenius theorem*, the maximum eigenvalue of M is less than $\max_i \sum_j M(i,j) = 1$. Thus, $I - (1-\alpha)M^T$ is positive definite and invertible. Eq. (4) takes a similar form with respect to two existing works: random walk with restart (RWR) method [23] and PageRank algorithm [22]. In the former, **h** is a vector with all entries to be 0 except one entry to be 1 indicating the vertex where the random walk could be restarted; while in the latter, **h** is a constant vector called as damping factor [22]. In contrast, the biased distribution vector **h** in Eq. (4) is a generic probability distribution, which is flexible thereby more powerful. Most importantly, through **h** we can assess group-to-vertex relevance, while RWR and PageRank methods measure vertex-to-vertex relevance. Similar to RWR [23], when we set the **h** to be a probability distribution in which all the entries are 0 except for those corresponding to $\mathcal{G}_k$, $\mathbf{p}^*(i)$ measures how relevant the k-th group is to the i-th vertex on $\mathcal{G}$.

3.3 Biased Random Walk on Tri-relational Graph

In order to classify unlabeled data points using the equilibrium probabilities in Eq. (5) of the BRW on TG, we need to construct the transition matrix M and the biased distribution **h** from $\mathcal{G}$.

Construction of the Transition Matrix M

Let

$$M = \begin{bmatrix} M_X & M_{XC} & M_{XA} \\ M_{CX} & M_C & M_{CA} \\ M_{AX} & M_{AC} & M_A \end{bmatrix}, \tag{6}$$

where M_X, M_C and M_A are the intra-subgraph transition matrices of $\mathcal{G}_X$, $\mathcal{G}_C$ and $\mathcal{G}_A$ respectively, and the rest 6 sub-matrices are the inter-subgraph transition matrices among $\mathcal{G}_X$, $\mathcal{G}_C$ and $\mathcal{G}_A$. Let $\beta_1 \in [0,1]$ be the jumping probability, *i.e.*, the probability that a random walker hops from $\mathcal{G}_X$ to $\mathcal{G}_C$ and vice versa. And let $\beta_2 \in [0,1]$ be the jumping probability from $\mathcal{G}_X$ to $\mathcal{G}_A$ or vice versa. Therefore, β_1 and β_2 regulates the reinforcement between $\mathcal{G}_X$ and one of the other two subgraphs. When both $\beta_1 = 0$ and $\beta_2 = 0$, the random walk are performed independently on $\mathcal{G}_X$, which is equivalent to existing graph-based learning methods using the data graph $\mathcal{G}_X$ only. Similarly, we define λ, $\lambda \in [0,1]$ as the jumping probability from $\mathcal{G}_C$ to $\mathcal{G}_A$ or vice versa.

During a random walk process, if the random walker is on a vertex of the data subgraph which has at least one connection to the either of the other two subgraphs, she can hop to the category class subgraph with probability β_1 or annotation terms subgraph with probability β_2, or stay on the data subgraph with probability $1 - \beta_1 - \beta_2$ hopping to other vertices of the data subgraph. If the random walker is on a vertex of the data subgraph without any connection to the category class subgraph or annotation terms subgraph, she stays on the data subgraph, hopping to other vertices on the same subgraph as the case of standard random walk process. To be more precise, let $d_i^{Y_c^T} = \sum_j Y_c^T(i,j)$, the transition probability from $\mathbf{x}_i$ to $\mathbf{c}_j$ is defined as following:

$$p(\mathbf{c}_j|\mathbf{x}_i) = M_{XC}(i,j) = \begin{cases} \beta_1 Y_c^T(i,j)/d_i^{Y_c^T}, & d_i^{Y_c^T} > 0, \\ 0, & otherwise. \end{cases} \tag{7}$$

Similarly, let $d_i^{Y_c} = \sum_j Y_c(i,j)$, the transition probability from $\mathbf{c}_i$ to $\mathbf{x}_j$ is:

$$p(\mathbf{x}_j|\mathbf{c}_i) = M_{CX}(i,j) = \begin{cases} \beta_1 Y_c(i,j)/d_i^{Y_c}, & d_i^{Y_c} > 0, \\ 0, & otherwise. \end{cases} \tag{8}$$

Following the same definition, the rest four inter-subgraph transition probability matrices are defined as:

$$p(\mathbf{a}_j|\mathbf{x}_i) = M_{XA}(i,j) = \begin{cases} \beta_2 Y_a^T(i,j)/d_i^{Y_a^T}, & if\ d_i^{Y_a^T} > 0, \\ 0, & otherwise. \end{cases} \tag{9}$$

$$p(\mathbf{x}_j|\mathbf{a}_i) = M_{AX}(i,j) = \begin{cases} \beta_2 Y_a(i,j)/d_i^{Y_a}, & if\ d_i^{Y_a} > 0, \\ 0, & otherwise. \end{cases} \tag{10}$$

where $d_i^{Y_a^T} = \sum_j Y_a^T(i,j)$ and $d_i^{Y_a} = \sum_j Y_a(i,j)$, and

$$p(\mathbf{c}_j|\mathbf{a}_i) = M_{AC}(i,j) = \begin{cases} \lambda R^T(i,j)/d_i^{R^T}, & if\ d_i^{R^T} > 0, \\ 0, & otherwise. \end{cases} \tag{11}$$

$$p(\mathbf{a}_j|\mathbf{c}_i) = M_{CA}(i,j) = \begin{cases} \lambda R(i,j)/d_i^{R}, & if\ d_i^{R} > 0, \\ 0, & otherwise \end{cases} \tag{12}$$

where $d_i^{R^T} = \sum_j R^T(i,j)$ and $d_i^R = \sum_j R(i,j)$. Let $d_i^X = \sum_j W_X(i,j)$, $d_i^Y = \sum_j Y(i,j)$, $d_i^{Q_a} = \sum_j Q_a(i,j)$, $d_i^{Q_c} = \sum_j Q_c(i,j)$ where $Q_a = R + Y_a$ and $Q_c = R^T + Y_c$.
The data subgraph intra transition probability from $\mathbf{x}_i$ to $\mathbf{x}_j$ is computed as:

$$p(\mathbf{x}_j|\mathbf{x}_i) = M_X(i,j) = \begin{cases} (1-\beta_1-\beta_2)W_X(i,j)/d_i^X,\ if\ d_i^{Y^T} > 0 \\ W_X(i,j)/d_i^X,\ otherwise \end{cases} \tag{13}$$

Similarly, let $d_i^A = \sum_j W_A(i,j)$, the annotation terms subgraph intra transition probability from $\mathbf{a}_i$ to $\mathbf{a}_j$ is:

$$p(\mathbf{a}_j|\mathbf{a}_i) = M_A(i,j) = \begin{cases} (1-\beta_2-\lambda)W_A(i,j)/d_i^A,\ if\ d_i^{Q_a^T} > 0 \\ W_A(i,j)/d_i^A,\ otherwise \end{cases} \tag{14}$$

let $d_i^C = \sum_j W_C(i,j)$, the category class subgraph intra transition probability from $\mathbf{c}_i$ to $\mathbf{c}_j$ is:

$$p(\mathbf{c}_j|\mathbf{c}_i) = M_C(i,j) = \begin{cases} (1-\beta_1-\lambda)W_C(i,j)/d_i^C,\ if\ d_i^{Q_c^T} > 0 \\ W_C(i,j)/d_i^C,\ otherwise \end{cases} \tag{15}$$

It can be easily verified that, $\sum_j M(i,j) = 1$, *i.e.*, M is a stochastic matrix.

Construction of the Biased Distribution H
The biased distribution vector specifies a group of vertices to which the random walker prefers to moving in every iteration step. The relevance between this group and an vertex

is measured by the equilibrium distribution of the random walk process. Therefore, we construct $K = K_c + K_a$ biased distribution vectors, one for each semantic group G_k:

$$\mathbf{h}^{(k)} = \begin{bmatrix} \gamma \mathbf{h}_{\mathcal{X}}^{(k)} \\ (1-\gamma)\mathbf{h}_{\mathcal{L}}^{(k)} \end{bmatrix} \in \mathbb{R}_{+}^{n+K} \tag{16}$$

where $\mathbf{h}_{\mathcal{X}}^{(k)}(i) = 1/\sum_i \mathbf{y}_i(k)$ if $\mathbf{y}_i(k) = 1$ and $\mathbf{h}_{\mathcal{X}}^{(k)}(i) = 0$, otherwise; $\mathbf{h}_{\mathcal{L}}^{(k)}(i) = 1$, if $i = k$, $\gamma \in [0, 1]$ controls how much the random walker prefers to go to the data subgraph $\mathcal{G}_{\mathcal{X}}$ and other two subgraphs $\mathcal{G}_{\mathcal{C}}$, $\mathcal{G}_{\mathcal{A}}$. It can be verified that $\sum_i \mathbf{h}^{(k)}(i) = 1$, *i.e*, $\mathbf{h}^{(k)}$ is a probability distribution. Let I_K be the identity matrix of size $K \times K$, we write

$$H = [\mathbf{h}^{(1)}, \cdots, \mathbf{h}^{(K)}] = \begin{bmatrix} \gamma H_X \\ (1-\gamma) I_K \end{bmatrix} \tag{17}$$

BRW on TG

Given the TG of a data set, using the transition matrix M defined in Eq. (6) and the biased probability matrix H defined in Eq. (17), we can perform BRW on the TG. According to Eq. (5), its equilibrium distribution matrix P^* is computed as:

$$P^* = \alpha[I - (1-\alpha)M^T]^{-1}H, \tag{18}$$

$P^* = [\mathbf{p}_1^*, \cdots, \mathbf{p}_K^*] \in \mathbb{R}^{(n+K)\times K}$, and $\mathbf{p}_k^*$ is the equilibrium distribution of the BRW taking the k-th semantic group as preference. Therefore, $\mathbf{p}_k^*(i)$ $(l+1 \leqslant i \leqslant n)$ measures the relevance between the k-th class and an unlabeled test image $\mathbf{x}_i$. We can predict the category class from the block P_{nc} obtained from the matrix P^* using Eq. (18) and select annotation terms for $\mathbf{x}_i$ using the adaptive decision boundary method [21] on block P_{na} which is calculated from matrix P^*, where $P_{nc} = \begin{bmatrix} P^*(l+1,1) & \cdots & P^*(l+1,K_c) \\ \vdots & \ddots & \vdots \\ P^*(n,1) & \cdots & P^*(n,K_c) \end{bmatrix}$, $P_{na} = \begin{bmatrix} P^*(l+1,K_c+1) & \cdots & P^*(l+1,K_c+K_a) \\ \vdots & \ddots & \vdots \\ P^*(n,K_c+1) & \cdots & P^*(n,K_c+K_a) \end{bmatrix}$
Since the category class prediction is a single-label classification problem, we select the category class $y(i)^{c*}$ with the maximum probability as the predicted category class for image x_i.

$$y(i)^{c*} = \arg\max_k(\widetilde{\mathbf{p}}_i^c) \tag{19}$$

where $\widetilde{\mathbf{p}}_i^c$ is the i-th row vector of matrix P_{nc}. Up to here, we are able to achieve the goal, that is, to predict the category class and annotation terms for the given test image simultaneously.

4 Experiment and Results

In this section, we will first briefly introduce the two datasets that we used in our experiment, followed by the feature extraction and experiment setup. After that, we will demonstrate the results of image classification and image annotation using our approach with the comparison of the state-of-art methods.

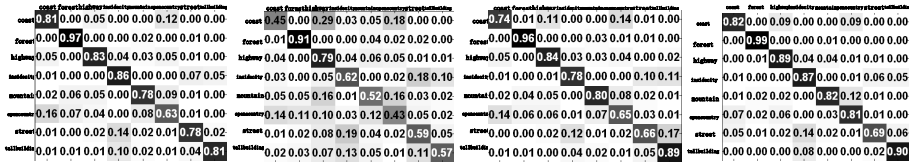
4.1 Data Description and Feature Extraction

We used two benchmark colorful image datasets, *i.e.* LabelMe [5] data and the UIUC-sport data [4]. Each image in our experiment contains one category class and several annotation terms. LabelMe dataset consists of eight category classes: coast, forest, highway, inside-city, mountain, open-country, street and tall-building. In order to keep the balance of the number of images for each class, followed [15] we unified the size of each image as $256 \times 256 \times 3$ and then randomly selected 200 images for each category class. Thus, the total number of images is 1600. The UIUC-sports dataset is an event dataset composed of eight classes as well: badminton, bocce, croquet, polo, rockclimbing, sailing and snowboarding. The number of images in each category class varies from 136 (bocce) to 250 (rowing). The total number of images is 1578. In both datasets, we remove the annotation terms that occurred less than 3 times. We refined the annotation terms based on the following two reasons. On one hand, if the number of data point with a certain annotation term is too small, from statistic point of view, it is hard let the machine learn this annotation efficiently. On the other hand, since we will do two-fold cross validation in our experiment, if the annotation terms correspond to too small number of data points, it is hard for us to evenly split the data, obtaining the training sets and testing sets. At last, we get a refined LabelMe data set with 58 distinct annotation terms and a refined UIUC-sports data set with 90 distinct annotation terms. On average, there are 4 terms per annotation in the refined LabelMe data and 6 terms per annotation in the refined UIUC-sports data.

Following the setting in [15], we used the 128-dimensional SIFT region descriptors to represent a sliding grid (5×5) and choose 256 as the number of codewords created by K-means algorithm. Therefore, after the visual quantization, we get a histogram with 256 dimension to be the descriptor for the corresponding image data.

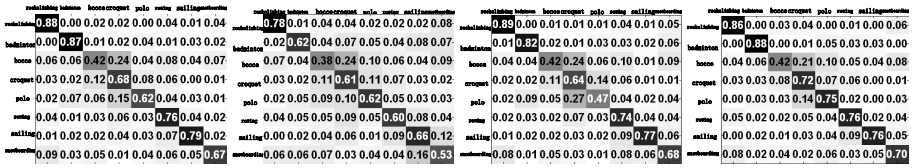
4.2 Experimental Setup

In our experiments, we found the following five parameters are not sensitive in certain ranges with good performance. β_1, β_2 and λ controls the jumping between different subgraphs and could not affect the result much if they are assigned in the range from 0.1 to 0.4. α controls initial bias of the random walker and will get stable result if it is assigned in the range of $(0, 0.2)$. γ controls how much the random walker prefers to go to the data subgraph or to go to other two subgraphs. We set it as 0.5 since we consider data subgraph and the other two subgraphs are equally important. Besides these parameters, we also need to initialize the category class as well as annotation terms for the test data. Practically, we can use any single-label multi-class classification method to get the initialized category class $\hat{\mathbf{y}}_i^c$ and use any multi-label multi-class classification approach to initialize annotation terms $\hat{\mathbf{y}}_i^a$. Although the initializations are not completely correct, a big portion of them should be (assumed to be) correctly predicted. Our joint classification framework will self-consistently amend the incorrect labels for class and annotation, which will be shown in the coming experimental results. In our experiments, we use k-nearest neighbor (KNN) method to do the above initializations because of its simplicity and clear intuition. We use $k = 1$ and abbreviate it as 1NN. We use 2-fold cross validation in our experiment to calculate the average results.



(a) SVM: avg. accuracy: 81% (b) KNN: avg. accuracy: 61% (c) JLDA: avg. accuracy: 79% (d) BRW: avg. accuracy: 84%

Fig. 3. Classification results in terms of confusion matrices on LabelMe data



(a) SVM: avg. accuracy: 71% (b) KNN: avg. accuracy: 60% (c) JLDA: avg. accuracy: 68% (d) BRW: avg. accuracy: 74%

Fig. 4. Classification results in terms of confusion matrices on UIUC data

4.3 Joint Classification and Annotation Results

We evaluate the performance of our method on both classifications and annotations tasks. We firstly train our model with classified and annotated images. And then, we evaluate the prediction performance of our proposed method on unlabeled as well as unsegmented test images.

Image Classifications. To evaluate our method's classification performance, we compared our method with the state-of-the-art method JLDA [15], support vector machine (SVM) with radial basis function (RBF) kernel [24] and 1NN on benchmark datasets. As to JLDA, we downloaded the published code from the author's website and reported the best classification results using different number of topics (from 60 to 120). For SVM, we tuned the parameters C and γ based on the training data only and used the trained model using optimal parameters to do the prediction. We demonstrated the category class prediction results by confusion matrices. From the resultant confusion matrices as shown in Fig. 3 and Fig. 4, we can see that the prediction accuracy of our proposed method is higher than the results of the other methods for more than 2 percent on both datasets.

Image Annotations. We also validate our method by predicting the annotation terms on these two benchmark datasets. Two standard multi-label classification performance metrics precision and F1 score are used to evaluate image annotation performances. Again, we used LIBSVM [24] but considering annotation terms independently by one-vs-all strategy. We compared the performance of our method with four famous multi-label classification methods as well, *i.e.*, Harmonic function (HF) [25], random walk (RW) [26] which considers the data graph and annotation terms graph only, local shared subspace (LS) [9] and LGC [27]. Table 1 shows that our method can consistently beat the other methods evaluated by all the metrics on both datasets.

Table 1. The comparison of average annotation prediction results using 2-fold cross validations

Dataset	Method	Metric			
		Macro avg. Precision	Macro avg. F1	Micro avg. Precision	Micro avg. F1
LabelMe	SVM[24]	0.4253	0.4179	0.1394	0.2013
	HF[25]	0.2569	0.3942	0.1273	0.1061
	RW[26]	0.2569	0.3942	0.1273	0.1061
	LS[9]	0.3688	0.3603	0.1369	0.1385
	LGC[27]	0.4592	0.4578	0.1402	0.1967
	BRW(our)	**0.4923**	**0.5038**	**0.1749**	**0.2071**
UIUC-sports	SVM[24]	0.4376	0.4451	0.1946	0.2018
	HF[25]	0.1349	0.2259	0.1514	0.1449
	RW[26]	0.2569	0.3942	0.1273	0.1061
	LS[9]	0.2576	0.2469	0.1657	0.1791
	LGC[27]	0.4328	0.4357	0.1932	0.1937
	BRW(our)	**0.4679**	**0.4508**	**0.2108**	**0.2052**

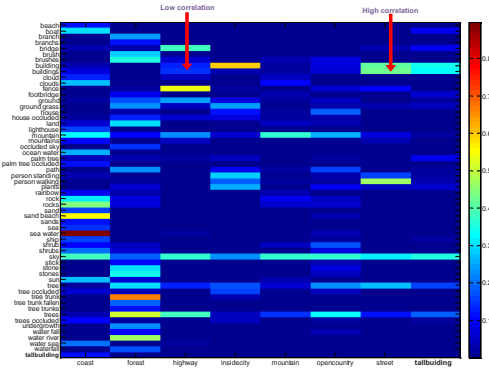


Fig. 5. The annotation-class correlation on LabelMe data

5 Discussion

Fig. 5 demonstrates the annotation-class correlation matrix. Highlighted by the red arrows on the top, we can see that the annotation term "building" has higher correlation with the category class "street". However it has lower correlation with the category class "highway". In other words, the annotation "building" can give us a hint to some extent, implicitly indicating that it has a higher probability with category class "street" than category class "highway". Fig. 7 further illustrates that with the help of annotation terms, we can boost the classification performance. Since we initialize W_C and W_A as the symmetric matrices before the random walk starts, we assume the hidden class-to-class and annotation-to-annotation relationship are both symmetric. However, in practice, they are not. For example, boat is usually seen in the image with sea background but not vice versa. Our learned P^*_{cc} and P^*_{aa} from $P^* = \begin{bmatrix} P^*_{nc} & P^*_{na} \\ P^*_{cc} & P^*_{ca} \\ P^*_{ac} & P^*_{aa} \end{bmatrix}$ can reflect

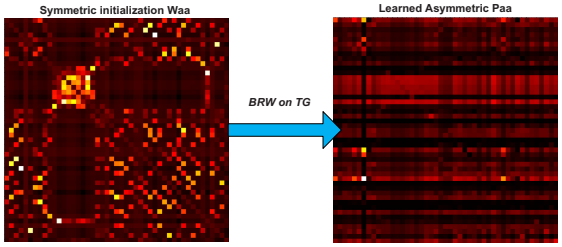
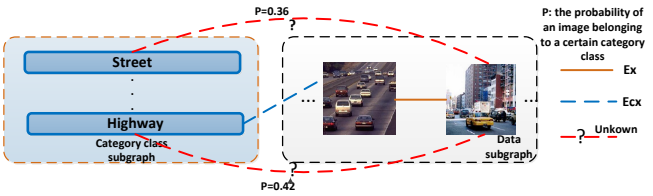
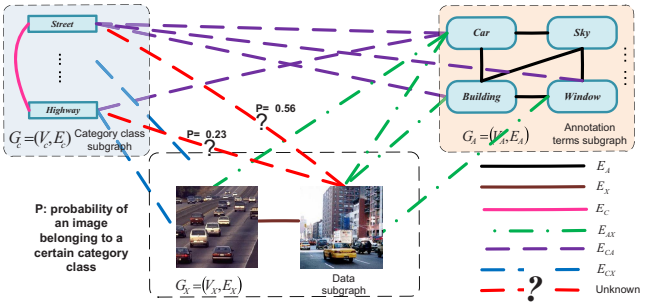


Fig. 6. The learned asymmetric annotation-to-annotation matrix by BRW on TG on labelMe data



(a) without the help of annotation



(b) with the help of annotation

Fig. 7. Illustration on how clarification of *annotation terms* help the classification of the *category* of a test image (shown on the right in the data subgraph). Shown on the left in the data subgraph is the image (from the database) which is the nearest neighbor of the test image. Two main features/characteristics in both images are road and car. Based on these two main features **only**, the category of the test image is ambiguous — it could be either street or highway. However, with the clarification of the test image's additional *annotation terms* of sky, building, windows and *etc*, our system assigns street *category* to the test image.

more accurate asymmetric relationships, which is shown in Fig. 6. And using such kind of asymmetric information, both of our classification and annotation tasks achieve superior performance than the state-of-art methods. The sample joint prediction results are shown in Fig. 8.

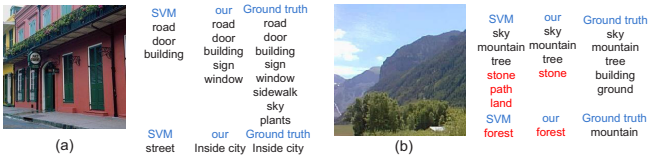


Fig. 8. The prediction results of two test images by SVM (left column), by our method (middle) and ground truth (right column). Annotation predictions are on the top, while the category class prediction result is on the bottom (incorrect annotation or category class is marked as red). (a) illustration of good prediction results and (b) failure cases where incorrect annotations affect the category class prediction and vice versa.

6 Conclusion

In this paper, we proposed a novel *Tri-relational Graph* (TG) model to jointly learn the interrelations between category class prediction and terms annotation inspired by the intuition that image classification and image annotation are related. The standard bag-of-word features were used to describe the images. A new *Biased Random Walk* (BRW) method was introduced to simultaneously propagate the category class and annotation terms information via TG model. Both category class prediction and terms annotation tasks are jointly completed. Besides that, our method can also exploit the more accurate intrinsic asymmetric annotation-to-annotation, class-to-class intra-relationships. We evaluated the proposed method on two popular datasets. The experimental results demonstrated our joint learning method can achieve superior prediction results on both tasks than the state-of-the-art methods.

Acknowledgement. This research was partially supported by NSF CCF-0830780, CCF-0917274, DMS-0915228, and IIS-1117965.

References

1. Wang, H., Huang, H., Ding, C.H.Q.: Image annotation using bi-relational graph of images and semantic labels. In: CVPR, pp. 793–800 (2011)
2. Cai, X., Wang, H., Huang, H., Ding, C.H.Q.: Joint stage recognition and anatomical annotation of *drosophila* gene expression patterns. Bioinformatics 28(12), 16–24 (2012)
3. Fei-Fei, L., Perona, P.: A bayesian hierarchical model for learning natural scene categories. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 2, pp. 524–531. IEEE (2005)
4. Li, L., Fei-Fei, L.: What, where and who? classifying events by scene and object recognition. In: IEEE 11th International Conference on Computer Vision, ICCV 2007, pp. 1–8. IEEE (2007)
5. Russell, B., Torralba, A., Murphy, K., Freeman, W.: LabelMe: a database and web-based tool for image annotation. International Journal of Computer Vision 77(1), 157–173 (2008)
6. Hwang, S.J., Sha, F., Grauman, K.: Sharing features between objects and their attributes. In: CVPR, pp. 1761–1768 (2011)
7. Bosch, A., Zisserman, A., Muñoz, X.: Scene Classification Via pLSA. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part IV. LNCS, vol. 3954, pp. 517–530. Springer, Heidelberg (2006)

8. Hofmann, T.: Unsupervised learning by probabilistic latent semantic analysis. Machine Learning 42(1), 177–196 (2001)
9. Ji, S., Tang, L., Yu, S., Ye, J.: Extracting shared subspace for multi-label classification. In: Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 381–389. ACM (2008)
10. Yu, K., Yu, S., Tresp, V.: Multi-label informed latent semantic indexing. In: SIGIR (2005)
11. Wang, H., Ding, C.H.Q., Huang, H.: Multi-label classification: Inconsistency and class balanced k-nearest neighbor. In: AAAI (2010)
12. Wang, H., Huang, H., Ding, C.: Multi-label Feature Transform for Image Classifications. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 793–806. Springer, Heidelberg (2010)
13. Wang, H., Ding, C., Huang, H.: Multi-label Linear Discriminant Analysis. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part VI. LNCS, vol. 6316, pp. 126–139. Springer, Heidelberg (2010)
14. Wang, H., Huang, H., Ding, C.: Function-Function Correlated Multi-Label Protein Function Prediction over Interaction Networks. In: Chor, B. (ed.) RECOMB 2012. LNCS, vol. 7262, pp. 302–313. Springer, Heidelberg (2012)
15. Wang, C., Blei, D., Li, F.: Simultaneous image classification and annotation (2009)
16. Blei, D., McAuliffe, J.: Supervised topic models. In: Advances in Neural Information Processing Systems, vol. 20, pp. 121–128 (2008)
17. Blei, D., Jordan, M.: Modeling annotated data. In: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval, pp. 127–134. ACM (2003)
18. Zelnik-Manor, L., Perona, P.: Self-tuning spectral clustering. In: Advances in Neural Information Processing Systems, vol. 17(1601-1608), p. 16 (2004)
19. Kang, F., Jin, R., Sukthankar, R.: Correlated label propagation with application to multi-label learning. In: CVPR (2006)
20. Zha, Z., Mei, T., Wang, J., Wang, Z., Hua, X.: Graph-based semi-supervised learning with multi-label. In: ICME (2008)
21. Wang, H., Huang, H., Ding, C.: Image annotation using multi-label correlated Green's function. In: ICCV (2009)
22. Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. In: WWW (1998)
23. Tong, H., Faloutsos, C., Pan, J.: Fast random walk with restart and its applications. In: ICDM (2006)
24. Chang, C., Lin, C.: LIBSVM: a library for support vector machines (2001)
25. Zhu, X., Ghahramani, Z., Lafferty, J.D.: Semi-supervised learning using gaussian fields and harmonic functions. In: ICML, pp. 912–919 (2003)
26. Zhou, D., Schölkopf, B.: Learning from Labeled and Unlabeled Data Using Random Walks. In: Rasmussen, C.E., Bülthoff, H.H., Schölkopf, B., Giese, M.A. (eds.) DAGM 2004. LNCS, vol. 3175, pp. 237–244. Springer, Heidelberg (2004)
27. Zhou, D., Bousquet, O., Lal, T.N., Weston, J., Schölkopf, B.: Learning with local and global consistency. In: NIPS (2003)