Image Annotation Using Bi-Relational Graph of Images and Semantic Labels

Hua Wang, Heng Huang and Chris Ding
Department of Computer Science and Engineering
University of Texas at Arlington, Arlington, Texas 76019, USA
huawangcs@gmail.com, heng@uta.edu, chqding@uta.edu

Abstract

Image annotation is usually formulated as a multi-label semi-supervised learning problem. Traditional graph-based methods only utilize the data (images) graph induced from image similarities, while ignore the label (semantic terms) graph induced from label correlations of a multi-label image data set. In this paper, we propose a novel Bi-relational Graph (BG) model that comprises both the data graph and the label graph as subgraphs, and connect them by an additional bipartite graph induced from label assignments. By considering each class and its labeled images as a semantic group, we perform random walk on the BG to produce group-to-vertex relevance, including class-to-image and class-to-class relevances. The former can be used to predict labels for unannotated images, while the latter are new class relationships, called as Causal Relationships (CR), which are asymmetric. CR is learned from input data and has better semantic meaning to enhance the label prediction for unannotated images. We apply the proposed approaches to automatic image annotation and semantic image retrieval tasks on four benchmark multi-label image data sets. The superior performance of our approaches compared to state-of-the-art multi-label classification methods demonstrate their effectiveness.

1. Introduction

Image annotation is a challenging but important computer vision task to understand digital multimedia contents for browsing, searching, and navigation. In a typical image annotation problem, each picture is usually associated with a number of different semantic keywords. This poses so-called *Multi-Label Classification (MLC)* problem, in which each object may be associated with more than one class label. MLC problem is more general than traditional *Single-Label Classification (SLC)* problem, in which each object belongs to one only one class.

An important difference between single-label classification and multi-label classification lies in that the classes in the former are assumed to be mutually exclusive while those in the latter are normally interdependent on one another. For example, "sea" and "ship" tend to appear in a same image, whereas "sky" usually does not appear together with "indoor". As a result, many MLC algorithms have been developed to exploit label correlations to improve the overall classification performance. Two ways are broadly used to employ label correlations: incorporating label correlations into the existing graph-based label propagation learning algorithms, as either part of graph weight [3, 5] or an additional constraint [14, 17, 19]; or utilizing label correlations to seek a more discriminative subspace [4, 13, 15, 18]. Besides, a variety of different mechanisms are also used to take advantage of label correlations, such as matrix factorization [6], maximizing label entropy [20], directed graph [11, 16], and many others [7, 9, 12].

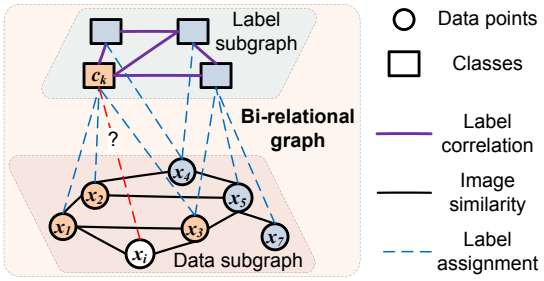


Figure 1. The Bi-relational Graph (BG) constructed from a multi-label image data set. The orange vertices, including the class vertex c_k and image vertices $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ form the k-th semantic group G_k. Our task is to determine the relevance between G_k and an unannotated image $\mathbf{x}_i$.

Due to the lack of labeled data in real world applications, image annotation is usually formulated as a semi-supervised learning problem. Traditional graph-based semi-supervised image annotation methods [3, 14, 19] only make use of the data graph induced from image similarities. However, multi-label image data present a new opportunity to improve classification accuracy through label correlations, which induce another graph among the semantic classes as shown in Figure 1.

In order to utilize both data graph and label graph, in this paper we present a new perspective for semi-supervised image annotation using Bi-relational Graph (BG) constructed from a multi-label image data set. As schematically illustrated in Figure 1, in the constructed BG both data graph and label graph exist as subgraphs, which are connected by an additional bipartite graph induced from label assignments. Consequently, both images and semantic classes are equally regarded as vertices, and image annotation is transformed to a new problem to measure how closely a class is related to an image. Toward this end, we further develop the random walk with restart (RWR) [8] model that measures vertex-to-vertex relevance. We consider a class vertex and its training image vertices as a semantic group, such as the orange vertices in Figure 1, and assess group-to-vertex, *i.e.*, class-to-image, relevance between the semantic group and a vertex. We use the resulted class-to-image scores to predict labels for unannotated images.

Because the proposed Bi-relational Graph (BG) approach performs random walks on the BG that comprises both image vertices and class vertices, the resulted equilibrium distributions measure the relevances not only between class and image but also between class and class. Namely, our approach is able to learn relationships between classes from input data. Different from symmetric label correlations used in existing MLC methods, the learned class relationships by our approach are asymmetric, which, though, are more close to real semantic relations. For example, as shown in Figure 2 an image with label "car" usually has label "road", whereas an image with label "road" may not also has label "car", because it could have other objects such as "bicycle". As a result, the relationship from an object class to a background class (green arrow) should be greater than that of the reverse (red arrow). That is, the learned asymmetric class relationships, called as *Causal Relationships (CR)* , by our approach is able to better reflect the true semantic relationships. Thanks to the nature of random walk formulation, the asymmetric CR can be naturally used in the proposed BG approach, while most, if not all, existing MLC methods can only work with symmetric label correlations.

To summarize, our main contributions include:

- We present a new perspective for multi-label semi-supervised learning to use a novel Bi-relational Graph (BG) model that consists of data graph as well as label graph. The proposed BG model considers both class and image equally as vertices, such that the learning model can be built upon existing single-label classification methods while label correlations are still leveraged.
- We show that asymmetric label correlations learned by our approach are more close to real semantic relationships. By making use of them, the image annotation performance by our approach is improved.

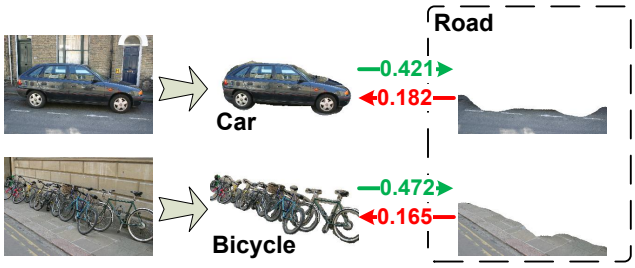


Figure 2. "Car" often co-occurs with "Road" in images. The proposed method learns more than this simple symmetric relationship. For example, the presence of "Car" induces the presence of "Road" with probability 0.421, whereas the presence of "Road" causes the presence of "Car" with probability of 0.182 (see Section 5.1). This surprising asymmetric relationship is in fact not difficult to understand, because an image with label "Car" usually has "Road" beneath the car, whereas "Road" could be part of city landscape or a road with walking people. In other words, "Road" is a concept which is not necessarily connected with "Car", whereas a "Car" usually runs on a "Road".

- Promising experimental results in both automatic image annotation task and semantic concept retrieval task on four benchmark multi-label image data sets validate the effectiveness of the proposed method.

2. Label prediction using BG approach

In this section, we propose a novel Bi-relational Graph (BG) construction method for multi-label image data. Then we present our semi-supervised learning method using the resulted BG for image annotation.

Problem formalization. For an image annotation task, we have n images $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ and K semantic classes $\mathcal{C} = \{c_1, \ldots, c_K\}$. Each image is abstracted as a data point $\mathbf{x}_i \in \mathbb{R}^d$, which is associated with a number of labels $\mathcal{L}_i \subseteq \mathcal{C}$ represented by a binary vector $\mathbf{y}_i \in \{0,1\}^K$, such that $\mathbf{y}_i(k) = 1$ if $\mathbf{x}_i$ belongs to the k-th class, and 0 otherwise. We write $Y = [\mathbf{y}_1, \ldots, \mathbf{y}_n]$. We are given the pairwise similarities between the images, denoted as $W_X \in \mathbb{R}^{n \times n}$ with $W_X(i,j)$ measuring how closely $\mathbf{x}_i$ and $\mathbf{x}_j$ are related. Suppose that the first l images are annotated. Our goal is to predict labels $\{\mathcal{L}_i\}_{i=l+1}^n$ for the unannotated images.

Throughout this paper, we denote a vector as boldface lowercase character and a matrix as an uppercase character. The entry of a vector $\mathbf{v}$ is denoted as $\mathbf{v}(\cdot)$ and the entry of a matrix M is denoted as $M(\cdot,\cdot)$.

2.1. Bi-relational graph for multi-label image data

Traditional graph-based semi-supervised learning methods [5, 14, 19] only consider the data graph $\mathcal{G}_X = (\mathcal{V}_X, \mathcal{E}_X)$ induced from W_X, where $\mathcal{V}_X = \mathcal{X}$ and $\mathcal{E}_X \subseteq \mathcal{V}_X \times \mathcal{V}_X$. Different from single-label data, the classes in multi-label data are interrelated to each other, and label correlations W_L

induces another graph $\mathcal{G}_L = (\mathcal{V}_L, \mathcal{E}_L)$, where $\mathcal{V}_L = \mathcal{C}$ and $\mathcal{E}_L \subseteq \mathcal{V}_L \times \mathcal{V}_L$. We aim to leverage the both graphs.

Specifically, as in Figure 1, we consider a graph $\mathcal{G} = (\mathcal{V}_X \cup \mathcal{V}_L, \mathcal{E}_X \cup \mathcal{E}_L \cup \mathcal{E}_R)$, where $\mathcal{E}_R \subseteq \mathcal{V}_X \times \mathcal{V}_L$. Obviously, both the data graph $\mathcal{G}_X$ and the label graph $\mathcal{G}_L$ are subgraphs of $\mathcal{G}$, and they are connected by a bipartite graph $\mathcal{G}_R = (\mathcal{V}_X, \mathcal{V}_L, \mathcal{E}_R)$. $\mathcal{G}_R$ abstracts the association between the images and the semantic classes, and its adjacency matrix is Y^T. Because $\mathcal{G}$ characterizes two types of entities, images and semantic classes, with $\mathcal{E}_X$ and $\mathcal{E}_L$ describing the respective intra-type relations and $\mathcal{E}_R$ describing the inter-type relations, we call $\mathcal{G}$ as Bi-relational Graph (BG).

Transition probability matrix M on BG. Given a BG, say $\mathcal{G}$, abstracted from a multi-label image data set, we may construct the transition probability matrix M for a random walk on it as following:

$$M = \begin{bmatrix} M_X & M_{XL} \\ M_{LX} & M_L \end{bmatrix}, \tag{1}$$

where M_X and M_L are the intra-subgraph transition probability matrices of $\mathcal{G}_X$ and $\mathcal{G}_L$ respectively, and M_{XL} and M_{LX} are the inter-subgraph transition probability matrices between $\mathcal{G}_X$ and $\mathcal{G}_L$. Let $\beta \in [0, 1]$ be the jumping probability, *i.e.*, the probability that a random walker hops from $\mathcal{G}_X$ to $\mathcal{G}_L$ or vice versa. Thus, β regulates the reinforcement between the two subgraphs. When $\beta = 0$, the random walk is performed on one of the two subgraphs, which is equivalent to existing graph-based semi-supervised learning methods only using the data graph $\mathcal{G}_X$.

As shown in Figure 1, not all the images are associated with sematic classes. During a random walk process, if the random walker is on a vertex of the data subgraph that has at least one connection to the label subgraph, such as vertex $\mathbf{x}_1$ in Figure 1, she can hops to the label subgraph with probability β, or stay on the data subgraph with probability $1 - \beta$ and hop to other vertices of the data subgraph. If the random walker is on a vertex of the data subgraph without a connection to the label subgraph, she stays on the data subgraph and hops to other vertices of it as standard random walk. To be more precise, let $d_i^{Y^T} = \sum_j Y^T(i, j)$, the transition probability from $\mathbf{x}_i$ to c_j is defined as:

$$p(c_j|\mathbf{x}_i) = M_{XL}(i, j) = \begin{cases} \beta Y^T(i, j) / d_i^{Y^T}, & \text{if } d_i^{Y^T} \geq 0, \\ 0, & \text{otherwise .} \end{cases} \tag{2}$$

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

$$p(\mathbf{x}_j|c_i) = M_{LX}(i, j) = \begin{cases} \beta Y(i, j) / d_i^Y, & \text{if } d_i^Y \geq 0, \\ 0, & \text{otherwise .} \end{cases} \tag{3}$$

Let $d_i^X = \sum_j W_X(i, j)$, the intra-subgraph transition probability inside $\mathcal{G}_X$ from $\mathbf{x}_i$ to $\mathbf{x}_j$ is computed as:

$$p(\mathbf{x}_j|\mathbf{x}_i) = M_X(i, j) = \begin{cases} W_X(i, j) / d_i^X & \text{if } d_i^{Y^T} = 0, \\ (1 - \beta) W_X(i, j) / d_i^X, & \text{otherwise .} \end{cases} \tag{4}$$

Similarly, let $d_i^L = \sum_j W_L(i, j)$, the intra-subgraph transition probability inside $\mathcal{G}_L$ from c_i to c_j is:

$$p(c_j|c_i) = M_L(i, j) = \begin{cases} W_L(i, j) / d_i^L & \text{if } d_i^Y = 0, \\ (1 - \beta) W_L(i, j) / d_i^L, & \text{otherwise .} \end{cases} \tag{5}$$

We write Eqs. (2–5) together in a concise matrix form following the definition of M in Eq. (1) as following:

$$M = \begin{bmatrix} (1 - \beta) D_X^{-1} W_X & \beta D_{Y^T}^{-1} Y^T \\ \beta D_Y^{-1} Y & (1 - \beta) D_L^{-1} W_L \end{bmatrix}, \tag{6}$$

where

$$\begin{aligned} D_{Y^T} &= \operatorname{diag}\left(d_1^{Y^T}, \ldots, d_n^{Y^T}\right), \\ D_Y &= \operatorname{diag}\left(d_1^Y, \ldots, d_K^Y\right), \\ D_X &= \operatorname{diag}\left(d_1^X, \ldots, d_n^X\right), \\ D_L &= \operatorname{diag}\left(d_1^Y, \ldots, d_K^Y\right) \quad . \end{aligned} \tag{7}$$

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

2.2. Semi-supervised learning on BG

Given the BG constructed from a multi-label image data set, the image annotation problem is transformed to measure how relevant a class vertex is to the unlabeled image vertices. We consider the random walk with restart (RWR) model that performs a random walk process as following:

$$\mathbf{p}^{(t+1)}(j) = (1 - \alpha) \sum_i \mathbf{p}^{(t)}(i) M(i, j) + \alpha \mathbf{e}_j, \tag{8}$$

where $0 \leq \alpha \leq 1$ is a fixed parameter, and $\mathbf{e}_j$ is a vector with all its entries to be 0 except that the j-th one to be 1. At the equilibrium state, the stationary distribution $\mathbf{p}^*$ of this random walk process measures the relevance between the j-th vertex and other vertices. Because $\mathcal{G}$ contains both images and classes as its vertex, when selecting j to be one class vertex, the stationary distribution of the image vertices give the how closely they are related to the j-th class, by which semi-supervise classification can be conducted.

Although the above straightforward method may be feasible, it does not fully make use of the available information. Thus we improve it as following.

Because each semantic class has a set of associated training images, which convey the same semantic meaning as the class itself, we consider both a class vertex and its labeled training image vertices as a semantic group:

$$G_k = c_k \cup \{\mathbf{x}_i \mid \mathbf{y}_i(k) = 1\} \quad . \tag{9}$$

As a result, instead of measuring vertex-to-vertex relevance between a class vertex and an unannotated image vertex as in RWR model, we measure the group-to-vertex, or to be more precise class-to-image, relevance between the semantic group and the image. We construct K distribution vectors, one for each semantic group G_k $(1 \le k \le K)$:

$$\mathbf{h}^{(k)} = \begin{bmatrix} \gamma \mathbf{h}_X^{(k)} \\ (1-\gamma)\,\mathbf{h}_L^{(k)} \end{bmatrix} \in \mathbb{R}_+^{n+K}, \tag{10}$$

where $\mathbf{h}_X^{(k)}(i) = 1/\sum_i \mathbf{y}_i(k)$ if $\mathbf{y}_i(k) = 1$ and $\mathbf{h}_X^{(k)}(i) = 0$ otherwise; $\mathbf{h}_L^{(k)}(i) = 1$ if $i = k$ and $\mathbf{h}_L^{(k)}(i) = 0$ otherwise; $\gamma \in [0, 1]$ controls to what extent the random walker prefers to going to the data subgraph $\mathcal{G}_X$. It can be verified that $\sum_i \mathbf{h}^{(k)}(i) = 1$, *i.e.*, $\mathbf{h}^{(k)}$ is a probability distribution.

Now we consider the following random walk process

$$\mathbf{p}_k^{(t+1)}(j) = (1-\alpha)\sum_i \mathbf{p}_k^{(t)}(i)\, M(i,j) + \alpha \mathbf{h}^{(k)}(j), \tag{11}$$

which describes a random walk process in which the random walker hops on the graph $\mathcal{G}$ according to the transition matrix M with probability $1-\alpha$, and meanwhile it takes a preference to go to the vertices specified by $\mathbf{h}^{(k)}$ with probability α. The equilibrium distribution of this random walk process $\mathbf{p}_k^*$ is determined by $\mathbf{p}_k^{(\infty)} = (1-\alpha) M^T \mathbf{p}_k^{(\infty)} + \alpha \mathbf{h}^{(k)}$, which leads to:

$$\mathbf{p}_k^* = \alpha \big[I - (1-\alpha) M^T\big]^{-1} \mathbf{h}^{(k)} \quad . \tag{12}$$

According to Perron-Frobenius theorem, the maximum eigenvalue of M is less than $\max_i \sum_j M(i,j) = 1$. Therefore, $I - (1-\alpha) M^T$ is positive definite and invertible.

Let I_K be the identity matrix of size $K \times K$, we write

$$H = \left[\mathbf{h}^{(1)}, \ldots, \mathbf{h}^{(K)}\right] = \begin{bmatrix} \gamma H_X \\ (1-\gamma) I_K \end{bmatrix} \quad . \tag{13}$$

Then we write Eq. (12) in matrix form for all K classes and compute the equilibrium distribution matrix P^* as following:

$$P^* = \alpha \big[I - (1-\alpha) M^T\big]^{-1} H, \tag{14}$$

where $P^* = [\mathbf{p}_1^*, \ldots, \mathbf{p}_K^*] \in \mathbb{R}^{(n+K)\times K}$. Thus $\mathbf{p}_k^*(i)$ $(l+1 \le i \le n)$ measures the relevance between the k-th class and an unannotated image $\mathbf{x}_i$, from which we can predict labels for $\mathbf{x}_i$ using the adaptive decision boundary method proposed in our previous work [14].

3. Beyond symmetric label correlations — causal relationships between classes

Label correlations play an important role in MLC to improve the overall classification performance [3, 7, 9, 14, 17]. All existing methods use symmetric label correlations. Our contribution is first to recognize that the true relationships between semantic classes could be **asymmetric** and provide a framework to learn these asymmetric relationships.

Symmetric label correlations used in existing works are often formulated as a symmetric correlation matrix $C \in \mathbb{R}^{K\times K}$ using a variety of measurements, including label co-occurrence [3, 14, 17], normalized mutual information between pairwise classes [7], the Pearson product moment correlation coefficient for label variables [9], *etc*. For example, label co-occurrence based correlations assess how closely two classes are related using cosine similarity as [3]:

$$C(k,l) = \cos\big(\mathbf{y}_{(k)}, \mathbf{y}_{(l)}\big) = \frac{\langle \mathbf{y}_{(k)}, \mathbf{y}_{(l)} \rangle}{\|\mathbf{y}_{(k)}\|\, \|\mathbf{y}_{(l)}\|}, \tag{15}$$

where $\mathbf{y}_{(k)}$ is the k-th row of Y, thus $\langle \mathbf{y}_{(k)}, \mathbf{y}_{(l)} \rangle$ counts the common images annotated to both the k-th and l-th classes.

Our framework starts with a symmetric correlation matrix, such as the one defined in Eq. (15), and gradually learns an asymmetric causal relationship matrix, which has a more realistic semantic meaning.

By a careful look at the equilibrium distribution matrix P^* in Eq. (14), we can write it in a block form as following:

$$P^* = \begin{bmatrix} P_X^* \\ P_L^* \end{bmatrix}, \quad \text{where } P_X^* \in \mathbb{R}^{n\times K}, P_L^* \in \mathbb{R}^{K\times K} \quad . \tag{16}$$

P_L^* is asymmetric, and its entry $P_L^*(i,k)$ assesses the class-to-class relevance from the k-th semantic group G_k to the i-th class c_i. This is same as $P_X^*(i,k)$ that assesses the class-to-image relevance from the G_k to the i-th image $\mathbf{x}_i$, because we consider both images and semantic classes equally as vertices on the BG. We call the learned P_L^* as *Causal Relationships (CR)* .

Given a symmetric label correlation matrix computed from Eq. (15), the learned asymmetric (from columns to rows) CR matrix of MSRC data are listed in Table 1. (The details to obtain Table 1 will be described later in Section 5.1.) We can see that the relationship from the object class "aeroplane" to the background class "sky" is 0.393 while that from "sky" to "aeroplane" is 0.108, *i.e*., the former is greater than the latter. Thus, the learned CR matrix P_L^* is able to better reflects the true semantic relationships.

4. An iterative semi-supervised learning approach

As in Eq. (6), the transition matrix M of a BG is constructed from the data pairwise similarity W_X, label assignments Y and label correlations W_L. Because our BG

Table 1. Asymmetric (column → row) causal relationships of MSRC data set learned by the proposed BG approach. The relationship from a background class such as "sky" to an object class such as "aeroplane" is higher the relationship of the reverse.

	building	grass	tree	cow	horse	sheep	sky	mountain	aeroplane	water	face	car	bycycle	flower	sign	bird	book	chair	road	cat	dog	body	boat
building	-	0.271	0.299	0.280	0.303	0.269	**0.319**	0.297	0.306	0.285	0.278	0.349	0.319	0.287	0.347	0.266	0.263	0.273	0.326	0.269	0.288	0.278	0.293
grass	0.434	-	**0.475**	**0.598**	**0.549**	**0.616**	0.471	0.452	0.501	0.443	0.438	0.442	0.444	**0.472**	0.419	**0.421**	0.423	0.458	0.426	0.424	0.455	0.438	0.432
tree	0.269	**0.275**	-	0.265	0.323	0.251	**0.299**	0.267	0.276	0.257	0.248	0.267	0.261	0.288	0.246	0.240	0.240	0.250	0.260	0.244	0.258	0.249	0.264
cow	0.139	**0.174**	0.141	-	0.132	0.126	0.139	0.155	0.134	0.138	0.137	0.138	0.126	0.131	0.129	0.125	0.124	0.124	0.138	0.124	0.123	0.139	0.129
horse	0.173	**0.171**	0.172	0.163	-	0.172	0.172	0.162	0.166	0.153	0.146	0.154	0.173	0.176	0.179	0.175	0.176	0.179	0.170	0.180	0.178	0.151	0.155
sheep	0.146	**0.175**	0.147	0.134	0.145	-	0.145	0.129	0.135	0.137	0.140	0.138	0.136	0.142	0.140	0.140	0.142	0.140	0.144	0.141	0.138	0.142	0.129
sky	**0.351**	0.333	**0.364**	0.315	0.296	0.302	-	0.356	**0.393**	0.325	0.305	0.323	0.310	0.318	0.378	**0.291**	0.290	0.298	0.316	0.293	0.307	0.307	0.332
mountain	0.128	0.126	0.126	0.135	0.115	0.107	0.124	-	0.120	0.174	0.125	0.128	0.110	0.117	0.119	0.125	0.111	0.109	0.129	0.110	0.110	0.126	0.193
aeroplane	0.096	0.097	0.097	0.088	0.084	0.080	**0.108**	0.090	-	0.100	0.100	0.099	0.083	0.088	0.091	0.088	0.084	0.080	0.096	0.081	0.079	0.102	0.092
water	0.196	0.194	0.196	0.197	0.178	0.183	0.201	0.278	0.200	-	0.193	0.203	0.189	0.181	0.182	0.380	0.175	0.179	0.196	0.176	0.191	0.195	**0.350**
face	0.164	0.161	0.163	0.161	0.145	0.153	0.162	0.161	0.169	0.165	-	0.169	0.157	0.234	0.161	0.146	0.215	0.148	0.162	0.145	0.154	**0.364**	0.162
car	0.125	0.106	0.108	0.103	0.092	0.095	0.107	0.103	0.109	0.111	0.108	-	0.102	0.094	0.097	0.093	0.088	0.091	**0.144**	0.087	0.095	0.109	0.106
bycycle	0.118	0.116	0.118	0.103	0.111	0.104	0.116	0.101	0.105	0.111	0.110	0.110	-	0.111	0.110	0.110	0.107	0.107	**0.137**	0.108	0.106	0.112	0.103
flower	0.166	**0.165**	0.165	0.155	0.173	0.164	0.165	0.154	0.158	0.147	0.143	0.145	0.164	-	0.170	0.168	0.167	0.170	0.163	0.172	0.170	0.139	0.147
sign	0.182	0.183	0.183	0.174	0.192	0.184	0.182	0.172	0.178	0.161	0.152	0.164	0.183	0.186	-	0.186	0.186	0.190	**0.179**	0.191	0.190	0.155	0.165
bird	0.182	**0.182**	0.183	0.173	0.190	0.182	**0.182**	0.172	0.176	0.175	0.154	0.164	0.182	0.185	0.187	-	0.184	0.188	0.179	0.189	0.187	0.160	0.166
book	0.160	0.160	0.160	0.149	0.164	0.158	0.159	0.147	0.151	0.144	0.158	0.141	0.156	0.161	0.161	0.159	-	0.161	0.157	0.163	0.161	0.153	0.142
chair	0.153	0.154	0.154	0.142	0.156	0.149	0.152	0.139	0.141	0.143	0.143	0.140	0.147	0.152	0.150	0.152	0.152	-	0.156	0.153	0.150	0.145	0.137
road	0.348	0.288	0.312	0.297	0.284	0.291	0.309	0.300	0.316	0.303	0.296	**0.425**	**0.413**	0.306	**0.360**	0.287	0.282	0.340	-	0.342	0.347	0.297	0.303
cat	0.162	0.162	0.163	0.152	0.167	0.160	0.160	0.149	0.152	0.150	0.148	0.144	0.158	0.162	0.160	0.162	0.163	0.163	0.163	-	0.162	0.151	0.146
dog	0.130	0.128	0.128	0.114	0.126	0.119	0.128	0.114	0.113	0.121	0.116	0.118	0.119	0.124	0.125	0.124	0.122	0.122	0.130	0.123	-	0.118	0.120
body	0.176	0.173	0.175	0.175	0.159	0.167	0.174	0.176	0.185	0.178	**0.393**	0.183	0.173	0.244	0.236	0.161	0.224	0.164	0.174	0.160	0.170	-	0.175
boat	0.140	0.140	0.138	0.134	0.126	0.122	0.139	0.217	0.137	**0.247**	0.139	0.144	0.126	0.127	0.129	0.148	0.121	0.119	0.142	0.120	0.131	0.140	-

approach is able to learn a better CR matrix P_L^* from the label correlation matrix C directly estimated from input image data, instead of only using the input label correlations by setting $W_L = C$, we may replace it by the learned CR matrix by setting $W_L = P_L^*$. Repeat this process, we predict labels for unannotated images using the Iterative BG approach listed in Algorithm 1.

Algorithm 1: Iterative BG approach.

Data: 1. Image pairwise similarity: W_X,
2. Label assignment matrix: Y,
3. A pre-specified maximum iteration number: $\max\_\text{iter}$.
Result: Labels $\mathcal{L}_i$ assigned to $\mathbf{x}_i$ $(l+1 \leq i \leq n)$.
1. Compute the symmetric label correlation matrix C by Eq. (15) and set $W_L = C$;
2. Set $\text{iter} = 1$;
repeat
  1. Construct M by Eq. (6) and H by Eq. (13);
  2. Compute P^* by Eq. (14) and Eq. (16);
  3. Set $W_L = P_L^*$;
  4. $\text{iter} = \text{iter} + 1$;
until $\text{iter} > \max\_\text{iter}$
3. Predict labels for $\mathbf{x}_i$ using P^* by adaptive decision boundary method [14].

Note that, most, if not all, existing MLC methods [3, 14, 17] requires the correlation matrix to be symmetric, which is not necessary for our approach. In order to ensure Eq. (14) is solvable, M thereby W_L is only required to be full rank, which is automatically satisfied by construction.

5. Experimental results

We experimentally evaluate the proposed approaches in automatic image annotation task and image retrieval task using following four benchmark image data sets.

TRECVID 2005 data set[1] contains 61901 sub-shots labeled with 39 concepts. We randomly sample the data such that each concept (label) has at least 100 video key frames.

MSRC[2] data set is provided by the computer vision group at Microsoft Research Cambridge, which contains 591 images annotated by 22 classes.

PASCAL VOC 2010 data set[3] has 13321 images with 20 classes. We randomly sample at least 200 images for each class, and obtain 3679 images for experiments.

Following [14, 17, 19], we extract 384-dimensional block-wise (over 8×8 fixed grid) color moments (mean and variable of each color band) in Lab color space as features for the above three data sets.

Natural scene data set [1] contains 2407 images represented by a 294-dimensional vector, which are labeled with 6 semantic concepts (labels).

Parameter settings of our approach. The parameter α in Eq. (11) indicates how much the random walk process is biased by the pre-specified distribution, which acts similar to the restart probability of RWR method [8] and the damping factor of PageRank algorithm [2]. Following [2, 8], we fix it as a small constant to be 0.01 in all our experiments.

We set both the cross-graph probability parameter β in Eq. (6) and the preferential probability parameter γ in Eq. (13) as 0.5, because we consider both image subgraph and label subgraph are equally important.

Other implementation details. The proposed BG approach requires both image and label similarity matrices as input. We compute image similarity using the Gaussian kernel function as $W_X(i,j) = \exp\left(\|\mathbf{x}_i - \mathbf{x}_j\|^2/\sigma^2\right)$ if $i \neq j$ and $W_X(i,j) = 0$ otherwise, where we empirically set

[1] http://www-nlpir.nist.gov/projects/trecvid/
[2] http://research.microsoft.com/en-us/projects/objectclassrecognition
[3] http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2010/

$\sigma = \sum_{i \neq j} \|\mathbf{x}_i - \mathbf{x}_j\| / [n\,(n-1)]$. We use co-occurrence based label similarity defined in Eq. (15) as input and set $W_L = C$ as initialization, which is symmetric.

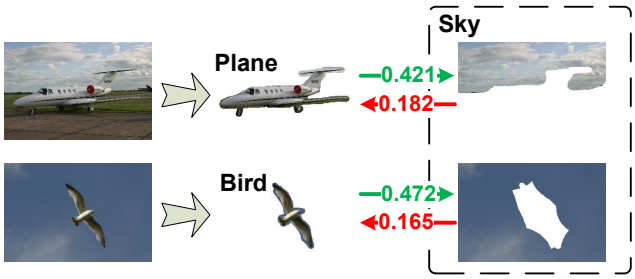
Following [17], we also initialize the labels of unannotated images by k-Nearest Neighbor (kNN) method where we set $k = 1$. Although the initializations are not completely correct, a big portion of them are (assumed to be) correctly predicted. Our approach will self-consistently amend the incorrect labels.

5.1. Evaluate the learned asymmetric causal relationships between classes

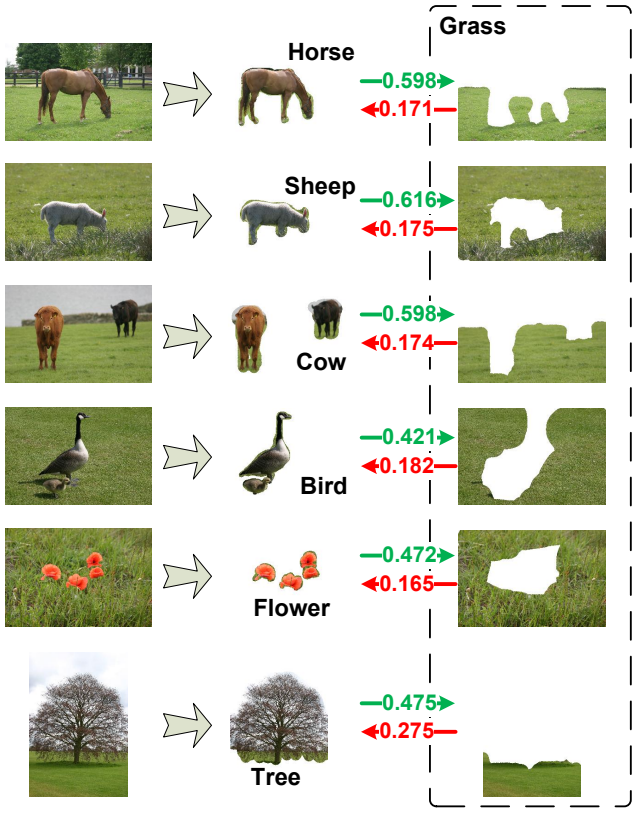
Because learning and using asymmetric CR matrix are one of important advantage of our approach, we first evaluate them on MSRC data. Given a symmetric label correlation matrix estimated from input data using Eq. (15), our approach learns a new CR matrix, which is shown in Table 1. We can see that the correlations from object classes to their corresponding background classes are mostly greater than those of the reverse. For example, the relationships of "car/bycycle"→"road" are greater than those of "road"→"car/bycycle" as in Figure 2, the relationships of "bird/aeroplane"→"sky" are greater than those of "sky"→"bird/aeroplane" as in Figure 3(a), the relationships of "horse/sheep/cow/bird/flower/tree"→"grass" are greater than those of "grass"→"horse/sheep/cow/bird/flower/tree" as in Figure 3(b), *etc*. All these observations are consistent with the real semantic meanings, which firmly support the correctness of the asymmetric causal relationships learned by our approach. In conclusion, asymmetric label correlation matrix has more freedom than its symmetric counterpart, therefore it is more flexible and able to better characterize the semantic relationships in real image data.

5.2. Results on automatic image annotation

We then evaluate the proposed approaches in automatic image annotation task. We perform standard 5-fold cross-validation and compare the classification performance averaged over the five trials of our approach against the following state-of-the-art MLC methods: (1) Multi-label informed Latent Semantic Indexing (MLSI) method [18], (2) Multi-Label Correlated Green's function (MLGF) method [14], (3) Random k-Labelsets (REKEL) method [10], (4) Multi-Label Least Square (MLLS) method [4]. We implement the first three methods following their original works. For MLGF method, we set $\alpha = 0.1$ as in [14]; for MLSI method, we set $\beta = 0.5$ as in [18], and kNN ($k = 1$) is used for classification after dimension reduction. For MLLS method, we use the codes posted by the authors. For our approaches, we report the results for (1) BG without iteration which uses symmetric label correlations, (2) iterative BG in Algorithm 1 which uses the learned asymmetric label correlations. The latter is denoted as I-BG(Iter) in Table 2 where Iter indicates the number of iterations.



(a) Object classes: "aeroplane/bird", background class: "sky".



(b) Object classes: "horse/sheep/cow/bird/flower/tree", background classes: "grass".

Figure 3. Asymmetric causal relationships between several classes of MSRC data learned by the proposed BG approach. The numbers are asymmetric causal relationships between the classes.

The conventional classification performance metrics in statistical learning, *precision* and *F1 score*, are used to evaluate the proposed algorithms. For every class, the precision and F1 score are computed following the standard definition for a binary classification problem. To address the multi-label scenario, following [10], macro average and micro average of precision and F1 score are computed to assess the overall performance across multiple labels.

Table 2 presents the classification performance compar-

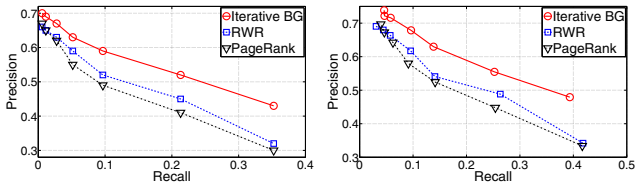
isons in 5-fold cross-validation, which show that the proposed BG approach and its iterative counterparts consistently outperform other compared methods, sometimes very significantly. In average, our approaches achieve more than 10% improvement compared to the best performance of compared methods, which demonstrate the effectiveness of our approaches in multi-label image classification. In addition, we also see that the iterative BG approach using the learned asymmetric label correlations is always superior to the BG approach using symmetric label correlations. More iterations lead to better performance. This provides one more concrete evidence of the usefulness of the learned asymmetric label correlations.

Some example annotation results on TRECVID 2005 data by our iterative BG approach are listed in Table 3, where all the labels are correctly predicted by our approach. Because REKEL method has shown best classification performance among the four other compared methods as in Table 2, we also list its annotation results, which, however, can only predict parts of the labels. Note that, the label "tree" predicted for the leftmost image by our approach is not in ground truth, which, nevertheless, can be clearly seen in the left part of the image. The similar is for the label "sky" in the second leftmost image.

5.3. Results on semantic retrieval

Because image annotation task ignores the ranking order of the resulted relevance scores, in this subsection we address this by evaluating the proposed approach in semantic retrieval task, in which, given a semantic keyword, a list of relevant images are expected to be returned. In our evaluations, we randomly split a data set into two parts with equal size, one for training and the other for testing. Our goal is to retrieve the relevant images in the testing set in response to a query semantic class. We compare our approach to two closely related methods: (1) Random Walk with Restart (RWR) method [8] and (2) PageRank [2] method.

As our approach measures class-to-image relevance, we can directly return the images with highest relevance scores with respect to a query class using Eq. (14). Because both RWR method and PageRank method measure image-to-image relevance, we apply the following retrieval strategy. Given a query keyword, we compute the image-to-image relevance scores between a test image and all the training images associated with the keyword. The maximum image-to-image relevance score is assigned as the image-to-class relevance score between the test image and the class. Repeat this process for all the test images and we retrieve images according to the resulted image-to-class relevance scores. Both the restart probability of RWR method and the damping factor of PageRank method are set as 0.01, which is same as α in our approaches. Note that, these two methods perform random walks on the data subgraph $\mathcal{G}_X$ while our approaches performs random walks on the BG of $\mathcal{G}$.



(a) TRECVID 2005 data. (b) PASCAL VOC 2010 data.

Figure 4. Image retrieval performance measured by precision-recall curves of the three compared methods.

Figure 4 shows the retrieval performance of the three compared methods measured by precision-recall curves. Given the computed image-to-class relevance scores, we retrieve the top 10, 20, 50, 100, 200, 500 and 1000 images for every class, upon which precision and recall are computed following the standard definitions. The averaged precisions and recalls over all the classes are plotted in Figure 4, which shows that the proposed iterative BG approach (with 3 iterations) is clearly better than the other two methods, especially when the number of retrieved images is big.

6. Conclusions

We proposed a novel Bi-relational Graph (BG) model to place both data graph and label graph of a multi-label image data set in a unified framework, upon which we considered a class and its training images as a semantic group and performed random walk to produce both class-to-image relevances and class-to-class relevances. Different from image-to-image relevance obtained by existing methods, the class-to-image relevance from our approach can be used to directly predict labels for unannotated images. The learned class-to-class relevances, called as causal relationships, describe the class relationships in an asymmetric way, which are more close to real semantic relationships. By applying the learned asymmetric yet better CR matrix in our approach, the annotation performance is improved. We applied the proposed approaches in automatic image annotation and semantic image retrieval tasks. Encouraging results in extensive experiments demonstrated their effectiveness.

Acknowledgments. This research was supported by NSF-CCF 0830780, NSF-CCF 0917274, NSF-DMS 0915228, NSF-CNS 0923494, NSF-IIS 1041637.

References

[1] M. Boutell, J. Luo, X. Shen, and C. Brown. Learning multi-label scene classification. *Pattern Recognition*, 37(9):1757–1771, 2004.

[2] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *WWW*, 1998.

Table 2. Classification performance comparison by 5-fold cross validations on the four multi-label image data sets.

Data	Metrics		MLSI	MLGF	REKEL	MLLS	BG	I-BG(1)	I-BG(2)	I-BG(3)
TRECVID 2005	Macro average	Precision	0.247	0.248	0.251	0.249	0.256	0.267	0.268	**0.269**
		F1	0.275	0.276	0.280	0.277	0.283	0.296	0.302	**0.303**
	Micro average	Precision	0.234	0.235	0.239	0.241	0.253	0.268	0.272	**0.274**
		F1	0.293	0.292	0.299	0.295	0.302	0.318	0.321	**0.324**
MSRC	Macro average	Precision	0.252	0.216	0.259	0.255	0.266	0.279	0.283	**0.286**
		F1	0.287	0.279	0.301	0.290	0.308	0.323	0.329	**0.331**
	Micro average	Precision	0.253	0.237	0.258	0.255	0.263	0.286	0.291	**0.293**
		F1	0.301	0.287	0.304	0.302	0.302	0.322	0.329	**0.332**
PASCAL VOC 2010	Macro average	Precision	0.357	0.348	0.362	0.359	0.371	0.386	0.394	**0.395**
		F1	0.401	0.395	0.413	0.405	0.420	0.431	0.438	**0.438**
	Micro average	Precision	0.403	0.392	0.409	0.406	0.415	0.431	0.435	**0.436**
		F1	0.415	0.411	0.422	0.421	0.429	0.430	0.437	**0.437**
Natural scene	Macro average	Precision	0.368	0.362	0.421	0.418	0.456	0.493	0.507	**0.511**
		F1	0.411	0.408	0.439	0.434	0.466	0.504	0.512	**0.515**
	Micro average	Precision	0.412	0.406	0.433	0.426	0.473	0.517	0.526	**0.531**
		F1	0.429	0.417	0.451	0.443	0.481	0.520	0.527	**0.530**

Table 3. Annotation results of several images in TRECVID 2005 data set by the proposed iterative BG approach and REKEL method. Our approach can predict all the labels for the images, while REKEL method can only predict part of the labels. The labels predicted by our approach but not in ground truth are in italic bold font, which, however, can be clearly seen in the images.

						
Iterative BG	building, car, outdoor, road, ***tree***	building, outdoor, waterscape, ***sky***	face, meeting, person, studio	building, outdoor, urban	military, outdoor, person, sky	building, outdoor, person
REKEL	building, outdoor	outdoor, waterscape	face, person, studio	building, outdoor	outdoor, person, sky	outdoor, person

[3] G. Chen, Y. Song, F. Wang, and C. Zhang. Semi-supervised multi-label learning by solving a sylvester equation. In *SDM*, 2008.

[4] S. Ji, L. Tang, S. Yu, and J. Ye. A shared-subspace learning framework for multi-label classification. *TKDD*, 4(2):1–29, 2010.

[5] F. Kang, R. Jin, and R. Sukthankar. Correlated label propagation with application to multi-label learning. In *CVPR*, 2006.

[6] Y. Liu, R. Jin, and L. Yang. Semi-supervised multi-label learning by constrained non-negative matrix factorization. In *AAAI*, 2006.

[7] G. Qi, X. Hua, Y. Rui, J. Tang, T. Mei, and H. Zhang. Correlative multi-label video annotation. In *ACM MM*, 2007.

[8] H. Tong, C. Faloutsos, and J. Pan. Fast random walk with restart and its applications. In *ICDM*, 2006.

[9] G. Tsoumakas, A. Dimou, E. Spyromitros, V. Mezaris, I. Kompatsiaris, and I. Vlahavas. Correlation-based pruning of stacked binary relevance models for multi-label learning. In *MLD Workshop of ECML PKDD*, 2009.

[10] G. Tsoumakas, I. Katakis, and I. Vlahavas. Random k-Labelsets for Multi-Label Classification. *TKDE*, 2010.

[11] H. Wang, C. Ding, and H. Huang. Directed graph learning via high-order co-linkage analysis. In *ECML/PKDD*, 2010.

[12] H. Wang, C. Ding, and H. Huang. Multi-Label Classification: Inconsistency and Class Balanced K-Nearest Neighbor. In *AAAI*, 2010.

[13] H. Wang, C. Ding, and H. Huang. Multi-label linear discriminant analysis. *ECCV*, pages 126–139, 2010.

[14] H. Wang, H. Huang, and C. Ding. Image annotation using multi-label correlated Green's function. In *ICCV*, 2009.

[15] H. Wang, H. Huang, and C. Ding. Discriminant Laplacian Embedding. In *AAAI*, 2010.

[16] H. Wang, H. Huang, and C. Ding. Image categorization using directed graphs. In *ECCV*, 2010.

[17] H. Wang, H. Huang, and C. Ding. Multi-label Feature Transform for Image Classifications. In *ECCV*, 2010.

[18] K. Yu, S. Yu, and V. Tresp. Multi-label informed latent semantic indexing. In *SIGIR*, 2005.

[19] Z. Zha, T. Mei, J. Wang, Z. Wang, and X. Hua. Graph-based semi-supervised learning with multi-label. In *ICME*, 2008.

[20] S. Zhu, X. Ji, W. Xu, and Y. Gong. Multi-labelled classification using maximum entropy method. In *SIGIR*, 2005.