

# Image Annotation Using Multi-label Correlated Green’s Function

Hua Wang, Heng Huang, Chris Ding  
Computer Science and Engineering Department  
University of Texas at Arlington

huawang2007@mavs.uta.edu {heng, chqding}@uta.edu

## Abstract

*Image annotation has been an active research topic in the recent years due to its potentially large impact on both image understanding and web/database image search. In this paper, we target at solving the automatic image annotation problem in a novel semi-supervised learning framework. A novel multi-label correlated Green’s function approach is proposed to annotate images over a graph. The correlations among labels are integrated into the objective function which improves the performance significantly. We also propose a new adaptive decision boundary method for multi-label assignment to deal with the difficulty of label assignment in most of the existing rank-based multi-label classification algorithms. Instead of setting the threshold heuristically or by experience, our method principally compute it upon the prior knowledge in the training data. We perform our methods on three commonly used image annotation testing data sets. Experimental results show significant improvements on classification performance over four other state-of-the-art methods. As a general semi-supervised learning framework, other local feature based image annotation methods could be easily incorporated into our framework to improve the performance.*

## 1. Introduction

Image retrieval plays an important role in information retrieval due to the overwhelming image and video data brought by modern technologies. One of notorious bottleneck in the image retrieval is how to associate an image or video with some semantic keywords to describe its semantic content. This poses a challenging computer vision topic, image annotation, which has attracted broad attentions in the recent years. However, manual image annotation is an expensive and labor intensive procedure and hence there has been great interest in coming up with automatic ways to retrieve images based on content.

Given a set of annotated images as training data, many methods have been proposed in the literature to find most

representative keywords to annotate new images [4, 14]. However, in most cases, the unlabeled data are insufficient. Compared to the large size of an image or video data set, the annotated images have a relative small number. The semi-supervised learning techniques leverage the unlabeled data in addition to labeled data to tackle this difficulty [10, 12].

In recent image annotation research, several researchers focus on multi-label problems [10, 20], because an image or video is usually associated with more than one concepts. For example, a natural image often includes “mountain”, “water”, and “sky” at the same time. As a result, it could be annotated with multiple labels. Thus, the inherent correlations among multiple labels should be also integrated into semi-supervised learning methods. Several researchers have proven that the performance of image annotation is improved by utilizing these correlations between labels.

In this paper, we propose a novel Multi-label Correlated Green’s Function (MCGF) approach to annotate images over graphs by employing the labels correlations. A new adaptive decision boundary method for multi-label assignment is also proposed and integrated into MCGF algorithm to solve the label assignment ambiguity problem in semi-supervised learning. Our new semi-supervised method can easily combine other image feature extraction and distance metric learning methods to improve the image annotation performance. We perform our new MCGF method to annotate image/video on three commonly used data sets TRECVID [3], Microsoft Research Cambridge data set [2], and Barcelona data set [1]. Over all three data sets, our new method outperforms four other state-of-the-art methods.

### 1.1. Related work

The most common solution to multi-label is to decompose the problem into multiple, independent binary classification problems and determine the labels for each data point by aggregating the classification results from all the classifiers [5]. However these approaches suffer from the problems of limited number of categories and unbalanced training data. Another group of approaches toward multi-label is label ranking [7, 16]. Though these approach deal

with the some of the disadvantages of the binary classification approach, they normally do not principally give out the thresholds for multi-label assignment which is gracefully resolved by our proposed method in section 3.

Recently more attention has been received for multi-label learning that consider the correlation among categories. Ueda *et al.* [18] suggests a generative model which incorporates the pairwise correlation between any two categories into multi-label learning. [9] introduces a Bayesian model to assign labels through underlying latent representations. Zhu *et al.* [23] employs a maximum entropy method for multi-labeled learning to model the correlations among categories. McCallum [13] and Yu *et al.* [19] apply approaches based on latent variables to capture the correlation among different categories. Despite these effort in exploiting label correlation of class labels, most of the research is limited to pairwise correlation of class labels.

Unlike most of previous works on multi-label classification problem, our method is built upon a graph based algorithm. Given a data set with pairwise similarities ( $W$ ), the semi-supervised learning can be viewed as label propagation from labeled data to unlabeled data. In its simplest form, the label propagation is like a random walk on a similarity-graph  $W$  [17]. Using the diffusion kernel, the semi-supervised learning is like a diffusive process of the labeled information [11]. The harmonic function approach [24] emphasizes the harmonic nature of the diffusive function; consistency labeling approach [22] emphasizes the spread of label information in an iterative way; and Green's function approach [8] focuses on the label information propagation. Our work is inspired by these prior works, especially by the work of Ding *et al.* [8].

## 1.2. Our contributions

We summarize our contribution as follows:

1) We propose a novel multi-label correlated Green's function approach which successfully incorporates label correlations to enhance the classification performance.

2) We propose an adaptive decision boundary method to deal with the difficulty of label assignment in most of the existing rank-based multi-label classification problem. Instead of setting the threshold heuristically or by experience, our method principally compute it upon the prior knowledge in the training data. Moreover, it also compensates the unbalanced distribution in the training data so as to improve the classification performance.

## 2. Class correlated reproducing kernel Hilbert space approach for multi-label semi-supervised classification

For a multi-class problem where we have  $K$  classes  $\{S_1, \dots, S_K\}$ , given input  $\{(\mathbf{x}_1, \mathbf{z}_1), \dots, (\mathbf{x}_n, \mathbf{z}_n)\}$  where

$\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_l, \mathbf{x}_{l+1}, \dots, \mathbf{x}_n\} \subset \mathbb{R}^p$  is a point set, the first  $l$  points are labeled as  $\{\mathbf{z}_1, \dots, \mathbf{z}_l\}$  where  $\mathbf{z}_i$  is a label indicator vector of size  $K$  containing the labels assigned to the point  $\mathbf{x}_i$ . Our goal is to predict the label sets  $\{\mathbf{z}_{l+1}, \dots, \mathbf{z}_n\}$  for the remaining unlabeled data points  $\{\mathbf{x}_{l+1}, \dots, \mathbf{x}_n\}$ . For  $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_n]^T$ , we have

$$\mathbf{Z}_{ik} = \begin{cases} 1, & \mathbf{x}_i \in S_k \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Consider a connected graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  with nodes set  $\mathcal{V} = \mathcal{L} \cup \mathcal{U}$  where  $\mathcal{L}$  corresponds to  $\{\mathbf{x}_1, \dots, \mathbf{x}_l\}$  and  $\mathcal{U}$  corresponds to  $\{\mathbf{x}_{l+1}, \dots, \mathbf{x}_n\}$ . The edge  $\mathcal{E}$  are weighted by the  $n \times n$  affinity matrix  $W$  with  $W_{ij}$  indicating the similarity measure between  $\mathbf{x}_i$  and  $\mathbf{x}_j$ . We define a  $K \times K$  matrix,  $C$ , to capture the label correlation information. Let  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_K] = \mathbf{Z}$ , by using cosine similarity,

$$C_{ij} = \cos(\mathbf{y}_i, \mathbf{y}_j) = \frac{\langle \mathbf{y}_i, \mathbf{y}_j \rangle}{\|\mathbf{y}_i\| \|\mathbf{y}_j\|} \quad (2)$$

According to our experimental results and previous research in the area of computer vision, we empirically choose cosine similarity due to its clear theoretical meaning and experimental performance.

Similarly, we define  $\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_K]$  to be the predicted decision values, which is a real-valued function  $f : \mathcal{V} \rightarrow \mathbb{R}$  on  $\mathcal{G}$ .

## 2.1. Background

Given a mesh/graph with edges weights  $W$ , the *combinatorial Laplacian* is defined to be

$$L = D - W \quad (3)$$

where the diagonal matrix contains row sum of  $W$ :  $D = \text{diag}(W\mathbf{e})$ ,  $\mathbf{e} = [1 \dots 1]^T$ . The Green's function for a generic graph is defined as the inverse of  $L = D - W$  with zero-mode discarded. We construct the Green's function using the eigenvectors of  $L$ :

$$L\mathbf{v}_k = \lambda_k \mathbf{v}_k, \mathbf{v}_p^T \mathbf{v}_q = \delta_{pq} \quad (4)$$

where  $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$  are the eigenvalues. Assuming the graph is connected, otherwise we deal with each connected component one at a time. The first eigenvector is a constant vector  $\mathbf{v}_1 = \mathbf{e}/n^{1/2}$  with zero eigenvalue and multiplicity one. Discarding this zero-mode, the Green's function is then the positive definite part of  $L$ :

$$G = L_+^{-1} = \frac{1}{(D - W)_+} = \sum_{i=2}^n \frac{\mathbf{v}_i \mathbf{v}_i^T}{\lambda_i} \quad (5)$$

where  $(D - W)_+$  indicates zero eigen-mode is discarded.

## 2.2. Multi-label classification framework by Green's function

Green's function approach builds up an effective framework for semi-supervised learning to address binary class and single-label multi-class problem, but still does not solve the multi-label multi-class problem. The most straightforward way is to transform the multi-label learning problem into a set of binary classification problem, where one vs. all or one vs. one are the two most popular ways, and the former is formulated as:

$$\mathbf{F} = G\mathbf{Y} \quad (6)$$

We name Eq. (6) as simple multi-label Green's function approach (MLGF), upon which we propose a novel multi-label correlated Green's function approach from the theory of Reproducing Kernel Hilbert Space (RKHS).

Despite its simplicity, MLGF approach treats the labels in isolation and ignores the correlations of among them. In multi-label scenarios, however, the labels usually interact with each other naturally. For example, "face" and "person" tend to appear simultaneously, while "seascape" typically does not appear with "indoor". To exploit these interactions among labels, we introduce a  $K \times K$  symmetric matrix,  $C$ , as in Eq.(2) to capture the label correlations, where  $C_{ij}$  represents the correlation between label  $i$  and label  $j$ . Therefore from the theory of RKHS, we can add an penalty term constructed by the label correlation matrix,  $\text{tr}[-\alpha(D-W)^{\frac{1}{2}}\mathbf{F}^T C \mathbf{F}(D-W)^{\frac{1}{2}}]$ , to impose smoothness to the loss function  $\text{tr}[\beta \|\mathbf{F} - \mathbf{Y}\|]$ [8], then we can derive the objective function to minimize as below:

$$J[\mathbf{F}] = \text{tr}[\beta|\mathbf{F} - \mathbf{Y}|^2 + \mathbf{F}^T \mathcal{K}^{-1} \mathbf{F} - \alpha \mathcal{K}^{-\frac{1}{2}} \mathbf{F}^T C \mathbf{F} \mathcal{K}^{-\frac{1}{2}}] \quad (7)$$

where  $\mathcal{K} = G = (D - W)^{-1}$  is the kernel,  $\alpha$  and  $\beta$  are two nonnegative small constant that balance the two regularizers. Differentiating  $J[\mathbf{F}]$  with respect to  $\mathbf{F}$ , we have

$$\frac{\partial J}{\partial \mathbf{F}} = 2\beta(\mathbf{F} - \mathbf{Y}) + 2(D - W)\mathbf{F} - 2\alpha(D - W)\mathbf{F}C = 0 \quad (8)$$

That is

$$\mathbf{F} = \frac{1}{\beta I + (D - W)}\beta\mathbf{Y} + \alpha \frac{1}{\beta I + (D - W)}(D - W)\mathbf{F}C \quad (9)$$

Because  $\beta$  is a small nonnegative constant, we have

$$\frac{1}{\beta}\mathbf{F} = \frac{1}{(D - W)}\mathbf{Y} + \alpha \frac{\mathbf{F}}{\beta}C \quad (10)$$

Let  $\tilde{\mathbf{F}} = \frac{\mathbf{F}}{\beta}$ ,

$$\tilde{\mathbf{F}} = \frac{1}{(D - W)}\mathbf{Y} + \alpha\tilde{\mathbf{F}}C = G\mathbf{Y} + \alpha\tilde{\mathbf{F}}C \quad (11)$$

Therefore we can build an iterative multi-label classifier according to the subspace iteration algorithm as follows

$$\begin{cases} \tilde{\mathbf{F}}^{(0)} = G\mathbf{Y} \\ \tilde{\mathbf{F}}^{(t+1)} = G\mathbf{Y} + \alpha\tilde{\mathbf{F}}^{(t)}C \end{cases} \quad (12)$$

Upon convergence,  $\tilde{\mathbf{F}}^{(\infty)}$  minimize the objective function (7) and hence the optimal solution is obtained.

This algorithm can be understood intuitively in terms of label information propagation on an network ( $\mathcal{G}$ ). In the initialization step, the label information is propagated from the labeled nodes into the entire network, which is equivalent to the MLGF approach as in Eq.(6). After initialization, during each iteration each data point receives the information from its neighbors (second term), and also retain its initial information (first term). The parameter  $\alpha$  specifies the relative amount of information from its neighbors and initial label information. Because the diagonal entries of affinity matrix are set to zero in its building process (15), self-reinforcement is avoided. Moreover, the information is spread symmetrically as  $C$  is a symmetric matrix. Finally, upon convergence the label of each unlabeled point is set to the class of which it has received most information during the iteration process.

It is easy to show when  $0 < \alpha < \min(1, \frac{1}{\max(\zeta_k)})$ , where  $\zeta_k (0 < k < K)$  are the eigenvalues of  $C$ , the sequence  $\{\tilde{\mathbf{F}}^{(t)}\}$  converges as below

$$\tilde{\mathbf{F}} = G\mathbf{Y} (I - \alpha C)^{-1} \quad (13)$$

By Eq. (13) we can computer  $\mathbf{F}$  for classification without iterations. This also shows that the iteration result is independent of the initial value for the iteration. Furthermore,  $(I - \alpha C)^{-1}$  in fact is another graph or diffusion kernel which propagates the label influence through the label correlations over the whole network.

## 3. Adaptive decision boundary

In many semi-supervised learning algorithms, the threshold for classification is usually selected as 0, which is similar to SVM where we use  $f(x) = 0$  as the decision boundary. However,  $f(x) = 0$  is not necessarily the best choice. We may fine tune the decision boundary to achieve better performance. In the semi-supervised setting, this decision boundary is adjusted such that the training errors of all positive and negative samples are minimized and the adaptive decision boundary is the following.

Consider the binary classification problem for the  $k$ th class, let  $b_k$  by the adaptive decision boundary,  $S_+$  and  $S_-$  be the number of positive and negative samples for the  $k$ th class, and  $e_+(b_k)$  and  $e_-(b_k)$  be the numbers of misclassified positive and negative training samples, the optimal

decision boundary is given by the Bayes rule:

$$b_k^{\text{opt}} = \arg \min_{b_k} \left[ \frac{e_+(b_k)}{|S_+|} + \frac{e_-(b_k)}{|S_-|} \right]$$

And the decision rule is given by:

$$x_i \text{ acquires label } k = \begin{cases} +1, & \text{if } \mathbf{F}_{ik} > b_k^{\text{opt}} \\ -1, & \text{if } \mathbf{F}_{ik} < b_k^{\text{opt}} \end{cases} \quad (14)$$

Figure 1 shows the adaptive decision boundary for the “outdoor” class in TRECVID 2005 data set where the areas (probability likelihood) of misclassification are minimized, which is different from 0.

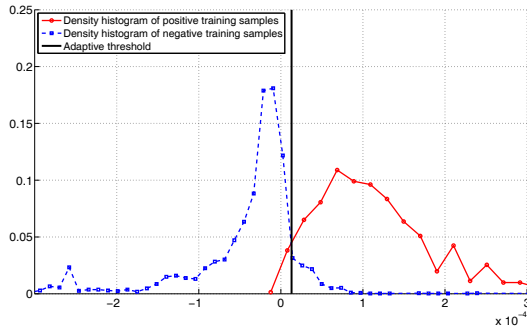


Figure 1: Optimal decision boundary to minimize misclassification for the “outdoor” class in TRECVID 2005 data set indicated by the black vertical line is different from 0.

## 4. Experiments

We apply the simple multi-label Green’s function (MLGF) approach and our proposed multi-label correlated Green’s function with adaptive decision boundary (MCGF) approach over three commonly used image annotation data sets, *i.e.* TREC Video Retrieval Evaluation (TRECVID) 2005 development set [3], Microsoft Research Cambridge (MSRC) data set [2], and Barcelona data set [1]. We compare our methods to four state-of-the-art methods: (a) the consistent framework (CF) approach [22], (b) the harmonic function (HF) approach [24], (c) multi-label harmonic function (MLHF) approach [21] and semi-supervised learning by Sylvester equation (SMSE) [6] approach.

### 4.1. Data sets and experimental setup

**TRECVID 2005** data set contains 137 broadcast videos from 13 different programs in English, Arabic and Chinese, which are segmented into 74523 sub-shot. According to LSCOM-Lite annotations [15], 39 concepts are labeled on each sub-shot, which consist of a wide range of genres including program category, setting/scene/site, people, object,

activity, event, and graphs. Many of these concepts have significant semantic dependence between each other. Many sub-shots (75.67%) have more than one label, and some sub-shots are even labeled with as many as 17 concepts. Figure 2 illustrates some sample images in this data set. In our experiments, the images in this data set are resized to half on both horizontal and vertical side.

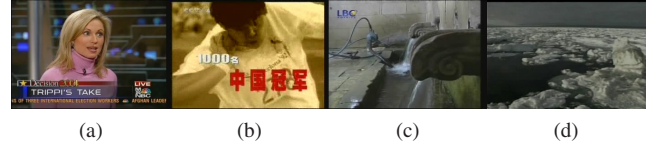


Figure 2: Sample images from TRECVID 2005 data set. (2a) face, person, studio; (2b) outdoor, person, sports; (2c) natural disaster, outdoor; (2d) sky, snow, waterscape.

**MSRC** data set contains 591 images with 23 classes. Around %80 of the images are annotated with at least one classes and around three classes per images on average. MSRC data set provides image annotation at pixel level, where each pixel is labeled as one of 23 classes or “void”. As suggested by MSRC, “horse” and “mountain” are treated as “void” because they have few labeled instances. Therefore, there are 21 classes in total. In our experiments, we use only image level annotation built upon the pixel level.

**Barcelona** data set contains 139 images with 4 categories, *i.e.*, “building”, “flora”, “people” and “sky”. Each image has at least two labels.

For each data set, we conduct 10-fold cross validation. Specifically, the images are randomly split into ten parts with equal size, we selected each of the ten parts as testing set and the rest as training set. The average performance over the ten iterations is reported for evaluation.

### 4.2. Affinity matrix

We use the Gaussian kernel function to calculate the affinity matrix  $W$ . When  $\mathbf{x} \in \mathbb{R}^p$ , the affinity matrix is

$$w_{ij} = \begin{cases} \exp \left( - \sum_{d=1}^p \frac{(x_{id} - x_{jd})^2}{\sigma_d^2} \right) & i \neq j \\ 0 & i = j \end{cases} \quad (15)$$

where  $x_{id}$  is the  $d$ th component of instance  $\mathbf{x}_i \in \mathbb{R}^p$ , and  $\sigma_1, \sigma_2, \dots, \sigma_p$  are length scale hyperparameters for each dimension, which are set equal to reduce parameter tuning according to most of the previous research. Thus nearby point in Euclidean space are assigned large edge weight.

### 4.3. Evaluation metrics

For performance evaluation, we adopt the widely-used performance metric, Average Precision (AP), as suggested

by TRECVID [3]. We compute the AP for each concept and average the APs over all the concepts to obtain the mean average AP (MAP) as the overall performance evaluation. In addition we also choose the F1 micro score to evaluate both the precision and recall altogether. The F1 micro score for  $k$ th category is defined as follows

$$F_1(k) = \frac{2p_k r_k}{p_k + r_k} \quad (16)$$

where  $p_k$  and  $r_k$  are the precision and recall of the  $k$ th category, respectively.

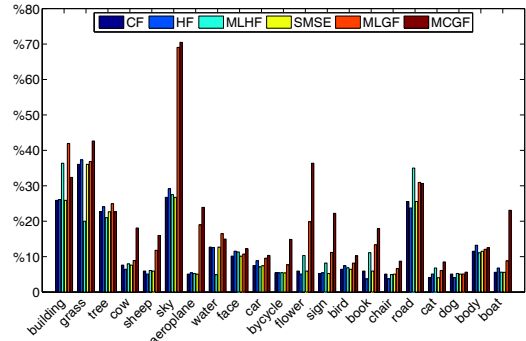
#### 4.4. Comparison and discussion

Table 1 shows MAP of the six approaches in comparison. The MLGF approach consistently outperforms the CF, HF, MLHF and SMSE approaches by more than 60% on the TRECVID 2005 data set, more than 50% on the MSRC data set and about 10% on the Barcelona data set. Furthermore the proposed MCGF approach outperforms the simple MLGF approach by another 40.69%, 19.78% and 11.47% on the TRECVID 2005 data set, MSRC data set and Barcelona data set respectively.

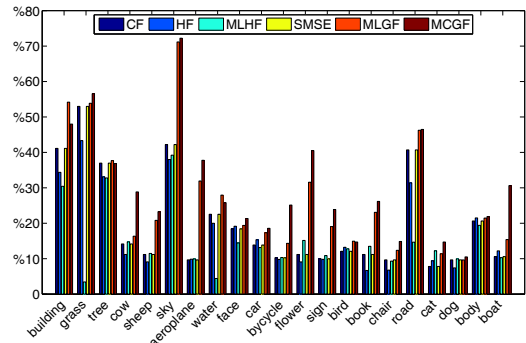
Approaches	MAP (TRECVID)	MAP (MSRC)	MAP (Barcelona)
CF	10.67%	11.72%	64.39%
HF	10.68%	11.91%	64.29%
MLHF	10.84%	12.29%	66.86%
SMSE	10.67%	11.27%	64.39%
MLGF	17.67%	18.05%	70.36%
MCGF	<b>24.86%</b>	<b>21.62%</b>	<b>78.43%</b>

Table 1: Comparison of MAP for the six approaches.

Figures 3–5 illustrates the AP and F1 micro score for the six approaches in comparison on TRECVID 2005 data set, MSRC data set, and Barcelona data set. Besides the overall performance superiority, the MLGF and MCGF consistently outperform the other four approaches for most of the classes and only degrade on few classes, e.g. “bus” in TRECVID data set. By examining the correlation matrix,  $C$ , we can see the average correlation of “bus” with other classes is among the lowest. As a result, the presence/absence of these concepts can not benefit from other classes as it has weak interaction with them. We also notice that the results of our proposed approaches in Barcelona data does not outperform other approaches as much as those in the other two data sets. Since there are only four classes in the Barcelona data set which is much less comparing with the other two data sets, there is actually no much label correlation information to be utilized.

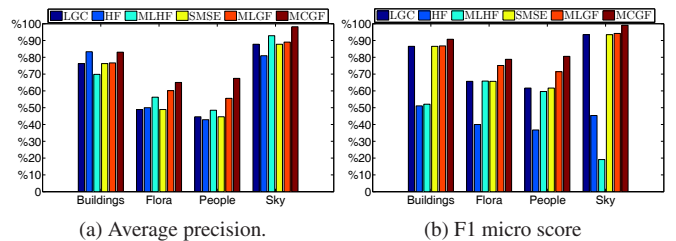


(a) Average precision.



(b) F1 micro score

Figure 4: The comparison of CF, HF, MLHF, SMSE and proposed methods on MSRC data set.



(a) Average precision.

(b) F1 micro score

Figure 5: The comparison of CF, HF, MLHF, SMSE and proposed methods on Barcelona data set.

In summary, the proposed approach outperforms the existing state-of-the-art semi-supervised multi-label classification approaches.

## 5. Conclusion

We proposed a semi-supervised learning framework to solve the automatic image annotation problem. A new multi-label correlated Green’s function approach is proposed to propagate image labels over a graph with consid-

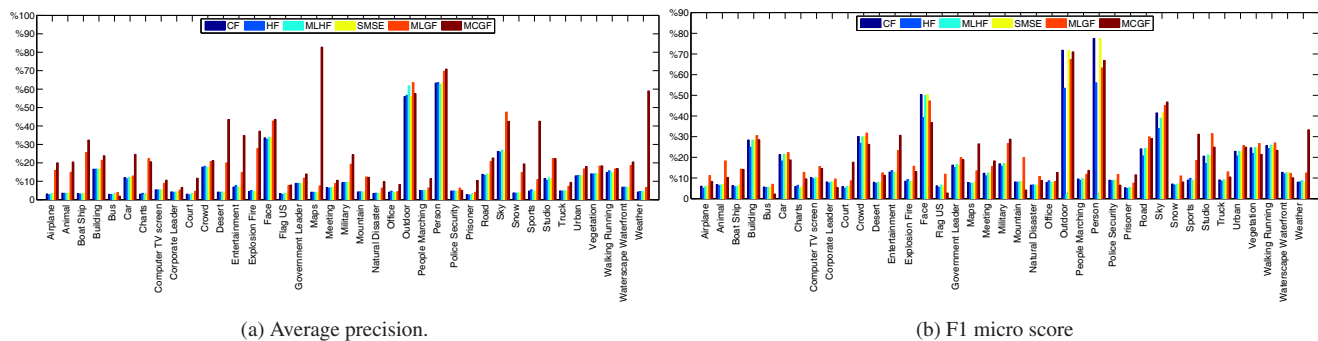


Figure 3: The comparison of CF, HF, MLHF, SMSE and proposed methods on TRECVID 2008 data set.

ering the correlations among labels. An adaptive decision boundary method is proposed to deal with the unbalanced distribution of the training data. Comprehensive experiments have demonstrated the effectiveness of the proposed approach. By incorporating the correlations among labels, the performance has been improved significant. Because we are focusing on a general semi-supervised image annotation framework, we use global features in experiments. Any local feature based image annotation methods can be integrated into our framework to improve the performance.

## References

- [1] Barcelona dataset for multi-label image annotation. <http://mlg.ucd.ie/content/view/61>.
- [2] Microsoft research cambridge: Research data. <http://research.microsoft.com/en-us/projects/objectclassrecognition/default.htm>.
- [3] Trecvid. <http://www-nlpir.nist.gov/projects/trecvid/>.
- [4] G. Carneiro and N. Vasconcelos. Formulating semantic image annotation as a supervised learning problem. In *Proc. of CVPR*, 2005.
- [5] E. Chang, K. Goh, G. Sychay, and G. Wu. CBSA: content-based soft annotation for multimodal image retrieval using Bayes point machines. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(1):26–38, 2003.
- [6] G. Chen, Y. Song, F. Wang, and C. Zhang. Semi-supervised Multi-label Learning by Solving a Sylvester Equation. In *Proc. of SDM*, 2008.
- [7] O. Dekel, C. Manning, and Y. Singer. Log-linear models for label ranking. *Proc. of NIPS*, 2003.
- [8] C. Ding, H. Simon, R. Jin, and T. Li. A learning framework using Green’s function and kernel regularization with application to recommender system. In *Proc. of ACM SIGKDD*, pages 260–269, 2007.
- [9] T. Griffiths and Z. Ghahramani. Infinite latent feature models and the Indian buffet process. *Proc. of NIPS*, 2006.
- [10] S. Hoi, R. Jin, J. Zhu, and M. Lyu. Semi-supervised SVM batch mode active learning for image retrieval. In *Proc. of CVPR*, 2008.
- [11] R. Kondor and J. Lafferty. Diffusion kernels on graphs and other discrete input spaces. In *Proc. of ICML*, 2002.
- [12] C. Leistner, H. Grabner, and H. Bischof. Semi-supervised boosting using visual similarity learning. In *Proc. of CVPR*, 2008.
- [13] A. Mccallum. Multi-label text classification with a mixture model trained by EM. In *AAAI 99 Workshop on Text Learning*, 1999.
- [14] T. Mei, Y. Wang, X. Hua, S. Gong, and S. Li. Coherent image annotation by learning semantic distance. In *Proc. of CVPR*, 2008.
- [15] M. Naphade, L. Kennedy, J. Kender, S. Chang, J. Smith, P. Over, and A. Hauptmann. LSCOM-lite: A light scale concept ontology for multimedia understanding for TRECVID 2005. Technical report, Technical report, IBM Research Tech. Report, RC23612 (W0505-104), 2005.
- [16] R. Schapire and Y. Singer. BoosTexter: A boosting-based system for text categorization. *Machine learning*, 39(2):135–168, 2000.
- [17] M. Szummer and T. Jaakkola. Partially labeled classification with Markov random walks. In *Proc. of NIPS*, 2002.
- [18] N. Ueda and K. Saito. Parametric metric models for multi-labelled text. In *Proc. of NIPS*, 2002.
- [19] K. Yu, S. Yu, and V. Tresp. Multi-label informed latent semantic indexing. In *Proc. of ACM SIGIR*, 2005.
- [20] Z. Zha, X. Hua, T. Mei, J. Wang, G. Qi, and Z. Wang. Joint multi-label multi-instance learning for image classification. In *Proc. of CVPR*, 2008.
- [21] Z. Zha, T. Mei, J. Wang, Z. Wang, and X. Hua. Graph-based semi-supervised learning with multi-label. In *Proc. of ICME*, 2008.
- [22] D. Zhou, O. Bousquet, T. Lal, J. Weston, and B. Scholkopf. Learning with local and global consistency. In *Proc. of NIPS*, 2004.
- [23] S. Zhu, X. Ji, W. Xu, and Y. Gong. Multi-labelled classification using maximum entropy method. In *Proc. of ACM SIGIR*, 2005.
- [24] X. Zhu, Z. Ghahramani, and J. Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In *Proc. of ICML*, 2003.