# LCMKL: Latent-Community and Multi-Kernel Learning based Image Annotation

Qing Li
Xi’an Jiaotong University
Xi’an, Shaanxi, China
liqing.91.10.01@stu.xjtu.edu.cn

Yun Gu
Xi’an Jiaotong University
Xi’an, Shaanxi, China
ygu@sei.xjtu.edu.cn

Xueming Qian*
SMILES LAB,Xi’an Jiaotong University
Xi’an, Shaanxi, China
qianxm@mail.xjtu.edu.cn

## ABSTRACT

Automatic image annotation is an important function for online photo sharing service. The concurrence of labels is pretty common in multi-label annotation. In this paper, we propose a novel approach called latent-community and multi-kernel learning (LCMKL). The established graph of labels is regarded as a semantic network. Community detection method is introduced that treats the label set as communities. Multi-kernel learning SVM is adopted for specifying communities and settling difficulty of extracting semantically meaningful entities with some simple features. Experiments on NUS-WIDE database demonstrate that LCMKL outperforms other state-of-the-art approaches.

## Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Content Analysis and Indexing-indexing methods; I2.10 [**Artificial Intelligence**]: Vision and Scene Understanding

## Keywords

Image Annotation, Multiple Kernel Learning, Community Detection

## 1. INTRODUCTION

With the explosive growth of web images, challenge about how to organize these resources draws wide attention. Image annotation, which specifies labels for the uploaded images, is an attractive service for the users and administrators of the online photo sharing websites like Flickr and Picasa.

In recent years, some research effort has been devoted to automatic image annotation [1]–[4], [12]-[14]. Some works focus on KNN method due to the simplicity and good performance in large scale data [2], [7], [12], [14]. However, this process rarely considers tag concurrences in multi-label annotation, which leads to low precision. Previous work [5] has shown that tag concurrence played a significant role on improving precision. Thus, tag concurrence should be considered for image annotation.

Besides, researchers have used SVM to solve image annotation due to its high accuracy. In recent years, community detection achieves great success in social network and researching its connections. We found that the connections between labels (also called ‘concepts’ or ‘tags’) similar with social network. Each label belongs to a community, like a human has its own social hub. Thus, community detection is adopted to research the concurrence between labels.

While just using simple features, it is difficult to extract semantically meaningful entities [12]. To specify communities for the images, it has demanding requirements of classification algorithms. It is often desirable to use multiple kernels instead of using single kernel [9]. Thus, MKL-SVM is proposed to specify the communities for the images.

In this paper, we propose a novel annotation method LCMKL by learning training sample based on latent community and multi-kernel learning. Figure.1 illustrates our framework, which contains two parts: offline training and online annotation.

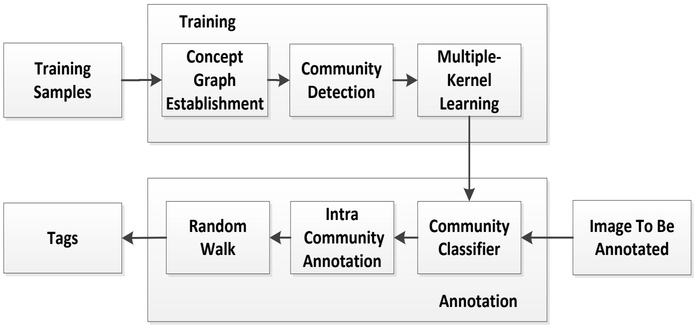

**Figure.1 The workflow of LCMKL**

- Training: Given the training samples, a concept graph is firstly established with the tagging information. Then concept communities are detected from concept graph. A community classifier is trained with Multiple-Kernel SVM based on the concept communities.
- Annotation: The corresponding community of the untagged image is firstly determined by the community classifier. Afterwards, inner-community annotation is performed with training samples according to the result of community classification. Random walk step is finally carried out to provide an extra complement for image annotation.

To testify the effectiveness of our proposed method, we perform our experiments on NUS-WIDE dataset [6], which is crawled from Flickr and has 55615 images. The experimental results demonstrate that the proposed method outperforms the state-of-the-art methods.

The main contributions of our work are as follows:

*CIKM’13*, Oct. 27–Nov. 1, 2013, San Francisco, CA, USA.
Copyright © 2013 ACM 978-1-4503-2263-8/13/10…$15.00.
DOI string from ACM form confirmation

- We construct a graph model based on posterior probability and introduce latent community concept. Each community contains related concepts and its elements are determined by community detection instead of K-means, etc. Thus, concept concurrence is fully considered and clustering reliability is drastically improved. To our knowledge, this is the first work to generalize community detection for image annotation.
- We first introduce MKL-SVM in the image annotation problem instead of utilizing SVM. The results are more accurate which guarantee the following step could be implemented successfully.

The rest of paper is organized as follows. Section 2 reviews related work on image annotation, community detection and MKL-SVM. Section 3 presents proposed method. The comparison experiments of our approaches and other state-of-the-art methods are given in Section 4. Finally, we give the conclusion in Section 5.

# 2. RELATED WORK

Recently, numerous approaches have been proposed to automatic image annotation. Given an untagged image, Zhang et al. [2] proposed ML-KNN to annotate concepts. Based on statistical information acquired from neighboring instances, the labels of given image could be determined. In 2011, Tang et al. [7] put forward a KNN sparse graph based semi-supervised learning approach. In 2012, Liu et al. [14] further proposed a graph-based dimensionality reduction for KNN-based image annotation. The aforementioned methods based on KNN fully consider the statistical information and express its simplification and efficiency [14], [15]. However, the precision of these methods was largely determined by the image set. Much training sample information was not mined. Besides, concept co-occurrence was not taken into consideration by these methods.

Apart from the mentioned approaches based on KNN, several other machine learning methods were also proposed to image annotation which considers similarity of image annotation and classification. Compared with single-label classification, which assigns an object to exactly one class, multi-label classification method should be able to assign an image to one or multiple classes. Zhang et al. [3] proposed multi-label naïve Bayes classification approach and gave the feature selection. In 2011, Zhang [4] proposed LIFT approach which constructed features specific to each label. These methods usually consider each concept as a class to label the given image. Compared with KNN, these approaches take full advantage of training sample. However, the connection between each two concepts is also ignored.

In previous work, Vailaya et al. [8] have proved that SVMs generally gain higher accuracy. Compared with traditional SVM, which has a single kernel, multi-kernel learning SVM is much better on classification accuracy [9]. Thus, MKL-SVM is used in this paper. While considering correlation between labels, several researchers ever utilize clustering algorithms and classification chains to cluster labels. However, the co-relation of concepts in image annotation has distinctive semantic network features. Thus, K-means, which is usually used for clustering, should be replaced. Here, we adopt community detection method [10], [11].

Thus, in our paper, we inherit high accuracy of MKL-SVM, efficiency of KNN and label-connection of community detection, then, propose LCMKL.

# 3. PROPOSED METHOD

## 3.1 Concept Graph Establishment

The first step of the proposed framework is the establishment of concept graph. In multi-labeling problems, the concurrence of some concepts (labels) is pretty common. A higher frequency of co-occurrences between two concepts implies a larger probability that this concept-pair will be tagged at the same time.

Based on the semantic correlations among the concepts, a directed-weighted graph $G = \{V, E\}$ is constructed. The elements of vertex set $V$ are tags from concept set $C = \{c_1, c_2, ..., c_m\}$. Two concepts $c_i$ and $c_j$ are connected with edge $e_{ij}$ if they are annotated to a tagged image at the same time. The weight of the edge implies the semantic correlation between two concepts. Let $w_{ij}$ denote the weight of $e_{ij}$ that is determined as follow:

$$w_{ij} = P(c_j \mid c_i) = \frac{N(c_i \wedge c_j)}{N(c_i)} \quad (1)$$

where $P(c_j \mid c_i)$ is the conditional probability of concept $c_j$ given $c_i$, $N(c_i)$ stands for the number of images tagged with concept $c_i$ in the image collection and $N(c_i \wedge c_j)$ stands for the number of images tagged with concept $c_i$ and $c_j$ simultaneously. Noted that $w_{i,j} \neq w_{i,j}$ , since the conditional probability of concept $c_j$ given $c_i$ is often not equivalent to the probability in reverse. For example, among the training data of NUS-WIDE's lite version, there are 4933 images tagged with "grass" and 19052 images tagged with "sky". The number of images tagged with "sky" and "grass" simultaneously is 3662. $P$('sky'|'grass') and $P$('grass'|'sky') are 0.193 and 0.733 respectively. The difference between two probabilities meets the common sense. In general, when an image is associated with 'grass', it is often related to an outdoor scene with blue sky and wide-open grassland. Conversely, the concept 'sky' may appear in other scenes like urban views or coastal landscapes which are not associated with 'grass'.

## 3.2 Community Detection

The concepts in image annotation have distinctive semantic network features. Based on the constructed graph, the densely connected concepts can be clustered into several communities which are the sets of highly inter-connected nodes. The connection between different communities should be sparse. The quality of the community detection is often measured by the modularity [10] of the partition. The modularity of a community is a real number between -1 and +1 that measures the density of intra-community links compared to the inter-community ones. Given a concept graph $G = \{V, E\}$ portioned into $m$ communities, which are denoted with $S = \{s_1, s_2, ..., s_m\}$, the modularity can be defined as follow:

$$Q = \sum_{s=1}^{m} \left[ \frac{l_s}{|E|} - \left(\frac{d_s}{2|E|}\right)^2 \right] \quad (2)$$

where $l_s$ denotes the numbers of inter-community links connected with community $s$ and $d_s$ denotes the sum of degree of intra-community concepts. Higher modularity of communities leads to a better partition quality.

A fast unfolding algorithm [11] is applied in this paper to realize the community detection. It has been proved a promising algorithm to generate proper communities under optimal time-complexity. The concept set $C$ will be clustered into $m$ communities and the tagged images will be re-annotated as follows: Given an image initially tagged with concepts $\{c_1, c_2, ..., c_k\}$ and the concepts belonging to community $\{s_1, s_2, .., s_m\}$.

Choose the community which includes the most part of the original tags. As a result, the tagged images of training set will be grouped into specific communities.

## 3.3 Multiple-Kernel SVM

After community detection and image re-annotation, the tagged images will be associated with only one community. In this paper, a multiple-kernel SVM model is applied to solve this multi-label classification problem.

Classifiers trained with only one visual feature is not robust and proper for predicting the community label of untagged images. Traditional SVM model combines all visual features into an entire vector which leads to a curse of dimensionality. Besides, features are likely to be treated differently in specific classifying scenes. For example, for two communities, one includes 'sky', 'water' and 'ocean' while the other one includes 'grass' and 'tree'. Color histogram features play more significant roles than edge detection histogram and wavelet texture. Hence, different visual features should have unique weights in classification. The multiple-kernel SVM model can be trained with adaptively-weighted combined kernels and each kernel is in accordance with a specific type of visual feature. The combined kernel is as follows:

$$K(x,y)=\sum_{j=1}^{n_{Kernel}}\beta_j K_j(x,y) \quad \beta_j \ge 0, \sum_{j=1}^{n_{Kernel}}\beta_j = 1 \quad (3)$$

where $K$ is the combined kernel, $K_j$ is the sub-kernel for the $j$ th visual feature and $\beta_j$ is the weight for $K_j$. and $x$, $y$ are visual features of images. The constraints are so-called '$l_1$-norm' constraints which can generate a sparse solution for sub-kernel weights. The binary decision function can be determined as follows:

$$f(x)=\sum_{i=1}^{k}\alpha_i \sum_{j=1}^{n_{Kernel}}\beta_j K_j(x_i,x)+b_i \quad (4)$$

Since the determination of community is a multi-class classification problem, the 'one-vs-one' strategy is adopted in this paper. For $q$ communities, there are $q(q-1)/2$ binary classifiers to be trained. Given an untagged image $x_u$, the top two communities will be determined by the trained community classifier. All visual features of the images can be adopted in multiple-kernel SVM with different weights.

## 3.4 Intra-community Annotation

The top two communities of an untagged image can be determined by the trained community classifiers. Let $X_i$ denote the initial tagged images belonging to community $i$ in training set. The visual features with largest weights in MKL-SVM are chosen and combined as a feature vector. ining set. A naïve KNN search is carried out in each community. We take the Euclidean distance between the image features of the images as the similartity measurements. For the $k$-nearest images in each community, tagging status of each can be represented as $m$-dimensional binary vector $t_j$,$(j=1,2...,k)$ where $m$ is the number of concepts. The concept-probability vector $T_p$ can be generated as follows:

$$\begin{cases} T_p = \{T_{p,1}, T_{p,2}, \ldots, T_{p,m}\} \\ T_{p,j} = \dfrac{1}{k}\sum_{i=1}^{k} t_{i,j} \end{cases} \quad (5)$$

where $t_{i,j}$ is a 0-1 value which implies the tagging status of $j$ th concept in image $i$. The final annotation for the untagged image is determined as follows:

$$t_i = \begin{cases} 0 & \text{if } T_{p,i} > Threshold; \\ 1 & otherwise \end{cases} \quad (6)$$

where the 'Threshold' is associated with the training set.

## 3.5 Random Walk on Concept Graph

Images are annotated with some concepts after intra-community annotation while it involves some problems. The intra-community annotation is carried out in top two communities separately. In another words, the potential tagged concepts will only come from these two communities but not others. Some concepts which are highly correlated with the tagged ones but not belong to the top two communities will not be included. Therefore, a Random Walk strategy is applied here to compensate for the deficiency.

For a concept $c_i$ tagged after intra-community annotations, find its directly connected concepts $\{c_{d,1}, c_{d,2}, \ldots, c_{d,t}\}$ based on the concept graph. If $c_{d,j}$ is not tagged to this image and the conditional probability $P(c_{d,j}\,|c_i)$ exceeds the confidence threshold (e.g. 0.8), this concept will be included.

# 4. EXPERIMENTS

The proposed method is tested on NUS-WIDE data set. [6] It is a large-scaled real-world data set crawled from Flickr. The data is composed with two parts: the training part, which contains 27807 images, and testing part, which contains 27808 images. All images are tagged with the concepts from 81 Ground Truth. The annotation model is trained from the training part and the evaluation of the model is based on the testing part. The low-level features extracted from the image including color histogram (64D), color correlation histogram (73D), edge-detection histogram (73D), block-wised color moments (256D) and wavelet texture (128D). All features are adopted in the community classifier and the most distinguished feature(s) are selected in inner-community annotation. In this paper, ***Precision*** , ***Recall and F1-score*** are used to measure the performance of image annotation. We compare our method with state-of-the-art approaches. ML-KNN [2], ML-NB [3] which are proved efficient methods for automatic image annotation will be carried out for comparison.

Based on the tagging information of training part, a concept graph is constructed. After the step of community detection, the graph is divided into nine communities The proper result of community detection may be helpful to train the community classifiers. Images in tagging parts are re-annotated according to the result of community detection and the initial tags.Five visual features are adopted in the training of community classifiers based on MKL-SVM model. The 'one-vs-one' strategy is used in this multi-class classifying problem. The optimal selection of sub-kernel weights and SVM parameters are solved with SimpleMKL algorithm by shogun-toolbox 2.0[9].

If the test images are tagged with the most possible community, the tagging precision is 0.613, not a satisfying result. When the tagged communities are expanded to the top two possible ones, the precision rises to 0.827 accordingly. Therefore, each test image is labeled with two communities in which the training images for intra-community annotation are selected.

According to the result of community classification, the optimal sub-kernel weights are also obtained. The most distinguished visual features are Color Moments and Color Correlation Histogram since their weights are 0.62 and 0.22 in average. They will be combined to a feature vector for intra-community annotation and random walk process. The result of annotation is presented in Table 1.

Table 1 and Figure 2 show the annotation result on NUS-WIDE-lite data set in comparison with ML-KNN and ML-NB. The performance of average recall with LCMKL is 0.3997 which is 42.4% higher than ML-KNN and 15.5% higher than ML-NB, which is the best-performed method. The average precision of LCMKL is 0.33, which indicates a remarkable performance compared with ML-KNN and ML-NB. It should be noted that the average precision of LCMKL suffers a slight loss after random work process while the average recall rises over 24.2%. In general, the recall of each concept rises after random walk process. For the most improved concept 'sky' and 'clouds', the recalls are respectively 0.675 and 0.493 higher than the result without random walk. The improvement is associated with the number of relevant instances and training part re-annotation. According to the procedure of intra-community annotation, only the test images are classified to **Community 3** can be tagged with 'sky'. With the help of random walk, the concepts semantic-correlated with 'sky' but not belonging to **Community 3** can lead to the recognition of 'sky'.

**Table.1 The performance comparison on NUS-WIDE**

| Measures | ML-KNN | ML-NB | LCMKL<br>No RW | LCMKL<br>RW |
|---|---|---|---|---|
| Avg. Recall | 0.2807 | 0.3460 | 0.3218 | 0.3997 |
| Avg. Precision | 0.0529 | 0.1453 | 0.3248 | 0.3030 |
| Avg.F1-score | 0.0561 | 0.1932 | 0.2725 | 0.2821 |

*RW is short for Random Walk.

## 5. CONCLUSION AND FUTURE WORK

In this paper, a Latent-Community and Multiple-Kernel Learning based framework for automatic image annotation is proposed. Our work integrates the community features of multi-labeled images and the multiple-kernel learning. A concept graph is constructed which implies a dense semantic intra-community correlation of concepts. The multiple-kernel SVM is applied for community classification. Further intra-community annotation is enhanced by Random Walk strategy. To evaluate the performance of our method, we conduct our method on NUS-WIDE library. From the result of experiment, it can be seen that our method outcomes the classical and state of art method for image annotation. In the future, we will study the optimal selection for visual features and MKL-SVM parameters. And the sparse-coded intra-community annotation will be applied to improve the performance of LCMKL.

## 6. ACKNOWLEDGEMENT

This work is supported by the NSFC No.60903121, No.61173109, and MSRA. This work is completed when Qing and Yun are in their visiting at SMILES LAB

## 7. REFERENCES

[1] A. Elisseeff, and J. Weston, A kernel method for multi-labeled classification. *Advances in neural information processing systems*, 14(2001-01-01 2001), 681-687.

[2] M. Zhang, and Z. Zhou, ML-KNN: A lazy learning approach to multi-label learning. *Pattern Recogn*, 40, 7 (2007-01-01 2007), 2038-2048.

[3] M. Zhang, J. Peña, and V. Robles, Feature selection for multi-label naive Bayes classification. *Inform Sciences*, 179, 19 (2009-01-01 2009), 3218-3229.

[4] M. Zhang, LIFT: Multi-label learning with label-specific features. AAAI Press, 2011.

[5] Y. Gao, J. Fan, X. Xue, and R. Jain, Automatic image annotation by incorporating feature hierarchy and boosting to scale up SVM classifiers. ACM, 2006.

[6] T. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y. Zheng, NUS-WIDE: a real-world web image database from National University of Singapore. ACM, 2009.

[7] T. Tang, R. Hong, S. Yan, T. Chua, G. Qi, and R. Jain, Image annotation by k NN-sparse graph-based label propagation over noisily tagged web images. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2, 2 (2011-01-01 2011), 14.

[8] A. Vailaya, H. Zhang, C. Yang, F. Liu, and A. Jain, Automatic image orientation detection. *Image Processing, IEEE Transactions on*, 11, 7 (2002-01-01 2002), 746-755.

[9] Sonnenburg, S., Rätsch, G., Schäfer, C. and Schölkopf, B. Large scale multiple kernel learning. *The Journal of Machine Learning Research*, 7(2006-01-01 2006), 1531-1565.

[10] M. Newman, Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103, 23 (2006-01-01 2006), 8577-8582.

[11] V. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre,. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008, 10 (2008-01-01 2008), P10008

[12] A. Makadia, V. Pavlovic, and S. Kumar, Baselines for image annotation. *Int J Comput Vision*, 90, 1 (2010-01-01 2010), 88-105.

[13] D. Kong, C. Ding, H. Huang, and H. Zhao, Multi-label ReliefF and F-statistic feature selections for image annotation. *IEEE*, 2012.

[14] X. Liu, R. Liu, F. Li, and Q. Cao, Graph-based dimensionality reduction for KNN-based image annotation. *IEEE*, 2012.

[15] X. Qian, X. Liu, C. Zheng, Y. Du, and X. Hou, "Tagging photos using users' vocabularies", Neurocomputing, vol.111, 2013, pp.144-153.

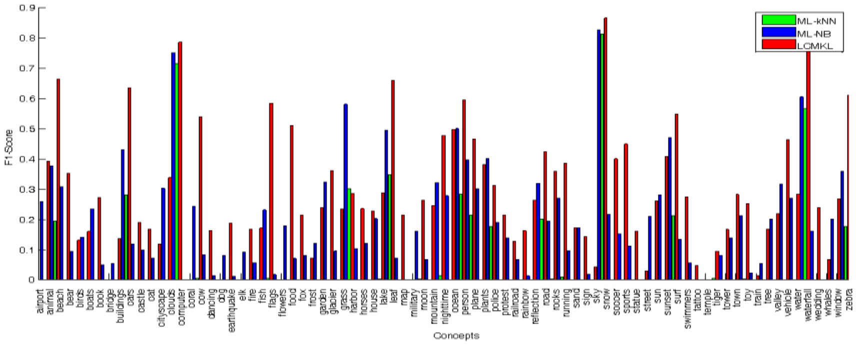

**Figure.2 Annotation Result(F1-score) of MLKNN,MKNB and LCMKL**