Multi-modal Multi-label Semantic Indexing of Images using Unlabeled Data

Wei Li and Maosong Sun
State Key Lab of Intelligent Technology and Systems
Department of Computer Science and Technology, Tsinghua University
wei.lee04@gmail.com, sms@mail.tsinghua.edu.cn

ABSTRACT

Automatic image annotation (AIA) refers to the association of words to whole images which is considered as a promising and effective approach to bridge the semantic gap between low-level visual features and high-level semantic concepts. In this paper, we formulate the task of image annotation as a multi-label multi class semantic image classification problem and propose a simple yet effective algorithm: hybrid self-learning with alternating space between uni-modality and bi-modality, which integrate multi-label boosting with asymmetric binary SVM-based active learning into a joint hierarchical classification framework to perform cross-modal image annotation by incorporating unlabeled images. We conducted experiments on a medium-sized image collection including about 15000 images from Corel Stock Photo CDs. The experimental results demonstrated that our proposed method can enhance a given annotation model by using unlabeled images to some extent, showing the effectiveness of the proposed algorithm and the feasibility of unlabeled data to help the annotation accuracy.

1. INTRODUCTION

Automatic image annotation refers to the association of words to whole images which has become a hot research topic in the domain of multi-modal indexing, semantic scene classification and medical image interpretation. Through the sustained efforts of experts and researchers, many approaches have been proposed to attack this problem, which, in general, can be categorized into three major classes: generative model based on EM (Expectation-Maximization) algorithm [1-4, 6-8, 13-18, 26]; discriminative approaches [5, 9-12, 20, 23] and search-based annotation [22]. Some of them have achieved state-of-the art performance. However, one of the major obstacles to automatic image annotation is the limited number of labeled training images, especially multi-labeled images, since it requires large amount of labeling effort of experienced annotators. How to incorporate the large pool of unlabeled images into the process of model training still remains an open issue. In the literature, there have been some works using semi-supervised learning and active learning to exploit the unlabeled data for image retrieval [27, 28] and automatic image annotation [24, 25]. H. Feng et al [25] applied co-training(semi-supervised multi-view learning) and decision tree to propagate the keywords to unlabeled images based on a labeled training set with explicit correspondence between image patch and keyword annotation. The key differences from previous works and our contributions are : 1) we propose to use the multi-label learning algorithm instead of traditional single-label multi-class classification model to solve the image annotation problem, which is able to assign multiple labels to images based on global image features which can further avoid the unstable image segmentation; 2) asymmetric binary SVM-based active learning is integrated with multi-label boosting to exploit the pool of unlabeled images; 3) the integrated hybrid self-learning framework based on multi-label boosting and asymmetric binary $\text{SVM}_{\text{active}}$ is learned through alternating space between uni-modality and bi-modality which is demonstrated to be more effective based on empirical results. To the best of our knowledge, the combination of multi-label boosting and asymmetric binary $\text{SVM}_{\text{active}}$ through alternating space has not been extensively studied for the task of automatic image annotation.

2. RELATED WORK

Recently, many models using machine learning techniques have been proposed for automatic image annotation and retrieval. In general, these methods can be categorized into three classes: generative models, discriminative approaches as well as search and mining-based techniques.

2.1. Generative Probabilistic Models

$$P(l, v) = \sum_{s} P(l|s)P(v|s)P(s) \quad l \subseteq L, v \in V \qquad (1)$$

where v dentoes the image data, l the subset of semantic concepts, s is the latent variable, L and V are concept lexicon and visual feature space respectively. By computing the joint distribution of visual features and associated concepts, the hidden correlation between these two modalities can be found and then is applied to annotate new images. Representative works are [1-4][6-8][13-18][26], especially R. Zhang et al[18] has achieved the state-of-the-art performance, G. Carneiro et al[26] proposed to use M-ary labeling and ignore the hidden variable which can reduce the model complexity.

2.2. Discriminative Models

$$P(w|v) \quad w \in L, v \in V \qquad (2)$$

where w is a concept from L. Instead of joint modeling of semantic concepts and visual features, discriminative approaches treat each concept as a single class label and directly model the posterior probability of w given v. Some attractive works are [5][9-12][20][23]. Among them, K.Goh et al[10] and Cees G.M. Snoek[23] can provide better results. M. Bouttell et al[19] proposed the cross-training

978-0-7695-3273-8/08 $25.00 © 2008 IEEE
DOI 10.1109/ALPIT.2008.107

method to conduct multi-label scene classification and introduced some specific evaluation metrics.

In short, generative models can handle a large number of classes and class imbalance problem in some degree, but the model complexity is a major hurdle. While discriminative approaches are computationally efficient, however, they are unable to scale well to a large number of classes since it requires one model to be built for each class.

2.3. Search and Mining-based Annotation

Apart from annotation by learning, Wang et al. [22] proposed annotation by search and mining techniques which can not only makes use of web-scale images but also allows for unlimited vocabulary.

The common characteristic of first two categories is that they all follow the idea of automatic classification of image and video content into one or more of a large number of predefined semantic concepts by using machine learning algorithms and the third category takes the visual matching and text mining strategy without constructing any learning or association models.

More recently, learning with unlabeled images has become an active research area due to fact that large amount of labeled training images is hard to obtain or create in large quantities while limited number of training images can hardly represent the visual distribution of target concepts and more information is contained in the large pool of unlabeled ones. Feng et al [25] and Song et al [24] introduced the use of co-training and combination of active learning together with semi-supervised ensembles to perform semantic annotation of images and video clips.

3 MULTI-MODAL MULTI-LABEL SEMANTIC INDEXING OF IMAGES

3.1. Formulation of Image Annotation Model

Automatic image annotation is the task of automatically generating the semantic labels for images to describe the image semantics. Given a training set of annotated images, where each image is associated with a number of semantic labels, we then make the assumption that each image can be considered as a multi-modal document containing both the visual component and semantic component. Visual component provides the image representation in visual feature space using low-level perceptual features including color and texture, etc. While, semantic component captures the image semantics in semantic feature space based on textual annotations derived from a generic vocabulary, such as "*sky*", "*ocean*", etc. Automatic image annotation is the task of discovering the association model between visual and semantic component from a labeled image corpus and then applying the association model to automatically generate annotations for unlabeled images. More formally, let *ID* denote the training set of annotated images:

1. $ID = \{I_1, I_2, \ldots, I_N\}$

2. Each image I_j in *ID* can be represented by the combination of visual features and semantic labels in a multi-modal feature space, i.e., $I_j = \{L_j; V_j\}$

3. Semantic components L_j, a bag of words described by a binary vector $L_j = \{l_{j,1}, l_{j,2}, \cdots, l_{j,m}\}$ where m is the size of generic vocabulary, $l_{j,i}$ is a binary variable indicating whether or not the i-th label l_i appears in I_j.

4. Visual component V_j may be more complex due to large variety of methods for visual representation, in general, it can also have the vector form $V_j = \{v_{j,1}, v_{j,2}, \ldots, v_{j,n}\}$, for patch-based image representation, image I_j is composed of a number of image segments or fixed-size blocks, each of them is described by a feature vector $v_{j,i}$, and n is the number of image components; for global image representation, $v_{j,i}$ only denotes a feature component and n is the dimension of the selected feature space.

For a given unseen image represented by v_u, the goal of automatic image annotation is to estimate:

$$l^* = \arg\max p(l|v_u), \quad l \subseteq L, v_u \in V \tag{3}$$

3.2. Multi-label Boosting for Image Annotation using Unlabeled Data

3.2.1. Multi-label Boosting for Image Annotation

In traditional classification problems, class labels are assumed to be mutually exclusive and each instance to be classified belongs to only one class. However, in the context of image annotation or semantic image classification, it is natural that one image belongs to multiple classes simultaneously since image semantics is represented by both multiple semantic entities contained in the image and the relationships between them, causing the actual classes overlap in the feature space. Motivated by R. Schapire et al [19]. We formalize the problem of image annotation as a multi-label multi-class semantic image classification task. More formally, given a training set of labeled images, let *D* be the collection of images to be classified and *L* be the finite set of semantic labels, each image $I_j \in D$ is associated with a binary label vector $L_j = \{l_{j,1}, l_{j,2}, \cdots, l_{j,m}\}$, then we convert the multi-label images into several single-label image pairs (I_j, l_i) using the criterion that each image serves as an observation for each of the classes to which it belongs, finally these converted single-label instances are taken as input to train a multi-label annotation model based on the boosting framework. The detailed boosting algorithm for multi-label, multi class semantic image classification is shown as follows.

Multi-label Boosting for Image Annotation Algorithm

Input: $I = \{(I_1, L_1), (I_2, L_2), \ldots, (I_n, L_n)\}$, a sequence of multi-labeled images; *T*, number of iterations

Output: $f(I, l)$ final accurate annotation model

1. initialize $D_1(j, l) = 1/(M)$ (M is the total number of single-label image pairs)

2. for t = 1 : *T*

3. pass distribution D_t to weak learner
4. get weak annotator $h_t : ID * L \to \Re$
5. choose $\alpha_t \in \Re$
6. update

$$D_{t+1}(j,l) = \frac{D_t(j,l)\exp(-\alpha_t l_j h_t(I_j,l))}{Z_t}$$

where Z_t is a normalization factor to ensure that D_{t+1} is a distribution

$$Z_t = \sum_{i=1}^{n}\sum_{l\in L} D_t(j,l)\exp(-\alpha_t l_j h_t(I_j,l))$$

7. output the final annotation model

$$f(I,l) = \sum_{i=1}^{T} \alpha_i h_i(I,l)$$

Distribution D_t denotes the importance weights over single-label image pairs. Initially, the distribution is set to uniform. For each iteration, the sequence of multi-labeled images together with the D_t are taken as input to compute a weak annotator, $h(I_j,l) : ID * L \to \Re$ then a parameter α_t is chosen and D_t is updated based on the actual labels and the predicted labels of the weak annotator. Through iterations, the next weak annotator will pay more attention to these example pairs with higher weights that are most difficult to classify by the preceding weak annotator. Finally, a highly accurate annotation model is constructed by combining all the weak models through weighted voting. To annotate an unlabeled image, the final annotation model can output a weight vector. In order to produce a reasonable ranking of the labels to be used as annotations, we map the associated weights in the weight vector via the following logistic function to get confidence scores or annotation probabilities for each label.

$$P(y=1|f) = \frac{1}{1+\exp(Af+B)} \qquad (4)$$

where f is the output of the final annotation model for a given unseen image, A and B are real-valued parameters estimated by Maximum Likelihood criterion.

3.2.2. Learning with the Unlabeled Images

The performance of the image annotation accuracy heavily depends on the size of the labeled training data. However, in most cases multi-labeled images are difficult to create or obtain in large quantities, while unlabeled images are easier to collect. Hence, there have been increasing interests in using semi-supervised and active learning algorithms which attempt to exploit the unlabeled images to improve the annotation performance. The key issue underlying semi-supervised is how to enhance the classification performance using both labeled and unlabeled simultaneously, while the goal of active learning is how to select the most informative unlabeled examples from a pool of unlabeled data, which aims to optimize the classification performance while minimizing the number of needed labeled examples for classifier training. In this paper, we propose an algorithm called hybrid self-learning with alternating space in which asymmetric binary SVM_{active} is integrated with multi-label boosting to build a two-level classification framework to perform the task of automatic image annotation by using both the labeled and unlabeled images. The main idea is, given a training set of multi-labeled images, two classifiers are built, the first-level multi-class multi-label classifier based on uni-modal image representation and second-level re-ranker model based on bi-modal image representations respectively. Here, first-level multi-label classifier and second-level re-ranker model serves as the base learning algorithm and data selection strategy. For each image in the unlabeled pool, the multi-label classifier is first used to predict the possible labels, and then the re-ranker model is responsible for determining whether or not each predicted label is appropriate to describe the image semantics. Finally, the predicted appropriate label together with the unlabeled image is added to the training set and the re-training of first-level multi-label classifier and second-level re-ranker classifier is performed. The iterative process is terminated until the unlabeled pool is empty or the annotation performance on a validation set can not be increased any more. To be more formal, let X be the image data, Y the finite set of predefined semantic labels and the size of Y is denoted by k. For the first-level multi-labeled classifier training, each training pair has the uni-modal form of (x, y) where $x \in X, y \subseteq Y$. While, for the second-level re-ranker model, the training data is derived using a natural reduction of multi-labeled data to binary data. To be more specific, each example is mapped to a k binary-labeled bi-modal meta-examples which takes the form $((x, l, r), y[l])$ for all $l \in Y$, where $y[l] = 1$ if $l \in y$ and -1 otherwise, r denotes the correlation between the label l and all the other labels. In this paper, the correlation among different labels is obtained by using latent semantic indexing. That is, the observation of each derived meta-example is (x, l, r), and the associated binary class label is $y[l] \in \{-1, 1\}$. To select the informative images from the unlabeled pool, the multi-label classifier is initially applied to each unlabeled image and a label list containing candidate labels is output. Each candidate label and the correlation between this label and other labels is then appended to the feature vector of the unlabeled image to form the above-mentioned bi-modal meta-example; this meta-example is then classified by the re-ranker model to examine if each predicted label is relevant to the unlabeled image. Finally, the relevant labels together with the image itself are added to the multi-labeled training set and the re-retraining of the whole annotation framework is performed. In other words, the main task of the re-ranker is to conduct meta-example identification, to identify the positive and negative ones, then the appended label in the positive meta-example is considered as the correct label for the corresponding image and is positioned front predicted label list while the appended label in the negative one is moved back in the label list. In our experiments BoosTexter is used as the first-level multi-label classifier while SVM_{active} ensemble serves as re-ranker model. In addition, in most multi-labeled image collections,

the number of semantic labels for each image is rather small compared to the total number of predefined semantic labels, the produced bi-modal training data is extremely imbalanced in the sense that the number of negative meta-examples is much larger than that of positive meta-examples. To avoid the performance degradation of SVM_{active} ensemble model due to the class imbalance problem, we propose to use the asymmetric bagging [21] to generate a classifier ensemble. The key idea behind asymmetric bagging is that keeping positive meta-examples the same for each base classifier and bootstrapping is only performed on the negative meta-examples to sample the same number as the positive meta-examples to construct a balanced training set. To build a desired SVM_{active} ensemble model, maximizing the diversity of each base SVM classifier while maintaining the consistency with the training data is known to be an important goal, so in our method, each sampled negative subset is different from each other to ensure the diversity of training data. Figure 3.1 and 3.2 show the uni-modal and bi-modal feature representation and the transformation as well as the asymmetric bagging algorithm.

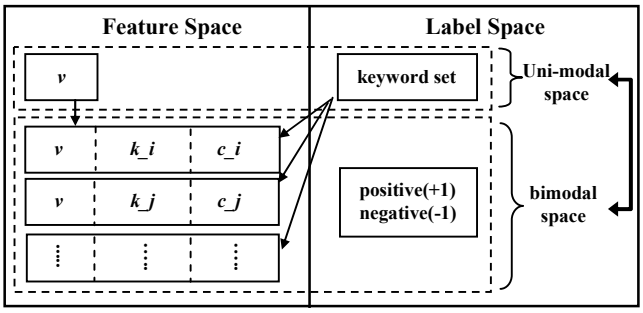

Fig. 3.1 uni-modal and bi-modal feature representation and transformation

In Figure 3.1 *v* is an abbreviation for visual features of a given image, *k_i* denotes one semantic label of the predefined lexicon and *c_i* represents the correlation between one label and the rest labels. Annotation keywords or semantic labels in uni-modal space serves as the class labels for images represented in uni-modal visual features while for bi-modal meta-examples the associated class labels range from either -1 to +1.

Asymmetric Bagging Algorithm:

Input: positive meta-examples S^+, negative meta-examples S^-, base classifier I, number of base classifiers N, sampling factor α and the test meta-example t.

Output: final label l and classifier ensemble C

1. for i = 1 to N
2. S_i^- bootstrap samples from S^- using the criterion that $\alpha|S_i^-| = |S^+|$.
3. $I_i = I(S^+, S_i^-)$
4. $l = majority\_voting(I_i(x, S^+, S_i^-))$, $C = \{I_i\}$

Fig. 3.2 Asymmetric Bagging Algorithm

In this proposed framework, the second-level binary SVM_{active} classifier plays an important role that exploits the pool of unlabeled images to select or identify the most informative images for ML-Boosting classifier re-training. Figure 3.1 illustrate the iterative training procedure through alternating space between uni-modality and bi-modality.

The reasons for using asymmetric binary SVM_{active} instead of direct bootstrapping the ML-Boosting to select the useful unlabeled images which can contribute to the performance improvement of ML-Boosting are: 1) since the predefined lexicon contains 16 keywords, transforming the instances in uni-modal space into bi-modal space result in a imbalanced training set in which the number of positive examples is much less than the number of negative examples due to that each image is usually associated with only 3-5 keywords. Asymmetric binary SVM_{active} have been demonstrated to be effective for tackling unstable generalization performance and biased hyperplane caused by the small-sized and imbalanced training data [20], its basic idea is to random sample the same number of negative examples as that of the positive examples, which can produce a balanced training set; 2) The essential idea of self-training or bootstrapping is to add the most confident predicted instances to training set and then perform classifier re-training, which heavily depends on the initial performance of the wrapped classifier. However, in most cases most-confident instances are not the most informative ones which can contribute to the classifier performance improvement, if the corresponding predicted labels are wrong, it can even degrade the classifier capability. Furthermore, for multi-label classification, top-ranked N ($N > 1$, in most cases) class labels should be selected for the most-confident examples to bootstrap the multi-label classifier re-training, but how to adaptively determine N is another difficult question. By using alternating space and asymmetric binary SVM_{active}, we can transform each multi-labeled example to several binary-labeled meta examples and identify the most-confident ones with positive label, which can avoid these above two problems to some extent. Figure 3.3 and 3.4 shows the hybrid self-learning classification model and the detailed algorithm for the proposed model.

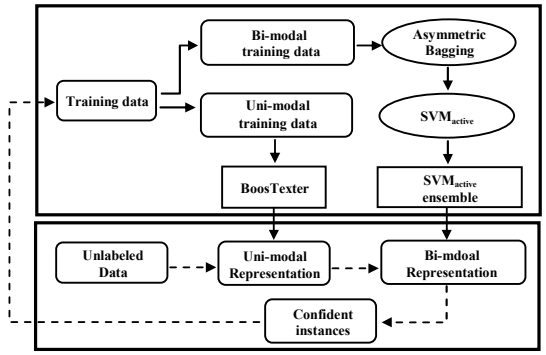

Fig. 3.3 Hybrid Self-Learning Classification Model

ML-Boosting based on Alternating Space and Asymmetric binary SVM$_{active}$

Input: L^T, Labeled data for training ML-Boosting; L^V, Labeled data for validation; U, Unlabeled data;

Output: a trained hybrid annotation model λ_*

1. λ_τ is trained using L^T
2. if λ_τ has no improvement on L^V
3. terminate training
4. else
5. create labeled training set L^E in bimodal space via transforming the instances with ground-truth multiple labels in L^T described in uni-modal space
6. perform the binary SVM B training using asymmetric bagging based on L^E
7. apply λ_τ to annotate the images in U, select top N class labels for each unlabeled instance and create the unlabeled pool U^E in bimodal space via transforming these instances with predicted N labels in U represented in uni-modal space
8. perform B re-training using active learning
9. apply final B to classify the instances in U^E and select the most-confident ones with positive predicted label
10. transform these instances from bimodal space into uni-modal space and add them to L^T
11. perform λ_τ re-training using the augmented L^T
12. return λ_τ as λ_*

Fig. 3.4 Hybrid Self-Learning Algorithm

4. EXPERIMENTAL RESULTS

In this paper, we conducted experiments on a subset of the Corel image collections which is widely used in the domain of image annotation. We have approximate 15000 images with the size of 192 * 128 pixels and 120 * 80 sizes, in which each image is associated with 3-5 keywords describing the image semantics and no correspondence between image patches and keywords is provided. Global image features like color moments in HSV color space and Gabor filters with 6 scales and 4 orientations are extracted to characterize the visual content of images. The predefined generic vocabulary contains 16 keywords. We divided the image dataset into 3 parts, 4500 images as the training data, 500 images as the validation set and the remaining 10000 images as the unlabeled pool. In our experiments, the weak classifier used for multi-label boosting is one-level decision tree and for each round of iterations in multi-label boosting, a threshold is generated from the corresponding feature value, decision is determined based on whether the observation value is above or below the given threshold and α_t is set to 1.For achieving computational efficiency, the number of iteration for ML-Boosting and active learning is fixed to 173 and 10 respectively and the selection strategy for SVM$_{active}$ is Kernel Farthest First (KFF) [29]. Top 4 keywords with highest probability are selected as the image annotation. In order to verify that asymmetric binary SVM$_{active}$ can provide better performance, we have tested other popular approaches which can be trained on both labeled and unlabeled examples including Transductive SVM and Graph-based Semi-supervised Learning (Graph-based SSL) [30]. Figure 4.1 shows the comparison of classification error of ML-Boosting on validation set using different binary classifiers at second level as well as direct bootstrapping of ML-Boosting.

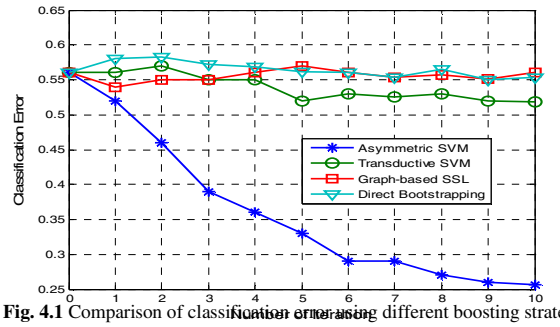

Fig. 4.1 Comparison of classification error using different boosting strategies

The proposed alternating space and asymmetric binary SVM$_{active}$ algorithm can also be integrated with other annotation models [4]. Table 4.1 shows some annotation result and Figure 4.2 shows the precision improvement using single word query to retrieve the multi-labeled image corpus.

Table. 3.1 Automatic image annotation results

Images	True Annotation	Automatic Annotation
	Sand grass water sky	sky water trees grass
	airplane sky	Sky airplane cloud
	grass trees house mountains sky	trees grass mountains rock

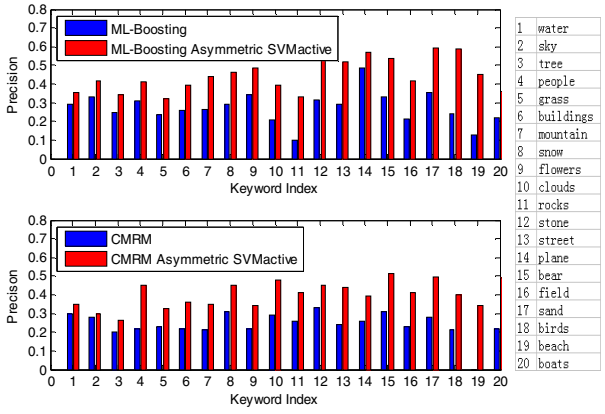

Fig. 4.2 Precision improvement of different annotation models

5. CONCLUSIONS

In this paper, we propose the algorithm: hybrid self-learning with alternating space which integrates asymmetric binary SVM_{active} with multi-label boosting classifier to conduct automatic image annotation. The advantage of this algorithm is that it can be integrated with other annotation models by exploiting the unlabeled pool, whereas the disadvantage also exist that it applies random asymmetric bagging to train binary SVM that may lose some representative negative examples of interest and if the predefined lexicon is too large, the transformed bi-modal meta-examples will produce an extremely imbalanced dataset which may result in computational complexity of SVM and degrade the performance of annotation model.

6. REFERENCES

[1]. K. Barnard, P. Dyugulu, N. de Freitas, D. Forsyth, D. Blei, and M. I. Jordan. Matching words and pictures. Journal of Machine Learning Research, 3: 1107-1135, 2003.

[2]. K. Barnard and D. A. Forsyth. Learning the Semantics of Words and Pictures. In Proceedings of International Conference on Computer Vision, pages 408-415, 2001.

]3]. P. Duygulu, K. Barnard, N. de Freitas, and D. Forsyth. Ojbect recognition as machine translation: Learning a lexicon fro a fixed image vocabulary. In Proc. of ECCV'02, 97-112, 2002.

[4]. J. Jeon, V. Lavrenko and R. Manmatha. Automatic image annotation and retrieval using cross-media relevance models. In Proc. of SIGIR'03, 119-126, 2003.

[5]. Edward Chang, Kingshy Goh, Gerard Sychay and Gang Wu. CBSA: Content-based soft annotation for multimodal image retrieval using bayes point machines. IEEE Transactions on CSVT 13(1): 26-38, 2003., 13(1): 26-38, 2003.

[6]. J. Li and J. A. Wang. Automatic linguistic indexing of pictures by a statistical modeling approach. IEEE Transactions on PAMI, 25(10): 175-1088, 2003.

[7]. V. Lavrenko, R. Manmatha and J. Jeon. A model for learning the semantics of pictures. In Proc. of the 16th Annual Conference on Neural Information Processing Systems, 2004.

[8]. D. Blei and M. I. Jordan. Modeling annotated data. In Proceedings of the 26th intl. SIGIR Conf, 127–134, 2003.

[9]. B. Li and K. Goh, Confidence-based dynamic ensemble for image annotation and semantics discovery, in Proc. of ACM MM'03, 195-206, 2003.

[10]. K.Goh, B. Li and E. Chang, Semantics and feature discovery via confidence-based ensemble, ACM Transactions on Multimedia Computing, Communications, and Applications, 1(2), 168-189, 2005.

[11]. K.Goh, E. Chang and B. Li, Using on-class and two-class SVMs for multiclass image annotation, IEEE Trans. on Knowledge and Data Engineering, 17(10), 1333-1346, 2005.

[12]. J. Fan, Y. Gao, and H. Luo, Multi-level annotation of natural scenes using dominant image components and semantic concepts," in Proc. of ACM MM, 540-547, 2004.

[13]. S. L. Feng, V. Lavrenko and R. Manmatha. Multiple Bernoulli Relevance Models for Image and Video Annotation. In Proc. of CVPR04, 2004.

[14]. R. Jin, J. Y. Chai, and L. Si. Effective Automatic image annotation via a coherent language model and active learning. In Proc. of ACM MM'04, 2004.

[15]. F. Kang, R. Jin, and J. Y. Chai. Regularizing Translation Models for Better Automatic Image Annotation. In Proc. of CIKM'04, 2004.

[16]. F. Monay and D. Gatica-Perez. On image auto-annotation with latent space models. In Proc. of ACM MM'03. Conf. on Multimedia,. 2003.

[17]. F. Monay and D. Gatica-Perez. PLSA-based image auto-annotation: Constraining the latent space. In Proc. ACM Int. Conf. on Multimedia, New York, Oct. 2004.

[18]. R. Zhang, Z. Zhang, M. Li, WY. M and HJ. Zhang. A probabilistic semantic model for image annotation and multi-modal image retrieval. Multimedia Systems, 12(1), 27-33,2006.

[19]. R. Schapire, Y. Singer, Boostexter: A boosting-based system for text categorization, Machine Learning 39, 135-168, 2000.

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

[21]. Dacheng Tao, Xiaoou, Tang, Xuelong Li and Xindong Wu, Asymmetric Bagging and Random Subspace for Support Vector Machines-based Relevance Feedback in Image Retrieval, IEEE trans on PRMI, 28(7), 1088-1099, 2006.

[22]. Xinjing Wang, Lei Zhang, Feng Jing, Wei-Ying Ma, AnnoSearch: Image Auto-Annotation by Search. In Proc. of CVPR'06.

[23]. Cees G. M. Snoek, Marcel Worring, Jan C. van Gemert, Jan-Mark Geusebroek, and Arnold W. M. Smeulders. The challenge problem for automated detection of 101 semantic concepts in multimedia. In Proc. Of ACM MM'06, 421–430, 2006.

[24]. Yan Song, Guo-Jun Qi, Xian-Sheng Hua, Li-Rong Dai, Ren-Hua Wang. Video Annotation by Active Learning and Semi-Supervised Ensembling. In Proc. of ICME'06, 933-936, 2006.

[25]. H. Feng, T.-S. Chua. A bootstrapping approach to annotating large image collection. MIR 2003. 55-62, 2003.

[26]. G. Carneiro, A. B. Chan, P.J. Moreno, and N. Vasconcelos, Supervised Learning of Semantic Classes for Image Annotation and Retrieval, IEEE trans on PAMI, 29(3), 394-410, 2007.

[27] Q. Tian, J. Yu, Q. Xue, and N. Sebe, A New Analysis of the Value of Unlabeled Data in Semi-Supervised Learning for Image Retrieval, IEEE Conf. ICME'2004, 2004.

[28] Z.-H. Zhou, K.-J. Chen, and Y. Jiang. Exploiting unlabeled data in content-based image retrieval. In Proc. 15th ECML, 2004.

[29] Baram, Y., El Yaniv, R. & Luz, K. Online Choice of Active Learning Algorithms. JMLR 5:255-291. 2004.

[30] Zhu, X., Lafferty, J. Semi-supervised learning using Gaussian fields and harmonic functions. In Proc. of ICML 2003.