Active Learning on Sparse Graph for Image Annotation

Minxian Li[1], Jinhui Tang[1,2] and Chunxia Zhao[1]
[1] School of Computer Science, Nanjing University of Science and Technology
Nanjing, China
[2] State Key Lab. for Novel Software Technology, Nanjing University
Nanjing, China
[e-mail: minxianli.njust@gmail.com, jinhuitang@mail.njust.edu.cn, zhaochx@mail.njust.edu.cn, jinhuitang@mail.njust.edu.cn]
*Corresponding author: Jinhui Tang

Received August 22, 2012; revised September 18, 2012; accepted September 25, 2012; published October 29, 2012

Abstract

Due to the semantic gap issue, the performance of automatic image annotation is still far from satisfactory. Active learning approaches provide a possible solution to cope with this problem by selecting most effective samples to ask users to label for training. One of the key research points in active learning is how to select the most effective samples. In this paper, we propose a novel active learning approach based on sparse graph. Comparing with the existing active learning approaches, the proposed method selects the samples based on two criteria: *uncertainty* and *representativeness*. The *representativeness* indicates the contribution of a sample's label propagating to the other samples, while the existing approaches did not take the *representativeness* into consideration. Extensive experiments show that bringing the representativeness criterion into the sample selection process can significantly improve the active learning effectiveness.

Keywords: image annotation, active learning, sparse graph

This research is supported by NSFC of China (Grant No:90820306, No: 61103059 and No: 61173104).

.

http://dx.doi.org/10.3837/tiis.2012.10.012

1. Introduction

In recent years there are some considerable research interests for automatic image annotation [1][2][3]. However, due to the well-known semantic gap problem, the performance of purely automatic image annotation techniques is still far from satisfactory. User interactions provide a possible solution to solve this problem. In order to fully utilize the human effort, active learning aims to actively select the most effective samples to present to the users for feedback [4].

A typical active learning framework is composed of two components, that is, a learning engine and a sample selection engine. **Fig. 1** shows the work procedure of the active learning, which works in an iterative way. In each round, the learning engine trains a model to predict the labels of unlabeled samples based on the the current training set.Then, the sample selection engine selects the most effective unlabeled samples based on a certain strategy for user feedback, i.e., manual labeling. And then, these samples are added to the training set for the next round of learning. The basic goal of the sample selection engine is to select the samples that are more useful than gahtered by random sampling.

The main challenge in active learning is how to select the most effective samples from the unlabeled data. Thus, several kinds of sample selection strategies were proposed. As discussed by Wang and Hua [2], these criteria can be classified into four categories.

- The most commonly used criterion is uncertainty, which means that the most uncertain samples should be selected. The information entropy is a typical measure that estimates uncertainty. The uncertainty criterion can be also viewed as a greedy strategy to reduce maximal expected risk without model updating. Due to its simplicity, this criterion has been widely applied [5][6][7].
- The second criterion is density, which selects samples in the regions of high density [8][9]. Wu et al. [8] defined a “representativeness” measure for each sample according to its distance to nearby samples, so this criterion also can be called as representativeness.
- The third criterion is diversity, which was first researched in batch mode active learning [10]. In many cases, instead of selecting just one sample, we need to select a batch of samples in an active learning iteration. Recent research shows that the selected samples in a batch or even all the labeled samples should be diverse [7][11].
- The last criterion is relevance, which means the degree of the selected image relevant to the query. It is usually used in multi-label image annotation and retrieval. In many applications, it is found that directly selecting the samples that have the highest probabilities to be relevant is more effective than using other criteria together [12][13].

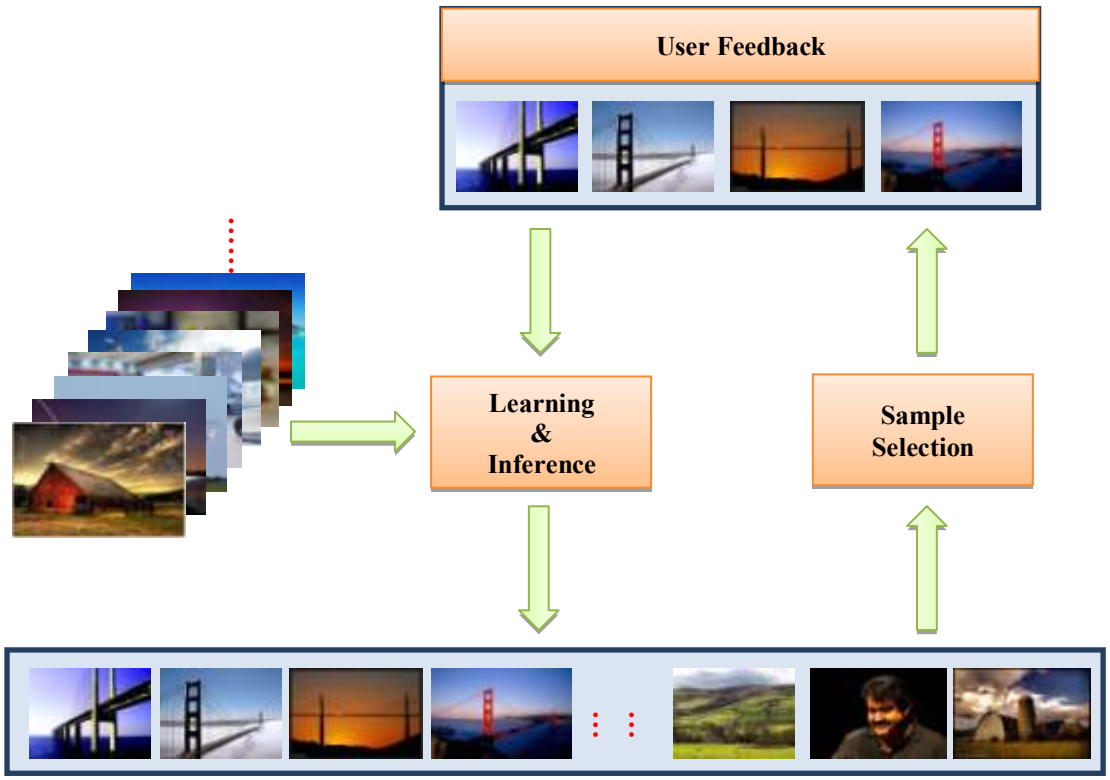
Meanwhile, graph-based learning methods are popularly applied to image annotation in recent years of strong performance [14][15][16]. And some active learning approaches based on graphs were proposed [17][18]. However, the computation cost of these methods is very large when the whole graph is huge. Therefore, active learning methods based on sparse graph has been proposed to in [19], which yielded impressive results. However, it does not consider the reconstruction contributions of each sample for other samples. The reconstruction contributions are indicated by the corresponding reconstruction coefficients, which are also used as the label propagation coefficients of this sample to others. Thus the reconstruction contributions should be a key point affecting the active sample selection performance. This motivates us to propose new criterion *representativeness* for sparse graph-based active learning, where *representativeness* indicates the total reconstruction contribution of each

sample for other samples.

In this paper, we propose a novel active learning approach based on *k*NN-sparse graph, which brings the *representativeness* criterion into the sample selection process. The main contributions of our work are follows:

- We sparsely reconstruct each sample from its *k* nearest neighbors in feature space instead of using all the other samples to improve the efficiency while maintaining its effectiveness. The reason is that, the one-vs-all reconstruction is very time-consuming and needs too much memory.
- Extensive experiments show that bringing the *representativeness* criterion into the sample selection process can significantly improve the active learning effectiveness.

The rest of this paper is organized as follows. In Section 2, we detail the proposed active learning approach by bringing the *representativeness* criterion into the sample selection process. Section 3 demonstrates the experimental evaluation on a real world image set. Finally, we provide some concluding remarks and suggestions for future work in Section 4.



2. Related Work

Actually active learning contains three different models: pool-based active learning [20], membership query learning [21], and stream-based active learning [22]. In this paper, we

focus on pool-based active learning, because this model allows selecting samples from an existing pool for labeling according to certain criteria.

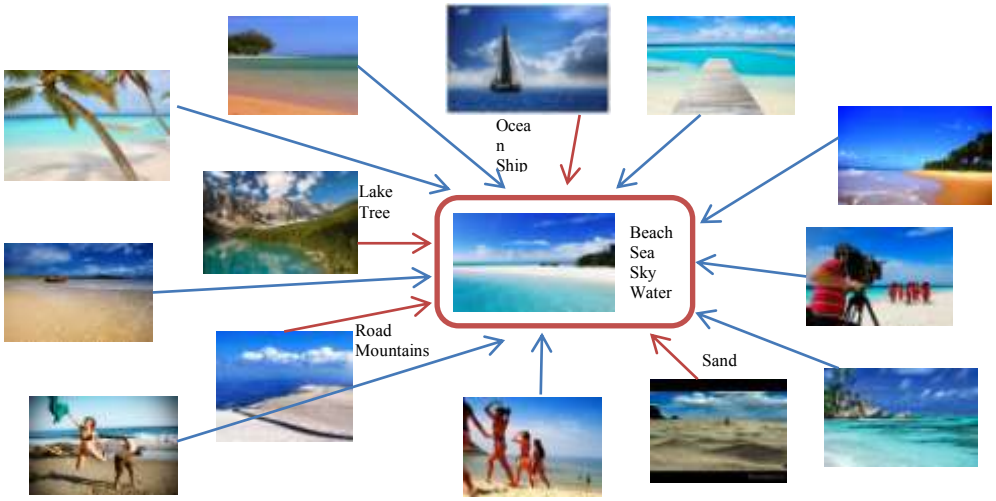
In practice most active learning methods adopt an uncertainty criterion, that is, to select the samples closest to the classification boundary [5][23], due to its simplicity. However, the experimental results [24][25] show that, the selected unlabeled data using uncertainty sampling cannot provide much help to the learner. Uncertainty sampling often fails by selecting the outliers, which indicates the high uncertainty samples. Some studies attempted to solve this problem. [20] and [24] proposed a method that directly optimizes reduced risk on future test examples. But, due to the high computational cost for selecting the most informative example from a large unlabeled pool, their methods are almost intractable in practical applications. [25] used a sampling scheme, namely "most uncertain per cluster", in which the learner selects the samples with the highest uncertain score from each cluster, and uses the density to weigh the selected samples. Actually, for uncertainty sampling, the scheme of using the most uncertain example per cluster is still unable to solve the outlier problem. [26] proposed to select samples based on informativeness, diversity and density criteria. In their work, the density of an unlabeled example is evaluated within a cluster, and multiple criteria are linearly combined with different factors. Nevertheless, in this approach, different values of the factors are associated with various applications, so it is difficult to determine those factors automatically.

Recently, graph-based methods have achieved much success in the applications of image and video analysis including image annotation. One of the typical works was developed by [27], which supported the keywords propagation from the labeled images to the unlabeled images by visual similarities for image retrieval. It is domain independent and the parameters are easy to tune, but this model directly adopted the pairwise relations over images to construct a *k*-NN similarity graph but did not consider the correlations between words.

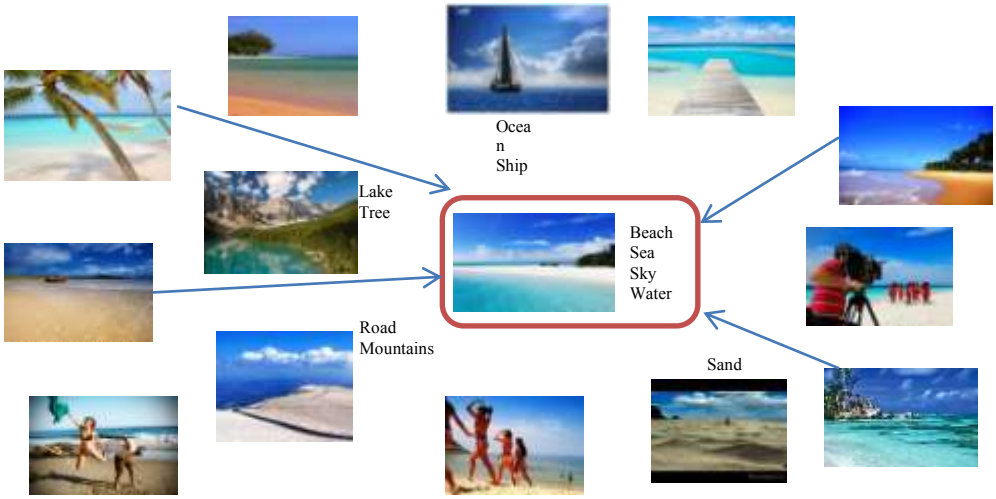
Image annotation, a solution to solve the insufficiency of CBIR systems, aims at automatically annotating image with some keywords. Machine learning techniques are adopted to develop the image annotation systems to map the visual features to semantic concepts. Ghoshal et al. [28] use a hidden Markov model for image annotation. Various model parameters and parameter estimations are examined to form the best image annotation model. Moreover, Wang et al. [29][30] adopt image annotation technique to video annotation. In addition, image annotation is also used to some special domain [31][32].

3. Active Learning Based On Sparse Graph Reconstruction

In this section, we introduce a novel active learning algorithm by bringing the *representativeness* criterion into the sample selection process based on *k*NN-sparse graph reconstruction.



(a) similarity-based graph**Fig. 1.** An overview of active learning framework



(b) sparse graph

Fig. 2. The exemplary comparison of conventional similarity-based graph and sparse graph

3.1 kNN-Sparse Graph Reconstruction

Most traditional graph-based learning algorithms construct the graphs only according to the visual distance, thus are very sensitive to the noise in visual features. Besides, constructing the graph only based on the visual distance will bring in semantically-unrelated links between data due to the semantic gap. A substitute way to construct a graph is to reconstruct each image from the other images by locally linear embedding [33] and linear neighborhood propagating [34]. However, they still cannot handle the semantically-unrelated links.

In neural science, it is reported that the human vision system searches for a sparse representation for the incoming image by using a few words in a feature vocabulary [35]. [36] demonstrated that the l_1-norm based linear reconstruction error minimization can naturally

lead to a sparse representation for the images. The sparse reconstruction is robust to the noise in features, and shows to enforce the images selected to reconstruct the test image are semantically-related to the test image [16]. This inspires us to construct the so-called sparse graph through the sparse reconstructions of the samples. Since the one-vs-all reconstruction is memory-consuming and time-consuming for large-scale computing, we choose to sparsely reconstruct each sample from its *k* nearest neighbors in feature space instead [16].

Fig. 2 shows an exemplary comparison of sparse graph and traditional similarity-based graph. In the similarity-based graph, there exists link between each pair of samples, and the weight is in inverse proportional to the distance measured in visual space. Therefore the information may be propagated between semantically unrelated samples. In the constructed sparse graph, only a few most probably semantically-related samples are selected to create links to the reference sample. Thus the sparse graph can remove most of those semantically-unrelated links between images to avoid propagation of incorrect information.

The recent study [37] shows that if the solution is sparse enough, the sparse representation can be recovered by convex l_1-norm optimization. Assume that we have an under-determined system of linear equations: $\boldsymbol{x} = \boldsymbol{D}\boldsymbol{w}$, where $\boldsymbol{x} \in \mathbb{R}^d$ is the feature vector of the image which needs be reconstructed, $\boldsymbol{D} \in \mathbb{R}^{d \times n} (d < n)$ is a matrix made by the feature vectors of the other images in the dataset, and $\boldsymbol{w} \in \mathbb{R}^{\mathrm{n}}$ is the vector of the unknown reconstruction coefficients. We can obtain the sparse solution for $\boldsymbol{w}$ by solving the following convex minimization problem [37]:

$$min_{\boldsymbol{w}} \parallel \boldsymbol{w} \parallel_1, \qquad s.t.\ \boldsymbol{x} = \boldsymbol{D}\boldsymbol{w} \tag{1}$$

This optimization task can be transformed into a general linear programming problem, since it is convex. There exists a globally optimal solution, which can be solved efficiently using many available l_1-norm optimization toolboxes, i.e. [38], which may convert the original constrained optimization problem into an unconstrained one. Meanwhile, practically an extra regularization coefficient can be tuned for optimality, but essentially does not exist in original problem formulation.

Let $\boldsymbol{\mathcal{X}} = \{\boldsymbol{x}_1, \dots \boldsymbol{x}_l, \boldsymbol{x}_{l+1}, \dots, \boldsymbol{x}_N\} \in \mathbb{R}^d$ be the set of feature vectors for the N images in the dataset, where $\boldsymbol{x}_i$ represents the i-th sample in the dataset, and $\boldsymbol{\mathcal{G}} = \{\boldsymbol{\mathcal{X}}, \mathbf{W} = \{w_{ij}\}\}$ be the sparse graph with the samples in set $\boldsymbol{\mathcal{X}}$ as graph vertices and $\mathbf{W}$ as the weight edges. Akin to [16], we summarize the construction of the *k*NN-sparse graph as follows:

1. Leverage approximate method [39] to search for k nearest neighbors $\mathcal{N}(\boldsymbol{x}_i)$ of each sample $\boldsymbol{x}_i$.
2. Form the matrix $\boldsymbol{D}_i$ with all samples $\boldsymbol{x}_{i_p} \in \mathcal{N}(\boldsymbol{x}_i)$: $\boldsymbol{D}_i = [\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \dots, \boldsymbol{x}_{ik}] \in \mathbb{R}^{d \times k}$, where $p \in \{1,2, \dots, k\}$ and $i_p \in \{1,2, \dots, N\}$. By solving the following l_1 -norm optimization problem, the vector of the reconstruction coefficients for $\mathrm{w_i}$ can be obtained:

$$min_{\boldsymbol{w}_i} \parallel \boldsymbol{w}_i \parallel_1, \qquad s.t.\ \boldsymbol{x}_i = \boldsymbol{D}_i \boldsymbol{w}_i \tag{2}$$

   where $\boldsymbol{w}_i \in \mathbb{R}^k$. Note that if we set $\mathcal{N}(\boldsymbol{x}_i) = \{\boldsymbol{x}_1, \dots \boldsymbol{x}_i, \boldsymbol{x}_{i+1}, \dots, \boldsymbol{x}_N\}$, then it becomes the one-vs-all sparse reconstruction and $k = N - 1$.
3. Set the edge weight w_{ij} from the data x_j to the data x_i as:

$$w_{ij} = \begin{cases} w_i(p), & if\ \boldsymbol{x}_j \in \mathcal{N}(\boldsymbol{x}_i)\ and\ j = i_p \\ 0, & if\ \boldsymbol{x}_j \notin \mathcal{N}(\boldsymbol{x}_i) \end{cases} \tag{3}$$

where $i, j \in \{1,2,\dots,N\}$, and $w_i(p)$ denotes the p-th element of vector $\boldsymbol{w}_i$

3.2 Sample Selection

Here we reorder the samples in set $\boldsymbol{\mathcal{X}}$ and have $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{L}} \cup \boldsymbol{\mathcal{U}}$, where $\boldsymbol{\mathcal{L}} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_l\}$ contains the first l samples labeled as $y_i \in \{1,0\}$ for every concept, and $\boldsymbol{\mathcal{U}} = \{\boldsymbol{x}_{l+1}, \boldsymbol{x}_{l+2}, \dots, \boldsymbol{x}_N\}$ contains the unlabeled ones. The label "1" indicates that the sample is relevant to a certain concept and "0" otherwise. And $\mathbf{f}$ is the vector of the predicted labels of all samples, which can be split into two blocks as $\mathbf{f} = [\mathbf{f}_{\mathcal{L}}; \mathbf{f}_{\mathcal{U}}]$.

Based on the label reconstruction assumption, we assume that the labels of each sample can also be reconstructed from the other samples' labels using the same reconstruction coefficients. Thus we can infer the labels of the unlabeled samples by minimizing the label reconstruction error as follows:

$$\min_{\mathbf{f}} \sum_{i=1}^{N} \left\| f_i - \sum_{j \neq i} w_{ij} f_j \right\|^2, \quad s.t.\ f_i = y_i, if\ \boldsymbol{x}_i \in \mathcal{L} \tag{4}$$

This formulation can be represented in matrix form as:

$$\min_{\mathbf{f}} [(\mathbf{I} - \mathbf{W})\mathbf{f}]^{\mathrm{T}} [(\mathbf{I} - \mathbf{W})\mathbf{f}], \quad s.t.\ \mathbf{f}_{\mathcal{L}} = \boldsymbol{y} \tag{5}$$

where $\mathbf{y} = \{y_1, y_2, \dots, y_l\}$ is the label vector for the first l sample, $\mathbf{f}$ is the vector of the predicted labels of all samples. Let $\mathbf{C} = (\mathbf{I} - \mathbf{W})^{\mathbf{T}}(\mathbf{I} - \mathbf{W})$ and differentiate the right side of Eqn.(5) with respect to $\mathbf{f}$, we obtain:

$$\left(\mathbf{C} + \mathbf{C}^{\mathrm{T}}\right)\mathbf{f} = \mathbf{M}\mathbf{f} = \mathbf{0} \tag{6}$$

where $\mathbf{M} = \mathbf{C} + \mathbf{C}^{\mathrm{T}}$ is a symmetric matrix. To compute the solution explicitly in terms of matrix operations, the matrix $\mathbf{M}$ can be split into four blocks after the l-th row and column:

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{\mathcal{L}\mathcal{L}} & \mathbf{M}_{\mathcal{L}\mathcal{U}} \\ \mathbf{M}_{\mathcal{U}\mathcal{L}} & \mathbf{M}_{\mathcal{U}\mathcal{U}} \end{bmatrix} \tag{7}$$

Thus, the solution is given by:

$$\mathbf{f}_{\mathcal{U}} = -\mathbf{M}_{\mathcal{U}\mathcal{U}}^{-\mathbf{1}} \mathbf{M}_{\mathcal{U}\mathcal{L}} \boldsymbol{y} \tag{8}$$

Based on the most popularly used *uncertainty* criterion, we prefer to select the most informative samples to present to the user for labeling. After the label propagation, each real-value label of the unlabeled data can be seen as the relevance score of the data to a certain concept. To convert the relevance score to the probability, we leverage the logic regression [40] as follow:

$$p_i = \frac{e^{a(\mathbf{f}_{\mathcal{U}}^i - b)}}{1 + e^{a(\mathbf{f}_{\mathcal{U}}^i - b)}} \tag{9}$$

where $\mathbf{f}_{ul}^{i}$ is the relevance score of the data $\boldsymbol{x}_i$, a and b are the coefficients. If the probability p_i is closer to 0.5, the data x_i is more uncertain to be classified, so is more informative. The information I_i of each unlabeled data x_i can be written as:

$$I_i = -p_i * \log(p_i) - p_i * \log(1 - p_i) \tag{10}$$

As aforementioned, besides the uncertainty, we should also take the reconstruction contribution of each sample into account for the sparse graph-based active learning. In the sparse reconstruction procedure, the contribution of each unlabeled sample to reconstruct the other unlabeled samples is different, which naturally depends on the reconstruction coefficient. The bigger the reconstruction coefficient is, the more contribution the unlabeled sample has. That is, these unlabeled data is more representative. Therefore, we define the *representativeness* score r_i of each unlabeled sample as follow:

$$r_i = \sum_{j=l+1}^{N} w_{ij} \tag{11}$$

where w_{ij} is the coefficient that x_j reconstructs x_i in Section 3.1.
To this end, we combine the *uncertainty* and *representativeness* criterions to select the most effective sample as follows:

$$x_i^* = argmin_x \{\alpha \times r_i + (1 - \alpha) \times I_i\} \tag{12}$$

where α belongs to [0,1].

4. Experimental Results and Analysis

To demonstrate the effectiveness of the proposed *representativeness* criterion, we conduct the experiments on a real-world data set to compare different sample selection criterion quantitatively:

- *Uncertainty*: the samples are selected according to the uncertainty only.
- *Representativeness*: the samples are selected according to the representativeness only.
- *Representativeness* + *Uncertainty*: the samples are selected according to both the uncertainty and representativeness.

4.1 Experimental Settings

The experiments are conducted on a large-scale real-world data set NUS-WIDE-Lite [41] on 81 concepts. The data set is divided into two parts: training data, which contains 27807 images, and testing data, which contains 27808 images. The basic model is trained from the training data, and the new training samples are actively selected from the testing data. For each image, we extract 426 dimensional features, including 225 dimensional 5×5 block-based color moments features, 73 dimensional edge direction histogram features, and 128 dimensional wavelet texture features.

There are four parameters in our algorithm, including the number of nearest neighbors k, the parameters of logic regression a and b in Eqn. (9), and the weight parameter α in Eqn. (12). These four parameters are empirically set to 300, 15, -0.2, and 0.95.

4.2 Experimental Results

For each concept, we use three different sample selection criterions to select 1000 images. **Fig. 3** illustrates the comparison of the average precisions (APs) [42] of 81 concepts based on the three different criterions. We can see that the performance of combining *representativeness* and *uncertainty* criterions is better than using *uncertainty* criterion and much better than using *representativeness* criterion. It is worth to mention that only using *representativeness* criterion is not very effective. The reason is that the uncertain of the label of the unlabeled sample is not considered. However, the result of the experiment shows that the combining use of *representativeness* criterion and *uncertainty* criterion increases the APs. It is more effective than using them separately.

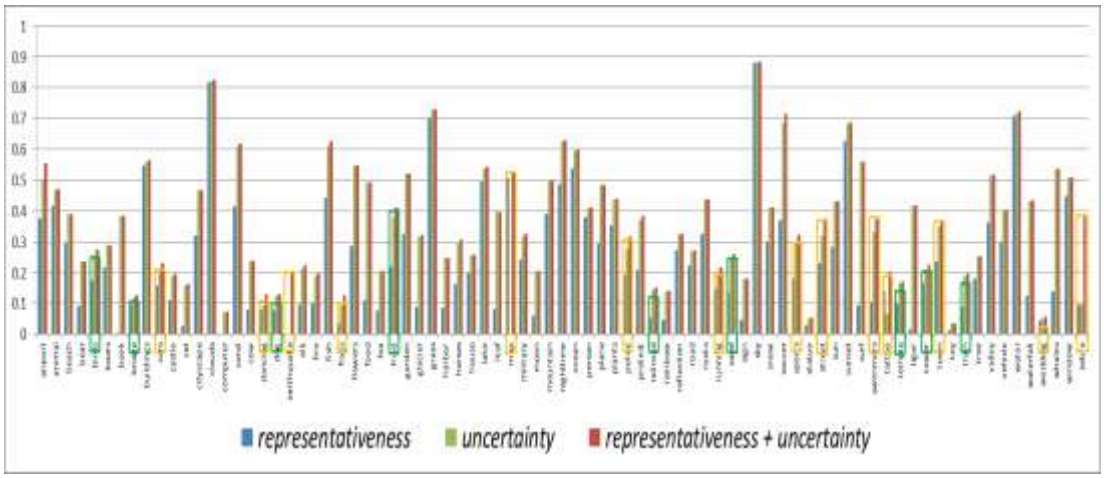


Fig. 3. Comparisons of average precisions for the 81 concepts, based on *uncertainty*, *representativeness* and *representativeness + uncertainty*.

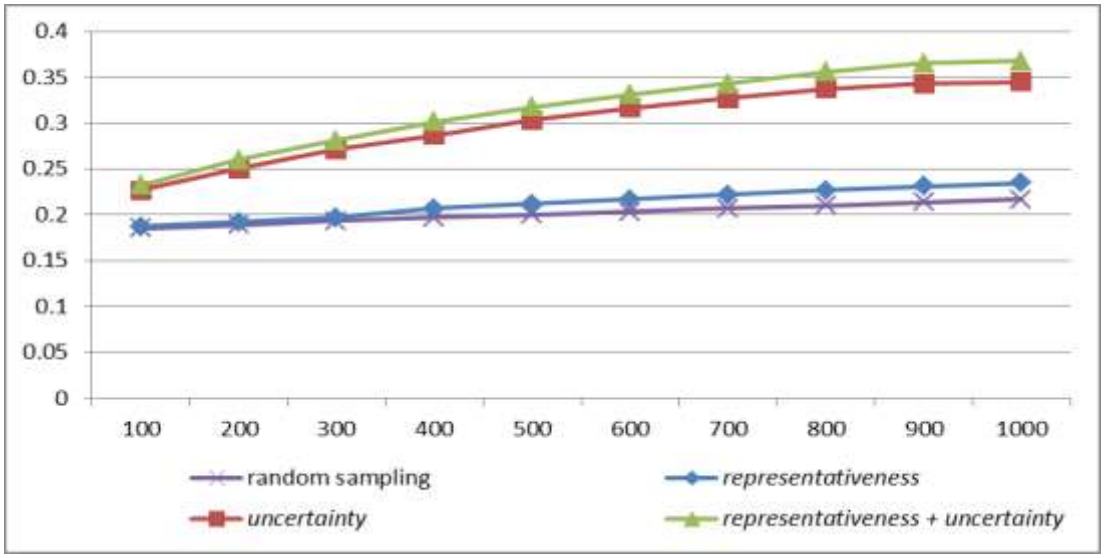


Fig. 4. Comparisons of MAP from three different principles and random sampling. The *x* axis indicates the number of sampled images, and the *y* axis indicates the MAP.

Compared to the *uncertainty* criterion, the APs of 78 concepts based on the combination principle improve; the improvement of APs of 24 concepts based on the combination principle exceeds 5%, which are marked by the green rectangles and the orange rectangles; the improvement of APs of 15 concepts based on the combination principle exceeds 10%, which are marked by the orange rectangles.

We select 1000 images for ten rounds with 100 sampled images for each round. **Fig. 4** compares the average MAP of each round obtained from the aforementioned three principles. Obviously, the sampling strategy by using active learning is better than random sampling. Moreover, we can see that by combining *representativeness* and *uncertainty* principles into the sample selection strategy, the performance of active learning significantly improves. For example, using 1000 sampled images, the selection strategy based on the combination principle of *representativeness* and *uncertainty* achieves a mean average precision (MAP) of 0.3678, which has improvements of 56.82% compared with *representativeness* principle, and 6.59% compared with *uncertainty* principle.

5. Conclusion

In this paper, we have presented a novel active learning method based on sparse graph reconstruction, which brings the *representativeness* criterion into the sample selection process, to select the most effective data. Given the labeled and unlabeled data, the *k*NN-sparse graph is proposed to reconstruct the whole data set. The most effective data are selected combining two principles: those are representative to reconstruct the whole data set and are uncertain to be classified. Extensive experiments have conducted on a real-world image set. The results show that bringing the representativeness criterion into the sample selection process can significantly improve the active learning effectiveness. In the future work, we will integrate the other criteria into the active learning framework to further improve the performance.

References

[1] S. Zhu, Y. Liu, "Semi-supervised learning model based efficient image annotation," *IEEE Signal Processing Letters*, vol. 16, no. 4, pp. 989-992, Nov 2009. Article (CrossRef Link)

[2] M. Wang and X.-S. Hua, "Active learning in multimedia annotation and retrieval: a survey," *ACM Transaction on Intelligent Systems and Technology*, vol. 2, no. 2, Feb 2011. Article (CrossRef Link)

[3] X.J. Wang, L. Zhang, X. Li, and W.Y. Ma, "Annotating images by mining image search results," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 11, pp. 1919-1932, Nov 2008. Article (CrossRef Link)

[4] T. Huang, C. Dagli, S. Rajaram, E. Chang, M. Mandel, G. Poliner, and D. Ellis, "Active learning for interactive multimedia retrieval," *In Proceedings of the IEEE*, vol. 96, no. 4, pp. 648-667, Apr 2008. Article (CrossRef Link)

[5] S. Tong and E. Chang, "Support vector machine active learning for image retrieval," in *Proc. of the 9th ACM Int. Conf. on Multimedia*, pp. 107-118, Sep 2001. Article (CrossRef Link)

[6] J. He, M.J. Li, H.J. Zhang, H. Tong, and C. Zhang, "Mean version space: a new active learning method for content-based image retrieval," in *Proc. of the 6th ACM SIGMM Int. Workshop on Multimedia Information Retrieval*, pp.15-22, Oct 2004. Article (CrossRef Link)

[7] A. Joshi, F. Porikli, and N. Papanikolopoulos, "Multi-class active learning for image classification," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 2372–2379, Jun 2009. Article (CrossRef Link)

[8] Y. Wu, I. Kozintsev, J. yves Bouguet, and C. Dulong, "Sampling strategies for active learning in personal photo retrieval," in *Proc. of IEEE Int. Conf. on Multimedia and Expo*, pp. 529-532, Jul 2006. Article (CrossRef Link)
[9] C. Zhang and T. Chen, "Annotating retrieval database with active learning," in *Proc. of IEEE Int. Conf. on Image Process*, vol. 3, no. 2, pp. 595-598, Sep 2003. Article (CrossRef Link)
[10] S. C. H. Hoi, R. Jin, J. Zhu, and M. R. Lyu, "Batch mode active learning and its application to medical image classification," in *Proc. of the 23rd Int. Conf. on Machine learning*, pp. 417-424, 2006. Article (CrossRef Link)
[11] C. Dagli, S. Rajaram, and T. Huang, "Leveraging active learning for relevance feedback using an information theoretic diversity measure," in *Proc. of Int. Conf. on Image Video Retrieval*, vol. 4071, pp. 123-132, 2006. Article (CrossRef Link)
[12] S. Ayache and G. Quénot, "Evaluation of active learning strategies for video indexing," *Image Communication*, vol. 22, no. 7, pp. 692-704, Aug 2007. Article (CrossRef Link)
[13] P. H. Gosselin and M. Cord, "A comparison of active classification methods for content-based image retrieval," in *Proc. of the* 1st *Int.Workshop on Computer Vision Meets Databases*, pp. 51-58, Jun 2004. Article (CrossRef Link)
[14] J. Tang, S. Yan, et al, "Inferring semantic concepts from community-contributed images and noisy tags," in *Proc. of the 17th ACM Int. Conf. on Multimedia*, pp. 223-232, Oct 2009. Article (CrossRef Link)
[15] J. Tang, H. Li, et al, "Image annotation by graph-based inference with integrated multiple/single instance representations," *IEEE Transactions on Multimedia*, vol. 12, no. 2, pp. 131-141, Feb 2010. Article (CrossRef Link)
[16] J. Tang, R. Hong, S. Yan, and et al, "Image annotation by knn-sparse graph-based label propagation over noisily-tagged web images," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 2, pp. 111-126, Feb 2011. Article (CrossRef Link)
[17] N. Cesa-Bianchi, C. Gentile, F. Vitale, and G. Zappella, "Active learning on trees and graphs," in *Proc. of the 23rd Conf. on Learning Theory*, pp. 320-332, Jun 2010. Article (CrossRef Link)
[18] X. Wang, W. Ma, L. Zhang, X. Li, "Multi-graph enabled active learning for multimodal web image retrieval," in *Proc. of the 7th ACM SIGMM Int.Workshop on Multimedia Information Retrieval*, pp. 65–72, Aug 2005. Article (CrossRef Link)
[19] J. Tang, Z. Zha, D. Tao, and T. S. Chua, "Semantic-gap-oriented active learning for multilabel image annotation," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp.2354-2360, Apr 2012. Article (CrossRef Link)
[20] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, "Active learning with statistical models," *Journal of Artificial Intelligence Research*, vol. 4, no.1, pp. 129-145, January, 1996. Article (CrossRef Link)
[21] R. D. King, K. E. Whelan, F. M. Jones, P. G. Reiser, C. H. Bryant, S. H. Muggleton, D. B. Kell, and S. G. Oliver, "Functional genomic hypothesis generation and experimentation by a robot scientist," *Nature*, vol. 427, no. 6971, pp. 247-252, Jan 2004. Article (CrossRef Link)
[22] D. Cohn, L. Atlas , and R. Ladner, "Improving generalization with active learning," *Machine Learning*, vol. 15, no. 2, pp. 201-221, May, 1994. Article (CrossRef Link)
[23] S. Tong, and D. Koller, "Support vector machine active learning with applications to text classification," *Journal of Machine Learning Research*, vol. 2, pp.45-66, Mar 2002. Article (CrossRef Link)
[24] N. Roy and A. McCallum, "Toward optimal active learning through sampling estimation of error reduction," in *Proc. of 18th Int. Conf. Machine Learning*, pp. 441-448, 2001. Article (CrossRef Link)
[25] M. Tang, X. Luo, and S. Roukos, "Active learning for statistical natural language parsing," in *Proc. of 40th Annu. Meeting Assoc. Comput. Linguist.*, pp. 120-127, 2002. Article (CrossRef Link)
[26] D. Shen, J. Zhang, J. Su, G. Zhou, and C. -L. Tan, "Multi-criteria-based active learning for named entity recognition," in *Proc. of 42nd Annu. Meeting Assoc. Comput. Linguist.*, pp. 589-596, 2004. Article (CrossRef Link)

[27] H. Tong, J. He, M. Li, W. Ma, H. J. Zhang, C. Zhang, "Manifold-ranking based keyword propagation for image retrieval," *EURASIP Journal on Applied Signal Processing*, vol. 21, pp. 1-10, Jan 2006. Article (CrossRef Link)
[28] A. Ghoshal, P. Ircing, S. Khudanpur, "Hidden Markov models for automatic annotation and content-based retrieval of images and video," in *Proc.of the 28th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 544-551, 2005. Article (CrossRef Link)
[29] M. Wang, X. Hua, J. Tang, R. Hong, "Beyond distance measurement: constructing neighborhood similarity for video annotation," *IEEE trans. On Multimedia*, vol.11, no.3, 2009. Article (CrossRef Link)
[30] M. Wang, X. Hua, R. Hong, J. Tang, G. Qi, Y. Song, "Unified video annotation via multi-graph learning," *IEEE trans. On Circuit System and Video Technology*, vol.19, no.5, 2009. Article (CrossRef Link)
[31] J. Park, S. Park, Y. Shin and D. Jang, "A novel system for detecting adult images on the internet," *KSII Transaction on internet and information system*, vol. 4, no. 5, pp. 910-924, 2010. Article (CrossRef Link)
[32] B. Wu and J. Juang, "Real-time vehicle detector with dynamic segmentation and rule-based tracking reasoning for complex traffic conditions," *KSII Transaction on internet and information system*, vol. 5, no. 12, pp. 2355-2373, 2011. Article (CrossRef Link)
[33] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, 2323-2326, 2000. Article (CrossRef Link)
[34] F. Wang and C. Zhang, "Label propagation through linear neighborhoods," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 1, pp. 55-67, 2008. Article (CrossRef Link)
[35] R. Rao and B. Olshausen and M. Lewicki, *Probabilistic models of the brain: perception and neural function*, MIT Press, 2002. Article (CrossRef Link)
[36] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210-227, February, 2009. Article (CrossRef Link)
[37] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics*, vol. 59, no. 6, pp. 797-829, 2006. Article (CrossRef Link)
[38] *l*1-magic. http://www.acm.caltech.edu/l1magic/. Article (CrossRef Link)
[39] D. Mount and S. Araya, "Ann: a library for approximate nearest neighbor searching," *in CGC 2nd Annual Fall Workshop on Computational Geometry*, 1997. Article (CrossRef Link)
[40] D.D. Lewis and W.A. Gale, "A sequential algorithm for training text classifiers," in *Proc.of the 17th annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, pp. 3-12, July 03-06, 1994. Article (CrossRef Link)
[41] T.-S. Chua and J. Tang et al, "NUS-WIDE: a real-world web image database from National University of Singapore," in *Proc.of ACM Int. Conf. on Image and Video Retrieval*, Jul 2009. Article (CrossRef Link)
[42] Trec-10 proceedings appendix on common evaluation measures. Article (CrossRef Link)

Minxian Li received the B.E. degree in the Department of Computer Science and Technology, in the HuaiHai Institute of Technology, Lianyungang, China, in 2006. Now he is a Ph.D. student in the Department of Computer Science and Technology, in Nanjing University of Science and Technology, Nanjing, China. His current research focuses on social multimedia analysis and retrieval.

Jinhui Tang received the B.E. and Ph.D. degrees from University of Science and Technology of China, Hefei, China, in 2003 and 2008, respectively, both in the Department of Electronic Engineering and Information Science. Since July 2008, he has been a Research Fellow in the School of Computing, National University of Singapore. Now he is a professor in the Department of Computer Science, Nanjing University of Science and Technology. His current research interests include content-based image retrieval, video content analysis, and pattern recognition. Prof. Tang is a member of the Association for Computing Machinery. He is a recipient of the 2008 President Scholarship of Chinese Academy of Science, and a co-recipient of the Best Paper Award in ACM Multimedia 2007.

Chunxia Zhao received the B.E., M.S. and Ph.D. degrees from Harbin Institute of Technology, Harbin, China, in 1985, 1988, 1998, respectively, both in the Department of Electrical Engineering and Computer. She is a professor in the Department of Computer Science, Nanjing University of Science and Technology. She had a stay of one year as research assistant at the Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong from 1997 to 1998. Her current interests are in the areas of robots, computer vision, and pattern recognition.