Graph-based Semi-Supervised Learning Framework for Medical Image Retrieval

Menglin Wu[1,2], Quansen SUN[1], Jin Wang[3], Yuhui Zheng[3]

[1]School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094
[2]College of Electronics and Information Engineering, Nanjing University of Technology, Nanjing 211816
[3]Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science & Technology, Nanjing, 210044
wumenglin@njut.edu.cn; qssun@126.com; jinwang@nuist.edu.cn; zhengyh@vip.126.com

Abstract: As low level features can not reflect the high level semantic in medical image search, in this paper, we propose an image retrieval algorithm to combine visual concept and local features by graph-based semi-supervised learning framework. More specific, we construct a graph model by distance between images, and add density similarity measure in the label propagation progress to get the membership degree of query images which is called visual semantic representation. Meanwhile, the dense SIFT feature of the image blocks is extracted and described with bag of visual words, in order to get the local feature. Besides, we design a combination of visual concept and local feature strategy for similarity measurement. Experimental results of ImageCLEFmed dataset demonstrate that the proposed algorithm represents the visual semantic of images effectively, and achieves a better retrieval performance than single low level feature.

Keywords: content-based medical image retrieval; graph-based semi-supervised learning; visual semantic; bag of words

1 Introduction

With the development of digit imaging technology in medical domain, hospitals and Medical research institutions produce a large amount of digit images every day. Content-based medical image retrieval (CBMIR) is an important tool for physicians in Clinical diagnosis and medical research [1]. In CBMIR system, feature vectors extracted by low level features, is a basis of similarity measurement in searching procedure. Some methods use global features for retrieval, in [2], Gabor feature was extracted for mammogram retrieval. Besides, other approaches are based on local features. Greenspan et al. [3] proposed a Gaussian mixture model-Kullback Leibler (GMM-KL) framework for matching and categorizing x-ray images by body regions feature. Lehman et al. [4] conducted a comparison of texture feature and multi-scale feature by various classifiers in medical image classification and retrieval. Different from direct low level feature extraction, Avni et al. [5] subdivided images into local blocks at multiple scales, and use bag of features to describe patch-based image content, while another algorithm in [6] improved the this image representation scheme via multiply assignment and visual words weighting of the patches. Fusion features are also applied to discriminate images of diverse classes. Tommasi et al. [7] adapted an integration of global and local features for medical image annotation. The local features were randomly sampled modified SIFT descriptors and the global features were downscaling raw pixels, then classification was done by support vector machine (SVM) by three alternative strategies. In spite of these methods demonstrate effective result in medical image retrieval, there is a semantic gap between low level features and high level semantic. In practice, radiographs with high variances may belong to same class, while others look similar but in different classes. For that reason, the low level features cannot be the complete description of image content.

Semi-supervised learning techniques, which attempt to leverage both labeled and unlabeled data, are applied in pattern classification. As a major family of semi-supervised learning, graph-based methods have attracted increasing research attention. Zhu et al. [8] introduce an approach based on Gaussian random fields and harmonic functions. The local and global consistency method is proposed in [9], which improved the energy function. Tang et al. [10] improved this algorithm and used for video annotation.

The major contributions of this paper are following: (1) introduce graph-based semi-supervised learning into medical image retrieval framework, and put sample point density similar in affinity matrix to enhance the effectiveness of visual concept extraction. (2) To solve the problem that low level features are failed to represent the image content, we apply graph model and label propagation method to obtain membership degree of the query image for semantic representation, and design a similarity measurement which combines local feature and visual semantic.

2 Visual Concept Extraction by Graph-based Semi-supervised Learning

2.1 Learning Framework

Given a point set $\chi = \{x_1, ..., x_l, x_{l+1}, ...x_n\}$ are n image samples in R^m feature space. The first l points of the sample set are labeled (training images in dataset for our application), while the rest ones are unlabeled (query images for our application). The goal of the learning method is to predict the label of unlabeled point by whole set χ. Let label set $L = \{1, ..., c\}$ be c concept labels, and the first labeled l samples are marked as $y_L = \{y_1^T, y_2^T, ..., y_L^T\}^T$ with $y_i \in R^c$. If a sample y_i has a concept j, $y_{ij} = 1$ and vice-versa. A $n \times c$ matrix $\mathrm{f} = (f_1^T, ..., f_l^T, ...f_n^T)^T$ denotes the membership degree of the each concept for whole sample set. As shown in equation (1), f can be split into two parts after the l-th row, then $\mathrm{f}_L = y_L$. f_U is the prediction of membership degree for the unlabeled samples, which we can treat as semantic similarity of each medical image class.

$$\mathrm{f} = \begin{pmatrix} \mathrm{f}_L \\ \mathrm{f}_U \end{pmatrix}$$

(1)

Construct an undirected connect graph $G = < V, E >$ with the vertex set $V = \chi$. $V = L \bigcup U$, where vertex set $L = \{1,...,l\}$ contains labeled sample points and another set $U = \{l+1,...n\}$ include unlabeled ones. The edge $w_{ij} \in E$ represents the relationship between point i and j:

$$w_{ij} = \exp(-\| x_i - x_j \|^2 / (2\sigma^2))$$

(2)

Where σ is bandwidth and x_i is the visual features of sample i. According to [8], the concept labels of the unlabeled points are inferred by minimizing the cost function in label propagation procedure:

$$Q(f) = \sum_{1 \le i \le j} w_{ij}(f_i - f_j) + \infty \sum_{i \in L}(f_i - y_i)$$

(3)

Membership degree prediction function is $f^* = \arg\min_f (Q(f))$. The left term in equation (3) describes the total variation of data labels with respect to neighborhood structure called smoothness term, and the right term called fit term present the invariant constrain of labeled data. Differentiating $Q(f)$ with respect to f, we have

$$\frac{\partial Q}{\partial f}\Big|_{\mathrm{f}=\mathrm{f}^*} = \mathrm{f}^* - S\mathrm{f}^* + \infty(\mathrm{f}_L^* - y_L) = 0$$

(4)

Here $S = D^{-1}W$ and W is affinity matrix with entry w_{ij}. $D = diag(d_i)$ is a diagonal matrix whose element is $d_i = \sum_{j=1}^{n} w_{ij}$, we can also split Matrix S after the l-th row and l-th column, So equation (4) can convert to:

$$\begin{cases} f_L = S_{LL} f_L + S_{LU} f_U \\ f_U = S_{UL} f_L + S_{UU} f_U \end{cases}$$

(5)

Finally, we can get the optimal solution f_U with constrain $f_L^* = y_L$:

$$f_U^* = (I - S_{UU})^{-1} S_{UL} y_L$$

(6)

2.2 Label Propagation with Density Similarity

Since the matrix S defined in (4) is a symmetric matrix, the label information is spread symmetrically, which means the neighborhood points have the same semantic label in feature space. But pairwise similarity measurement in (2) cannot effectively descript the complex structure of the true data set. According to a global consistency assumption, points on the same manifold might have same concept label. Therefore the label propagation in section 2.1 considers direct contribution from one sample to another through similarity weight, however, it ignores the structural influence.

In this paper, we embed structural similarity model into affinity matrix by sample points' density for improving the performance of our learning method and semantic extraction. Let p_i is the estimation density by Parzen window:

$$p_i = \frac{1}{N_i} \sum_{j=1}^{n} k(x_i - x_j)$$

(7)

Where N_i is the number of the sample x_i's neighborhoods, while $k(x)$ is kernel function satisfied $k(x) > 0$ and $\int k(x)dx = 1$, which we choose Gaussian kernel in this paper. The similarity between x_i and x_j is $\tilde{w}_{ij} = w_{ij} \times g_{ij}$, in which density difference g_{ij} is defined:

$$g_{ij} = \exp(-(p_i - p_j)^2 / 2\sigma_p^2)$$

(8)

σ_p is the bandwidth parameter controlling the significance of the influence. The formulation (8) indicates that the similarity of two samples not only get smaller by growing the distance of the feature space, but also by the increasing difference of density. Hence, S can be rewritten:

$$\tilde{S} = \tilde{D}^{-1} \tilde{W}$$

(9)

Here $\tilde{W}$ is a matrix with element $\tilde{w}_{ij}$, while diagonal matrix $\tilde{D} = diag(\tilde{d}_i)$ and $\tilde{d}_i = \sum_{j=1}^{n} \tilde{w}_{ij}$. Label propagation with density similarity enhances the label information spreading in same structure and suppresses it in different structure.

On the basis of anisotropic diffusion equation [10], formulation (6) equals to an iteration form:

$$\mathrm{f}_U(t+1) = \tilde{S}_{UL}\, \mathrm{f}_L(t) + \tilde{S}_{UU}\, \mathrm{f}_U(t)$$

(10)

t is the number of iterations. To sum up, the algorithm for solving the concept membership degree Matrix f_u of unlabeled sample (query images) is following:

1) Calculate the Euclidean distance w_{ij} and difference of density g_{ij} between arbitrary two samples by equation (2) and (8). In practice, we just need to consider k-nearest neighborhood to simplify calculation.

2) Calculate affinity matrix $\tilde{W}$, whose entry is $\tilde{w}_{ij} = w_{ij} \times g_{ij}$.

3) Construct $\tilde{S} = \tilde{D}^{-1} \tilde{W}$ according to (9)

4) Split Matrix $\tilde{S}$ into $\tilde{S}_{LL}$, $\tilde{S}_{LU}$, $\tilde{S}_{UL}$ and $\tilde{S}_{UU}$.

5) Iterate $\mathrm{f}_U(t+1) = \tilde{S}_{UL}\,\mathrm{f}_L(t) + \tilde{S}_{UU}\,\mathrm{f}_U(t)$ until convergence, and we get the concept annotations of unlabeled samples $\mathrm{f}_U^* = \{f_{l+1}^T, \ldots f_n^T\}^T$ and $f_i \in R^c (l < i \le n)$, which presents the concept membership degree of c classes.

3 Local Features Extraction

In this paper, we apply SIFT descriptor [11] to extract the local feature, which is robust to scale, rotation and illumination changes so that it can present detail features well. Due to low contrast of the medical images, we cannot use salient point detector directly. Unlike random sampling in [7], we utilize dense SIFT [12] by sampling regular grids of fixed size. Firstly, every image is downscaled to 512×512 and SIFT descriptors of 16×16 pixel patches are computed over a grid with 8 pixels spacing. Secondly, we built the vocabulary and create the visual words using unsupervised K-means clustering algorithm from these descriptors (K is equal to 500 in our paper). At final step, each patch of the image is assigned to the nearest word, and the frequency histogram of the visual words is subsequently normalized for “bag of features” representation of the image.

4 Similarity Measurement Combining Semantic and Local Features

In this section, we design a similarity distance metric based on visual concept in section 2 and local features in section 3:

$$sim(I, J) = \exp(-d_{JSD}(I, J) / 2\sigma_d^2) \times f_{ij}$$

(11)

Where I is a query image and J is a radiograph in dataset respectively. Since the image local features are presented as a kind of histogram, Jensen-Shannon divergence (JSD) is utilized to computer similarity between two visual words histogram:

$$d_{JSD}(I,J)=\sum_{m=1}^{M} I_m \log(\frac{2I_m}{I_m+J_m})+J_m \log(\frac{2J_m}{J_m+I_m})$$

(12)

In which I_m and J_m are m-th bin in histogram. The left multiplier in equation (11) indicates the local feature similarity of two images, which is decreasing with distance increasing, while right multiplier is the membership degree of I relative to the concept class of J . The bandwidth σ_d is a tradeoff between visual concept and local features. Experiments will show the influence of the retrieval performance when σ_d variances.

5 Experiments

To evaluate our proposed algorithm for medical image retrieval, we conduct experiments on ImageCLEFmed 2009 dataset [13] from Aachen University of Technology.

5.1 Results on ImageCLEFmed 2009 dataset

From this dataset, we choose 4471 images of 50 classes as sub set to ensure that the number of each class is approximately the same, while 1639 radiographs in original testing set are query images. A retrieval example is shown in the Fig. 1, where the top-left image is for querying and the rest of ones are searching result. The similarity distance is increasing from left to right and top to bottom.

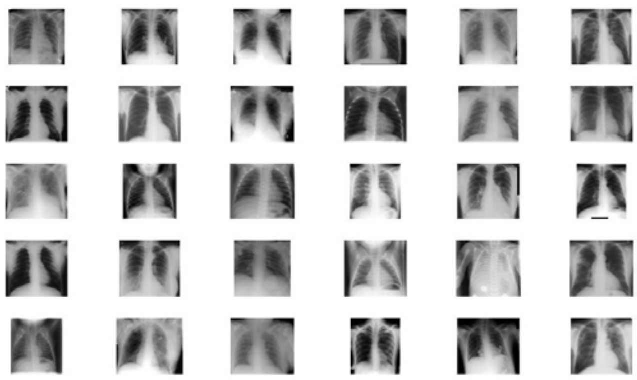

Fig1. An example of image retrieval

We compared the proposed retrieval framework with other methods in [2] and [6]. In paper [2], it extracted Gabor features and applied Euclidean distance for similarity measurement referred as Gabor+Euclidean, while patch-based visual words and Jensen-Shannon divergence used in [6] referred as BOW+JSD. Pension-Recall curve is utilized to test retrieval performance and results are illustrated in Fig. 2.

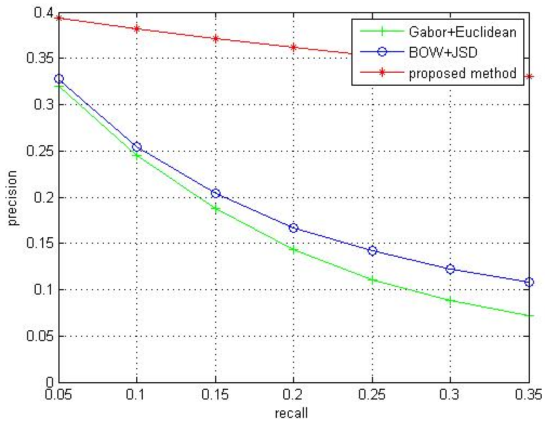

Fig2. Average precision-recall comparison between three algorithms

Our method is clearly superior to other two algorithms. More importantly, the precision of the proposed method is decreasing more slowly than other ones as recall number increasing. Therefore, our approach is getting better performance than single low level features. Besides, it also indicates that the local features obtain higher precision than global features for medical image retrieval.

We also create the Mean Average Precision (MAP) over all the 50 classes for evaluation. The MAP of three approaches is 0.374, 0.254, 0.26 respectively, when the number of return images is 30.

5.2 Comparison with other graph-based semi-supervised learning methods

Since the proposed label propagation algorithm is designed for semantic extraction, we test its retrieval performance by comparing with local and global consistency (LGC) method in [10]. Fig. 3 demonstrates the comparison of two methods on average precision of retrieval. To fairly evaluate our method, two approaches are employed the same feature extraction and similarity measurement described in this paper.

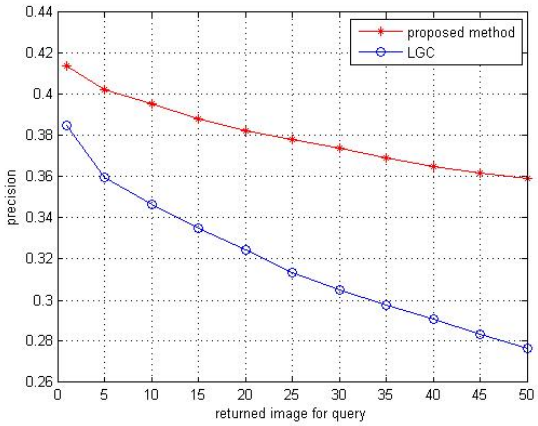

Fig3. Average retrieval precision between two algorithms

As a result, the method in our paper obtains higher accuracy than LGC. Since we follow the same steps in retrieval measurement, it is inferred that our approach benefits from the label propagation procedure in semantic extraction. Therefore, embedding structural assumption in label propagation, which means that the similarity between two samples rely on both local feature and sample density, can effectively compute the membership degree of the concept classes for query images.

5.3 Results in Different σ_d

In equation (11), bandwidth parameter σ_d is tradeoff between local features and visual concept for similarity measurement. The average precisions of different σ_d are shown in Fig. 4.

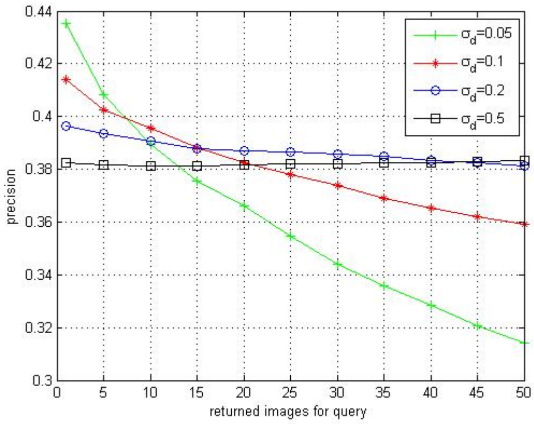

Fig4. Average retrieval precision between different σ_d

When $\sigma_d = 0.05$, initial accuracy is higher than other ones, but the retrieval performance drops down quickly as the number of return image increasing. On the contrary, we get lower accuracy in original, though the precision doesn't change too much. The figure infers that the initial precision is determined by local features while semantic similarity guarantees the stability of the retrieval performance. As aforementioned, we select 0.1 in this paper.

6 Conclusions and Future Work

We have demonstrated a novel medical image retrieval framework based on graph-based semi-supervised learning. Unlike the other retrieval method base on low level feature, the proposed algorithm first obtains visual concept by improved learning procedure, and then utilizes dense SIFT and "bag of features" method for local features description. A similarity measurement is also designed combining semantic and local features. Experiment results demonstrate the effectiveness of our method. In the future, we intend to explore the extensions of our algorithms, which allow radiologists to search medical images by interest regions.

References

1. Müller H, Michoux N, Bandon D, et al.: A review of content-based image retrieval systems in medical applications-clinical benefits and future directions. Medical Informatics, vol. 73, no. 1, pp. 1--23(2004).
2. C.-H. Wei, Y. Li, and C.-T. Li.: Effective extraction of Gabor features for adaptive mammogram retrieval. In: IEEE International Conference on Multimedia & Expo 2007 (ICME2007), pp. 1053--1056(2007)

3. Greenspan H, Pinhas A T.: Medical image categorization and retrieval for PACS using the GMM-KL framework. IEEE Transactions on Information Technology in Biomedicine, vol. 11, no. 2, pp. 190--202(2007).
4. Thomas M. Lehmann, Mark O. Glüd, et al.: Automatic categorization of medical images for content-based retrieval and data mining. Computerized Medical Imaging and Graphics, vol. 29, no. 2, pp. 143--155(2005).
5. Uri A, Hayit G, Eli K, et al.: X-ray Categorization and Retrieval on the Organ and Pathology Level, Using Patch-Based Visual Words. IEEE Transactions on Medical Image, vol. 30, no. 3, pp. 733--746(2011).
6. Jingyan Wang, Yongping Li, et al.: Bag-of-Features Based Medical Image Retrieval via Multiple Assignment and Visual Words Weighting. IEEE Transactions on Medical Image, vol. 30, no. 3, pp. 1996--2011(2011).
7. Tatiana T, Francesco O, Barbara C.: Discriminative cue integration for medical image annotation. Pattern Recognition Letters, vol. 29, no. 15, pp. 1996--2002(2009).
8. Xiaojin Zhu, Zoubin G, John L.: Semi-supervised learning using Gaussian fields and harmonic functions. International Conference on Machine Learning (ICML), pp. 912--919(2003).
9. Dengyong. Z, O. B, et al.: Learning with Local and Global Consistency. Advances in Neural Information Processing Systems (NIPS), pp. 321--328(2004).
10. Jinhui Tang, Guo-Jun Qi, et al.: Video semantic analysis based on structure-sensitive anisotropic manifold ranking. Signal Processing, vol. 89, pp. 2331--2323(2009).
11. David G. Lowe.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, vol. 60, no. 2, pp. 91--110(2004).
12. Svetlana L, Cordelia S.: Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. IEEE Computer Vision and Pattern Recognition (CVPR), pp. 2169--2178(2006).
13. Tatiana T, Barbara C, et al.: Overview of the CLEF 2009 medical image annotation track. The 10th international conference on Cross-language evaluation forum: multimedia experiments, pp. 85--93(2010).