REGRESSION AND CLASSIFICATION BASED DISTANCE METRIC LEARNING FOR MEDICAL IMAGE RETRIEVAL

Weidong Cai[1], *Yang Song*[1], *David Dagan Feng*[1,2,3]

[1]Biomedical and Multimedia Information Technology (BMIT) Research Group,
School of Information Technologies, University of Sydney, Australia
[2]Center for Multimedia Signal Processing (CMSP), Department of Electronic &
Information Engineering, Hong Kong Polytechnic University, Hong Kong
[3]Med-X Research Institute, Shanghai Jiao Tong University, China

ABSTRACT

Better utilizing the vast amount of valuable information stored in the medical imaging databases is always an interesting research area, and one way is to retrieve similar images as a reference dataset to assist the diagnosis. Distance metric is a core component in image retrieval; and in this paper, we propose a new learning-based distance metric design, based on regression and classification techniques. We design a weight learning approach by classifying the similar-dissimilar data samples, and a further optimization with a sparsity-constraint regression algorithm for feature selection. The learned distance metric is generally applicable for medical image retrievals. We evaluate the proposed method on clinical PET-CT images, and demonstrate clear performance improvements.

Index Terms— image retrieval, distance metric, regression, sparsity, classification

1. INTRODUCTION

As digital medical images are produced in ever increasing quantities, vast amount of valuable information is stored in the imaging databases, but remains largely unexplored. To better utilize such information, physicians can use a computerized content-based image retrieval (CBIR) system to retrieve images similar to the query image, and the retrieved images can then be used for medical researches, or as a reference dataset to assist the patient diagnosis [1].

Similarly to CBIR approaches in the general imaging domain, the core components of a medical CBIR system are the feature extraction and similarity measure [2]. Usually, an image is first represented as a vector of features, and then the degree of similarity between two images is computed based on a distance metric. Both components affect the retrieval performance greatly, and in this work, we focus on designing a new and effective distance metric.

For vector-based image features, basically any between-vector distance metric can be applied, such as the Euclidean and histogram-intersection distances. It is also well acknowledged that with high dimensional feature vectors, not all features should contribute equally to the distance computation. Therefore, feature weights are normally incorporated into the distance metric for a more effective similarity measure. The feature weights are often determined based on predefined constructs, such as the object volumes [3], and category [4] or level [5] specific values. Although these choices of feature weights are often intuitively suitable for the problem areas, there is no guarantee that such feature weights would lead to high retrieval performance.

Learning-based approaches have thus been proposed for obtaining feature weights that optimize the retrieval measures. The triplet-based learning methods [6, 7] are probably the most popular now, with their explicit modeling of the similar-dissimilar triplet relationships between data samples. Variations of such methods have also been proposed for medical images recently [8, 9, 10]. A common problem with these triplet-learning approaches is, however, the large number of training samples that is proportional to the possible combinations of all triplets, and the resultant long training time. A different type of learning-based techniques derives the feature weights by classifying the features into various categories [11, 12] rather than modeling the similar-dissimilar relationships, and can be trained with fewer training samples. However, since such feature weights are optimized for classification, they normally do not imply a good nearest-neighbor measure, which is the actual means used to measure distances between images and to retrieve similar images.

In this paper, we present a new distance metric learning method based on regression and classification. Our main contributions are three-folds. First, the feature weights are learned by classifying the between-vector distances into similar or dissimilar categories. Comparing to the triplet-learning approaches, our training sample formulation leads to a simpler solution with higher effectiveness and efficiency. The

This work was supported in part by ARC and PolyU grants.

classification procedure is also essentially different from [11, 12] that the optimization objective is to differentiate similar images from the dissimilar ones. Second, to further optimize the distance metric, we design a regression-based approach with sparsity constraints to select a subset of features before the weight learning. Comparing to the conventional dimension reduction techniques such as the principal component analysis (PCA), our approach effectively integrates the similar-dissimilar objectives during the feature selection in a supervised manner. Third, our proposed method is generally applicable to all image retrieval problems; for evaluation purposes, we test on clinical PET-CT images with our recently proposed feature vectors [10], and demonstrate clear performance improvements.

2. METHODS

2.1. Background

Let f_i and f_j be the n-dimensional feature vectors of image I and J. The distance value between them is computed as d_{ij}:

$$d_{ij} = \omega \cdot \Delta(f_i, f_j) \tag{1}$$

where $\Delta(f_i, f_j) \in R^{n\times 1}$ is a feature distance vector, with each element representing the difference of corresponding feature elements between f_i and f_j; and $\omega \in R^{n\times 1}$ is the feature weights. In our formulation, the distance vector $\Delta(f_i, f_j)$ can be computed using any vector distance functions (e.g. L1 and L2); and the objective is to optimize the feature weights ω to achieve high retrieval performance.

In a triplet-learning framework [6, 7], given three images I, J and K, with I similar to J but dissimilar to K, ideally the distance metric should satisfy $\omega \cdot \Delta(f_i, f_j) < \omega \cdot \Delta(f_i, f_k)$, which is equivalent to the following by scaling ω:

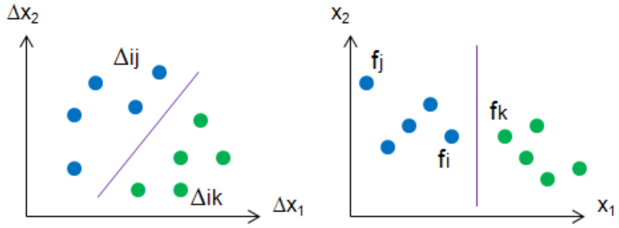
$$\omega \cdot \Delta(f_i, f_j) + 1 \leq \omega \cdot \Delta(f_i, f_k) \tag{2}$$

The goal is thus to minimize the total empirical loss given a set of T training triplets:

$$\sum_{t=1}^{T} [1 - \omega \cdot (\Delta(f_i, f_k) - \Delta(f_i, f_j))_t]_+ \tag{3}$$

where t denotes the tth training triplet $\langle f_i, f_j, f_k \rangle$. Such a goal is normally integrated into a large margin construct as a optimization constraint to solve for ω.

It can be seen that this triplet-learning naturally captures the similar-dissimilar relationships in retrieval. However, assuming there are N images, the maximum number of triplet samples is $N \cdot (N-1) \cdot (N-2)$, which is potentially quite large. If we ignore the ranking information between similar images, and assume each image has on average P similar and Q dissimilar images, the number of triplet samples reduces to $N \cdot P \cdot Q$. To further reduce the training size, bootstrapping approaches [10] have been used; however, the learned feature weights ω might then be sub-optimal.



(a) (b)

Fig. 1. Toy examples on a 2-dimensional feature space, with image I similar to J and dissimilar to K. Green and blue represent the two classes $\{-1, 1\}$, and purple indicates the separation line. (a) Feature distance space Δ_{x1} and Δ_{x2}, with Δ_{ij} and Δ_{ik} denoting $\Delta(f_i, f_j)$ and $\Delta(f_i, f_k)$. (b) Feature space $x1$ and $x2$.

2.2. Classification-based Weight Learning

Based on the above background, we propose a different weight-learning method. With the distance metric same as Eq. (1), we define label $l \in \{-1, 1\}$ for d_{ij}, with 1 indicating I similar to J and -1 otherwise. The problem of similarity measure can be thus rephrased as the following: given two images I and J, we want to obtain the optimal labeling for their feature distance d_{ij} by:

$$l^* = \underset{l}{\operatorname{argmax}} \;\; P(l|d_{ij}) \tag{4}$$

where $P(.)$ is the probability estimate. A large difference between $P(1|d_{ij})$ and $P(-1|d_{ij})$ helps with high labeling accuracy, and such an objective can be modeled as a large-margin optimization problem between the two classes $\{-1, 1\}$.

We solve the problem using a linear-kernel support vector machine (SVM) [13]:

$$\begin{aligned} \operatorname{argmin}_{\omega,\xi_t} \quad & \tfrac{1}{2} \parallel \omega \parallel^2 + C \textstyle\sum_t \xi_t \\ s.t.\ \forall t: \quad & l_t(\omega \cdot \Delta(f_i, f_j)_t) \geq 1 - \xi_t,\ \xi_t \geq 0 \end{aligned} \tag{5}$$

where t indexes the training sample $\langle f_i, f_j \rangle$. The empirical loss ξ_t and regulation parameter C follow the standard definitions in SVM, and the bias term b is merged into ω by appending $\Delta(f_i, f_j)_t$ with an additional dimension of constant value 1. With a set of T training samples, the feature weights ω are then derived.

Although our approach resembles closely to the triplet-learning methods without considering the ranking information between similar images, it exhibits two main advantages. First, since our optimization objective is to maximize the separation margin between the similar and dissimilar categories, it is easier to accommodate less discriminative feature vectors. To further clarify, as shown in Fig. 1a, large feature distances may exist between similar images Δ_{ij}; to enforce

Eq. (3) between all pairs of Δ_{ij} and Δ_{ik} is thus more complicated than just to establish a separation line between the two sets of samples. Second is the much reduced training size and shorter training time. With N images, and each having on average P similar and Q dissimilar images, the number of training samples is thus $N \cdot (P + Q)$, which is much smaller than the number of triplet samples $N \cdot P \cdot Q$.

We also illustrate that if the classification is performed in the original feature space f_i, rather than the feature distance space $\Delta(f_i, f_j)$, the derived weights would not be suitable for similarity measure. As shown in Fig. 1b, if the margin differences between f_i and f_j is larger than that between f_i and f_k, image I could be identified as more similar to K rather than J. This is also in accordance with Eq. (1) that ω is in fact applied to the feature distance vector.

2.3. Regression-based Feature Selection

To further optimize the feature weights for the distance metric, we propose to perform feature selections before training for the feature weights ω. To do this, a least squares regression with sparsity constraints [14] is incorporated:

$$\begin{aligned} \operatorname{argmin}_x \quad & \tfrac{1}{2}\|y - Dx\|^2 + \tfrac{1}{2}\alpha\|x\|_2^2 \\ s.t. \quad & \|x\|_1 \leq Z \end{aligned} \tag{6}$$

Here $\|y - Dx\|^2$ is the least squares term, $\|x\|_2^2$ is the L2 regularization term, and $\|x\|_1 \leq Z$ is the sparsity constraint with $Z < n$, where n is the feature dimension of f_i. The vector $y = \{y_t : t = 1, ..., T\} \in \{1, 2\}^{T\times 1}$ is the labeings of the data samples $D = \{\Delta(f_i, f_j)_t\} \in R^{T\times n}$, and $x \in R^{n\times 1}$ is the corresponding weight vector.

The idea is that we want to minimize the non-zeros elements in x, while achieving a low regression error simultaneously. The non-zero elements in x then indicate the features selected, and x is then used to refine the weight learning by modifying Eq. (5) as the following:

$$\begin{aligned} \operatorname{argmin}_{\omega,\xi_t} \quad & \tfrac{1}{2} \parallel \omega \parallel^2 + C \textstyle\sum_t \xi_t \\ s.t.\ \forall t: \quad & l_t(\omega \cdot x \cdot \Delta(f_i, f_j)_t) \geq 1 - \xi_t,\ \xi_t \geq 0 \end{aligned} \tag{7}$$

Note that here x is also appended with an additional dimension of constant value 1 to accommodate the bias term in ω.

2.4. Evaluation Method

The distance metric designed can be generally applied to any imaging modality. In this work, we test the proposed method based on our previous CBIR work on positron emission tomography – computed tomography (PET-CT) thoracic images from non-small cell lung cancer (NSCLC) studies [10].

In [10], a grid-based bag-of-words representation is first created with 16 feature words. The center of lung tumor is then estimated as the area with the highest standard uptake value (SUV), and a hierarchical bag-of-words feature vector f_i of dimension 336 is derived originated from the center of the tumor. Image similarities are then computed based on a weighted histogram-intersection distance metric between the feature vectors, and the feature weights are obtained using the triplet-learning approach. In this work, we adopt the same feature vector f_i, but replacing the distance metric with our proposed method. The retrieval performance is evaluated using precision and recall, for all levels of retrieval recalls.

The retrieval tests are performed on 40 PET-CT thoracic images, which are acquired using a Siemens TrueV 64 PET-CT scanner at the Royal Prince Alfred Hospital, Sydney. For each patient study, the other 39 patient studies were manually marked as similar or dissimilar as a benchmark for measuring the retrieval performance. The similarity of cases was determined based on the spatial contexts and appearance of primary lung tumors and abnormal lymph nodes. In particular, the location of lung tumors in the lung and relative to the mediastinum (e.g. with invasion) was an important factor of similarity. The number of similar cases for each case ranged from 1 to 11, with an average of 4.75 [10]. A preprocessing is performed on each image slice to remove the soft tissues outside of the lung field and mediastinum using thresholding and connected component analysis.

3. RESULTS

From our experiments, for weight learning, the choice of training images has high impact on the retrieval results. We choose to employ a bootstrapping approach for training image selection. The procedure is briefly described as follows. First, a number of images exhibiting typical lung tumor characteristics are selected. Then, training is conducted and based on initial testing results, images showing poor retrieval performances are added to the training set. The second step is executed iteratively until the total number of training images reaches 10 (i.e. $1/4$ of the total database size).

A separate training is also conducted for the regression-based feature selection, using the same training set constructed for weight learning. As a result of the feature selection, the feature dimension is reduced from 336 to 314. Further reduction is possible, but that leads to information loss and less optimized regression goal. With the 314-dimensional features selected, we obtain on average 1% improvement of retrieval precision for all levels of recalls, comparing to the original feature set.

We then compare our proposed distance metric with the baseline techniques: (i) feature weights are not used (No-W); (ii) a triplet-learning method trained using all triplet samples created from the training images (Trip-All); and (iii) a triplet-learning method trained with reduced triplet samples selected with a bootstrapping procedure (Trip-B) [10]. As shown in Fig. 2, our proposed method achieves the highest retrieval precisions. Without the feature weights, as expected, the retrieval precisions are much lower. The two triplet-based

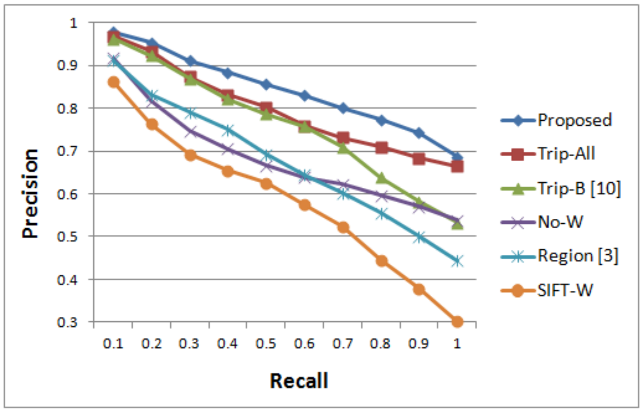


Fig. 2. The average precision and recall of retrieval results.

methods result in similar retrieval performances, but Trip-All outperforms the other for recalls $>$ 0.6. This can be explained by the optimization criteria during the bootstrapping procedure, which aims at maximizing the retrieval precisions for the first a few retrievals. The on average 5% improvements over Trip-All shows the advantage of our regression- and classification-based distance metric design. We believe that due to insufficient discriminative power of the feature vector, triplet-learning is affected negatively by training samples exhibiting contradicting similar-dissimilar relationships, which could be better handled with an SVM-type formulation. However, we also acknowledge that the selection of training images does affect the retrieval results. If too few images are used for training, both types of methods would show lower performances, and the advantage of the proposed method would be less obvious.

We then further test our proposed method using other types of features. First, we change to use the popular 128-dimensional SIFT descriptor with the proposed distance metric; although the retrieval precisions are quite low as shown in Fig. 2 (SIFT-W), it would have been on average 12% lower if without the learned feature weights. We also find that little improvement is obtained by integrating the new distance metric with [3]; and we believe it is due to its low feature dimension (i.e. 16). We thus suggest that a relatively high dimensional feature vector is helpful to achieve good similarity measures with our proposed method.

4. CONCLUSIONS

In this paper, we present a new distance metric design for medical image retrieval with regression-based feature selection and classification-based weight learning. Based on vector-based feature descriptors and distance function, the feature weights are learned by classifying the similar and dissimilar feature distance vectors. A feature selection procedure is also conducted for further optimization with a sparsity-constraint least squares regression model. The proposed distance metric is evaluated on PET-CT thoracic images, and promising results are observed comparing to the baseline techniques.

5. REFERENCES

[1] H. Muller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medical applications - clinical benefits and future directions," *Int J. Medical Informatics*, vol. 73, pp. 1–23, 2004.

[2] R. Datta, D. Joshi, J. Li, and J.Z. Wang, "Image retrieval: ideas, influences, and trends of the new age," *ACM Computing Surveys*, vol. 40, no. 2, pp. 1–60, 2008.

[3] Y. Song, W. Cai, S. Eberl, M.J. Fulham, and D. Feng, "Thoracic image case retrieval with spatial and contextual information," *in Proc. ISBI*, pp. 1885–1888, 2011.

[4] M.M. Rahman, S.K. Antani, and G.R. Thoma, "A biomedical image retrieval framework based on classification-driven image filtering and similarity fusion," *in Proc. ISBI*, pp. 1905–1908, 2011.

[5] P. Zhu, S.P. Awate, S. Gerber, and R. Whitaker, "Fast shape-based nearest-neighbor search for brains MRIs using hierarchical feature matching," *in MICCAI 2011, LNCS*, vol. 6892, pp. 484–491, 2011.

[6] K.Q. Weinberger, J. Blitzer, and L.K. Saul, "Distance metric learning for large margin nearest neighbor classification," *in Proc. NIPS*, pp. 1–8, 2006.

[7] A. Frome, Y. Singer, F. Sha, and J. Malik, "Learning globally-consistent local distance functions for shape-based image retrieval and classification," *in Proc. ICCV*, pp. 1–8, 2007.

[8] M. Liu, L. Lu, J. Bi, V. Raykar, M. Wolf, and M. Salganicoff, "Robust large scale prone-supine polyp matching using local features: a metric learning approach," *in MICCAI 2011, LNCS*, vol. 6893, pp. 75–82, 2011.

[9] W. Yang, Q. Feng, Z. Lu, and W. Chen, "Metric learning for maximizing map and its application to content-based medical image retrieval," *in Proc. ISBI*, pp. 1901–1904, 2011.

[10] Y. Song, W. Cai, and D. Feng, "Hierarchical spatial matching for medical image retrieval," *in Proc. ACM MM workshop MMAR*, 2011.

[11] J. Zhang, S.K. Zhou, S. Brunke, C. Lowery, and D. Comaniciu, "Detection and retrieval of cysts in joint ultrasound b-mode and elasticity breast images," *in Proc. ISBI*, pp. 173–176, 2010.

[12] A. Sridhar, S. Doyle, and A. Madabhushi, "Boosted spectral embedding (bose): applicatioins to content-based image retrieval of histopathology," *in Proc. ISBI*, pp. 1897–1900, 2011.

[13] C.C. Chang and C.J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.

[14] J. Liu, S. Ji, and J. Ye, "SLEP: sparse learning with efficient projections," http://www.public.asu.edu/jye02/Software/SLEP.