

LETTER

A New Re-Ranking Method Using Enhanced Pseudo-Relevance Feedback for Content-Based Medical Image Retrieval

Yonggang HUANG^{†a)}, Student Member, Jun ZHANG^{††}, Yongwang ZHAO[†], and Dianfu MA[†], Nonmembers

SUMMARY We propose a novel re-ranking method for content-based medical image retrieval based on the idea of pseudo-relevance feedback (PRF). Since the highest ranked images in original retrieval results are not always relevant, a naive PRF based re-ranking approach is not capable of producing a satisfactory result. We employ a two-step approach to address this issue. In step 1, a Pearson's correlation coefficient based similarity update method is used to re-rank the high ranked images. In step 2, after estimating a relevance probability for each of the highest ranked images, a fuzzy SVM ensemble based approach is adopted to re-rank the images. The experiments demonstrate that the proposed method outperforms two other re-ranking methods.

key words: CBIR, re-ranking, similarity update, fuzzy SVM ensemble

1. Introduction

With the development of the Internet, medical images are now available in large numbers in on-line repositories. In such a web-based environment, images are generally stored and accessed in common formats (such as JPEG, GIF, etc) other than DICOM format for convenience and anonymization purposes [1]. Since there is no textual description attached in this situation, the text-based approach is both expensive and ambiguous due to the fact that manually annotating these images is extremely time-consuming and highly subjective. This leads to the use of content-based image retrieval (CBIR) techniques which can search for medical images based on the modality, anatomic region and different acquisition views [1].

In CBIR, images are retrieved according to their visual similarities on extracted low-level features, such as color, texture and spatial location. Obviously, the effectiveness of CBIR strongly depends on the choices of features and distance metrics, both of which have been widely studied in previous works [2]. However, no matter how suitable the two are designed, CBIR still can't obtain satisfactory results. This is mainly due to the semantic gap existing between the low-level features and the high-level human perception [2]. A number of research efforts have been devoted in further improving the effectiveness of CBIR.

Relevance feedback (RF) [2] has shown certain success in CBIR. During the RF procedure, the user labels some re-

turned images as relevant or irrelevant, and the CBIR system adjusts the retrieval parameters according to the user's feedback. Then, the system displays the revised retrieval results. RF can go through one or more iterations until the user is satisfied with the results. However, RF needs users' intervention to provide relevance information, which imposes a burden on users.

Re-ranking (RR), which revises the retrieval results by exploiting the information encoded in the original retrieval results, has also been introduced into CBIR. RR carries out without the users' participation. Park et al. [3] proposed a re-ranking method using post-retrieval clustering (PC-RR). In their method, firstly the retrieved images are analysed by hierarchical agglomerative clustering, and the image ranks are adjusted according to the distances between the image clusters and the query image. Jing et al. [4] formulated the re-ranking process as a random walk over a visual similarity graph. The PageRank algorithm is used to assign numerical weight to each image on the visual similarity graph, and the images are re-ranked according to their weights. However, the re-ranking methods presented in references [3], [4] only take into account the high ranked images without consideration of the whole image collection, which limits their performance.

Pseudo-relevance feedback (PRF), which has been successfully used in text retrieval, is an effective technique exploiting original retrieval results for re-ranking. The basic idea of PRF is to assume that a small number of the highest ranked documents in the original retrieval results are relevant, and re-rank the documents with adjusted retrieval parameters. Yan et al. [5] made use of the low ranked images for re-ranking with negative pseudo-relevance feedback (NPRF-RR), since the low ranked images are very likely to be irrelevant. When multiple example images are available in the query for video retrieval task, NPRF-RR takes the low ranked images as negative examples and the example images in the query as positive examples. With different negative examples randomly selected from the low ranked images, multiple support vector machine (SVM) classifiers are constructed, and the images are re-ranked through combining the outputs of all classifiers. However, NPRF-RR is proposed to deal with multiple-example queries without taking single-example queries into account. For the single-example queries, the highest ranked images are not always relevant due to the limited accuracy of current CBIR systems, and a naive PRF is not capable of producing a satisfactory result because of these noisy images. Zhang

Manuscript received September 6, 2011.

Manuscript revised October 19, 2011.

[†]The authors are with the School of Computer Science and Engineering, Beihang University, Beijing, China.

^{††}The author is with the School of Information Technology, Deakin University, Victoria, Australia.

a) E-mail: yonggang.h@gmail.com

DOI: 10.1587/transinf.E95.D.694

et al. [6] and Le et al. [7] applied SVM classifier ensemble method for image re-ranking to leverage the noisy images existed in the example sets respectively. However, they treat the pseudo positive examples equally without exploiting the different relevance probabilities of them.

In this letter, we propose a new re-ranking method using enhanced pseudo relevance feedback (EPRF-RR) for content-based medical image retrieval. By comparison with naive PRF, EPRF-RR: (I) improves the accuracy of the highest ranked images during a new similarity update process; (II) alleviates the influence of the irrelevant images by associating each of the highest ranked images with a relevance probability; (III) further improves the retrieval performance of re-ranking using a fuzzy SVM ensemble approach.

The remainder of the letter is organized as follows. In Sect. 2, EPRF-RR is presented in detail. The experiments and results are reported in Sect. 3. Finally, Sect. 4 concludes this letter.

2. Re-ranking Method Based on Enhanced Pseudo Relevance Feedback

2.1 Overview of the Proposed Re-Ranking Method

Let us consider a medical image collection which contains n images, $\mathbb{I} = \{I_i\}_{i=1}^n$. Suppose the low-level feature $\mathbb{F}(\cdot)$ and distance metric $\mathbb{D}(\cdot, \cdot)$ are available. Given a query image Q , the distance between Q and I_i can be calculated as $\mathcal{D}_i = \mathbb{D}(\mathbb{F}(Q), \mathbb{F}(I_i))$. The distance can be normalized as: $\bar{\mathcal{D}}_i = \frac{\mathcal{D}_i - \mathcal{D}_{\min}}{\mathcal{D}_{\max} - \mathcal{D}_{\min}}$, where \mathcal{D}_{\max} and \mathcal{D}_{\min} denote the maximum and minimum distances between Q and images in \mathbb{I} . The normalized distances can be converted to similarities as $S_i = 1 - \bar{\mathcal{D}}_i$. The original similarity ranking list (SRL) \mathcal{R}^0 is constructed as:

$$\mathcal{R}^0 = [S'_1, \dots, S'_i, \dots, S'_n] \quad \text{and} \quad S'_1 \geq \dots \geq S'_n \quad (1)$$

where S'_i denotes the i -th largest similarities. The proposed re-ranking method is performed on \mathcal{R}^0 as:

- I. To improve the accuracy of the highest ranked images, a Pearson's correlation coefficient based similarity update method is applied to re-rank the high ranked K images in \mathcal{R}^0 , and a new SRL $\bar{\mathcal{R}}^0$ is constructed.
- II. A two-step strategy is used to re-rank images in \mathbb{I} as:
 - a. To alleviate the influence of the irrelevant images, a relevance probability estimating method is employed to associate a relevance probability for each of the highest ranked P images in $\bar{\mathcal{R}}^0$.
 - b. A fuzzy SVM ensemble based approach, which can take into account the highest ranked P images with different relevance probabilities and overcome the instability of single fuzzy SVM, is proposed to re-rank all the images in \mathbb{I} .



Fig. 1 A example query.

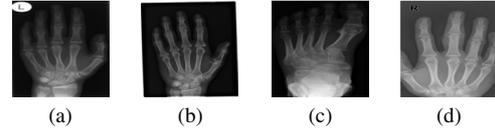


Fig. 2 The top four ranked images.

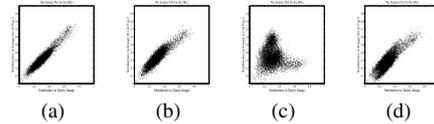


Fig. 3 The scatter plots for the SRLs of images in Fig. 2.

2.2 Similarity Update Based on Pearson's Correlation Coefficient

The similarities update method is proposed following the hypothesis that relevant images tend to have closer SRLs when they are taken as query images separately.

The hypothesis is confirmed by the experimental results on IRMA dataset [8]. Figure 1 shows a example query, and Fig. 2 presents the top four ranked images on the whole IRMA dataset using the selected low-level feature and distance metric in Sect. 3. As indicated in Fig. 2, image (a, b, d) is relevant to example image, and image (c) is visually similar but irrelevant to the example image. The SRLs are constructed for each of the top ranked images. Figure 3 reports the scatter plots[†] for the SRLs of the top four ranked images to the example image. The scatters for the SRLs of images (a, b, d) tend to concentrate along the line $y = x$, which indicate they have a close relationship. In contrast, the scatters for the SRLs of images (c) are not clustered around the line $y = x$, which shows they haven't a close relationship.

Considering the computation cost, the similarity update is performed only for the high ranked K images set \mathcal{Q} , and the high ranked N images set \mathbb{I}_N are used as reference images set to construct SRLs. For query image Q , the SRL constructed on \mathbb{I}_N can be represented as \mathbf{r}^0 . Taking all images in \mathcal{Q} as query images, the SRLs constructed on \mathbb{I}_N can be denoted as $\mathbb{R} = \{\mathbf{r}^1, \dots, \mathbf{r}^i, \dots, \mathbf{r}^K\}$. In this letter, the Pearson's correlation coefficient is used to characterize the relationships among SRLs as:

[†]Scatter plot is a type of mathematical diagram to display values of two data sets, and it can be used to show how two comparable data sets agree with each other. The more the two data sets agree, the more the scatters tend to concentrate in the vicinity of the line $y = x$.

$$\rho_{\mathbf{r}^0, \mathbf{x}^i} = \frac{\sum_{j=1}^N (\mathbf{r}_j^0 - \bar{\mathbf{r}}^0)(\mathbf{x}_j^i - \bar{\mathbf{r}}^i)}{\sqrt{\sum_{j=1}^N (\mathbf{r}_j^0 - \bar{\mathbf{r}}^0)^2} \sqrt{\sum_{j=1}^N (\mathbf{x}_j^i - \bar{\mathbf{r}}^i)^2}} \quad (2)$$

where $\bar{\mathbf{r}}^0$ and $\bar{\mathbf{r}}^i$ denote the means of similarities in \mathbf{r}^0 and \mathbf{r}^i respectively. The value of $\rho_{\mathbf{r}^0, \mathbf{x}^i}$ ranges from -1 to 1 , and it measures the strength of linear dependence of two SRLs. $\rho_{\mathbf{r}^0, \mathbf{x}^i}$ can be used to update the similarities of images in \mathbb{Q} as:

$$\bar{\mathcal{S}}'_i = \mathcal{S}'_i + \lambda \cdot \rho_{\mathbf{r}^0, \mathbf{x}^i} \quad (3)$$

where λ is an additional slack factor, which is simply set as 1 in this letter.

After that, a new SRL $\bar{\mathcal{R}}^0$ is constructed using the updated similarities.

2.3 Image Re-ranking with Fuzzy SVM Ensemble

The classifiers for PRF are constructed by treating the highest ranked \mathbb{P} images $\mathbb{P} = \{\mathcal{P}_1, \dots, \mathcal{P}_i, \dots, \mathcal{P}_p\}$ in $\bar{\mathcal{R}}^0$ as positive examples. Since the low ranked images are very likely to be irrelevant, the randomly sampled images from the lowest ranked images in $\bar{\mathcal{R}}^0$ are treated as negative examples. The number of negative examples is set as the same with the number of positive examples. Considering that irrelevant images may exist in \mathbb{P} , single classifier will be unstable. In order to overcome the instability of single classifier, Zhang et al. [6] applied SVM classifier ensemble method. The SVM ensemble method used different negative example sets to train SVM and get multiple classifier. After that, all classifiers were used to classify the images. In this paper, by incorporating the ideas of classifier ensemble and noise tolerant classifier, a fuzzy SVM ensemble based approach, is proposed to re-rank all the images in \mathbb{I} . Different from SVM ensemble method, the fuzzy SVM ensemble based approach can not only overcome the instability of single classifier, but also take into account the highest ranked \mathbb{P} images with different relevance probabilities.

2.3.1 Relevance Estimation

We associate with each image a relevance probability to alleviate the influence of irrelevant images in \mathbb{P} . The relevance probabilities are estimated based on the relationships among SRLs and the class-conditional probability of SVM ensemble.

The SVM classifier is constructed in the same way with the classifiers for PRF. Due to the irrelevant images in \mathbb{P} , single SVM classifier will be unstable. Based on the idea of classifier ensemble [6], we use different negative example sets to train SVM and obtain multiple classifiers. All classifiers are then applied to classify the images in \mathbb{P} . The sigmoid function combined with the output of an SVM classifier is used to estimate the class-conditional probability for image \mathcal{P}_i by

$$P(C, \mathcal{P}_i) = \frac{1}{1 + \exp(-f(\mathcal{P}_i))} \quad (4)$$

where $f(\mathcal{P}_i)$ is the decision value produced by the SVM classifier. The outputs of multiple classifiers are then combined to get the conditional probabilities using Bayes sum rule.

$$P(\mathcal{P}_i) = \frac{1}{T_1} \sum_{j=1}^{T_1} P(C_j, \mathcal{P}_i) \quad (5)$$

where \mathcal{P}_i is the conditional probabilities of SVM classifiers ensemble and T_1 is the number of SVMs.

The estimated relevance probability for image \mathcal{P}_i is then computed as:

$$R(\mathcal{P}_i) = \alpha \cdot P(\mathcal{P}_i) + (1 - \alpha) \cdot \rho_{\mathbf{r}^0, \mathbf{x}^i} \quad (6)$$

when $\rho_{\mathbf{r}^0, \mathbf{x}^i}$ is the Pearson's correlation coefficient between SRLs of image \mathcal{P}_i and the query image \mathcal{Q} , and α is weighting factor to determine the influence of conditional probabilities of SVM ensemble. In this letter, α is set to 0.4 based on our informal experiment results.

2.3.2 Image Re-Ranking

Since the images in \mathbb{P} have different relevance probabilities, the classifier for PRF is designed using the fuzzy SVM algorithm [9]. By comparison with regular SVM, each training example in fuzzy SVM has an assigned membership value μ_i according to its relative importance in the class. For positive examples, the fuzzy membership values are set according to their estimated relevance probabilities in formula (6) as $\mu_i = R(\mathcal{P}_i)$. For negative examples, the fuzzy membership values are set to 1 . With different negative example sets, multiple fuzzy SVMs are trained to overcome the instability of single fuzzy SVM. All the trained fuzzy SVMs are then used to classify images in \mathbb{I} . The final relevance $R(\mathcal{I}_i)$ of image \mathcal{I}_i is obtained by combining the outputs of all the fuzzy SVMs in the form of a logistic regression as:

$$R(\mathcal{I}_i) = \frac{\exp(\sum_{j=1}^{T_2} f_j(\mathcal{I}_i))}{1 + \exp(\sum_{j=1}^{T_2} f_j(\mathcal{I}_i))} \quad (7)$$

where $f_j(\mathcal{I}_i)$ is the decision value produced by the fuzzy SVM classifier and T_2 is the number of fuzzy SVMs.

Finally, the images in \mathbb{I} are re-ranked according to their final relevance.

3. Experiments and Results

A number of experiments were carried out on the IRMA medical image collection [8] which contains 9000 medical images and are subdivided into 57 classes. The images are classified manually by reference coding with respect to a mono-hierarchical coding scheme. The scheme describes the imaging modality, the body orientation, the body region examined and the biological system examined. To evaluate the content based medical image retrieval, the example images for single-example queries were randomly selected

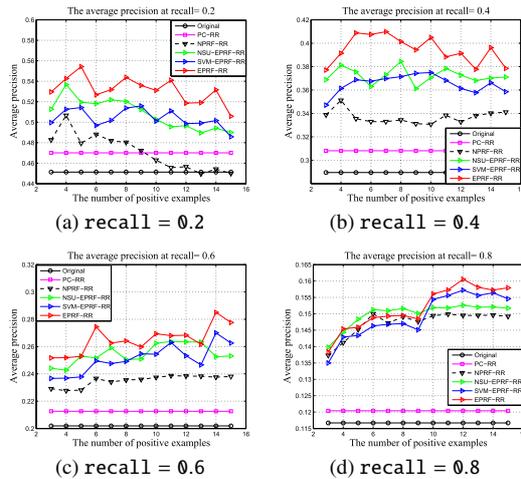


Fig. 4 The average precision at recall = 0.2, 0.4, 0.6, 0.8.

from each class and the remained images in the class were regarded as the corresponding ground truth set.

The Color Layout Descriptor, Tamura feature and Edge Histogram [2] were used to represent the medical images with a 168-dimensional vector.

The Euclidean distance metric was used to model the image similarities for original retrieval results.

To evaluate the performance of the proposed re-ranking method, the post-retrieval clustering based re-ranking method (PC-RR) [3] and negative pseudo-relevance feedback based re-ranking method (NPRF-RR) [5] were implemented as references. For PC-RR, the high ranked 200 images were clustered for re-ranking based on our experimental results. We used the LIBSVM [10] to solve SVMs and fuzzy SVMs. Gauss kernels were used in our experiments, and we applied grid search for optimal parameter set that produces the best retrieval performance. Since it is reported that the number of classifiers for ensemble does not affect the retrieval performance [6], ten SVMs and ten fuzzy SVMs are chosen for both NPRF-RR and EPRF-RR based on experimental results. For EPRF-RR, the similarity update was performed for the high ranked 60 images, and the high ranked 1000 images were chosen as reference images set to construct SRLs. In NPRF-RR and EPRF-RR, the number of the highest ranked images treated as positive examples varies from 3 to 15, and the 10% lowest ranked images were chosen as negative examples.

To evaluate the contributions of the similarity update and fuzzy SVM ensemble methods respectively, the proposed re-ranking method was implemented as NSU-EPRF-RR without using similarity update, and implemented as SVM-EPRF-RR using SVM ensemble.

The retrieval performance was measured by the average precision of 400 single-example queries, which were randomly created from the whole image collection. Figure 4 reports the average precisions at recall=0.2, 0.4, 0.6, 0.8, with the number of positive examples varying from 3 to 15. The experimental results show that the retrieval performance of EPRF-RR is better than PC-RR and NPRF-RR.

The precision of EPRF-RR is higher than that of the original method about 11 percent when the recall is less than 0.4. The precision of EPRF-RR is higher than that of PC-RR and NPRF-RR about 8 percent and 7 percent, respectively, when the recall is less than 0.4. We argue that ERPR-RR outperforms NPRF-RR since ERPR-RR improves the accuracy of the highest ranked images during a new similarity update process, and improves the retrieval performance by associating each of the highest ranked images with a relevance probability for classifier ensemble based image re-ranking.

One also can see that the precision of EPRF-RR is higher than that of NSU-EPRF-RR and SVM-EPRF-RR probably 4 percent and 5 percent, respectively, when the recall is less than 0.4. EPRF-RR is related to SVM-EPRF-RR, both of which employ the idea of classifier ensemble. However, EPRF-RR outperforms SVM-EPRF-RR because that EPRF-RR can take into account the different relevance probabilities of the highest ranked images while SVM-EPRF-RR does not.

In addition, the number of the highest ranked images treated as positive examples can affect the retrieval performance of NPRF-RR and EPRF-RR. For example, with the increase of the number of pseudo positive examples, the average precisions of the two methods decrease when recall is low (e.g., recall=0.2). However, the average precisions of the two methods increase when recall is high (e.g., recall=0.8). The reason could be twofold. Firstly, more highest ranked images are treated as positive examples, their ranks may not be changed after re-ranking and the precision at the low recall points cannot be improved. Secondly, more positive information will be obtained and used in the re-ranking process if more highest ranked images are treated as positive examples, so the precision at the high recall points can be improved. We will further investigate the effect of pseudo positive examples in the future work.

4. Conclusions

This letter proposed a new two-step re-ranking method using enhanced pseudo-relevance feedback for content-based medical image retrieval. In step 1, a similarity update method was used to re-rank the high ranked images. In step 2, a fuzzy SVM ensemble approach is adopted to re-rank the images. A number of experiments were carried out on a real-world medical image dataset and the results showed the proposed EPRF-RR method outperforms PC-RR and NPRF-RR. In addition, the retrieval performance of both NPRF-RR and EPRF-RR is influenced by the number of the highest ranked images treated as positive examples.

Acknowledgements

The authors thank Dr. Ken Nakamura and the anonymous reviewer for valuable feedback to this work. The authors also thank courtesy of TM Deserno, Dep. of Medical Informatics, RWTH Aachen, Germany, for providing IRMA dataset. This work was supported by Natural Science Foundation

(61003017) and Core-High-Base Project (2010ZX01042-002-001) of China.

References

- [1] M.M. Rahman, B.C. Desai, and P. Bhattacharya, "Medical image retrieval with probabilistic multi-class support vector machine classifiers and adaptive similarity fusion," *Computerized Medical Imaging and Graphics*, vol.32, no.2, pp.95–108, March 2008.
 - [2] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.22, no.12, pp.1349–1380, Dec. 2000.
 - [3] G. Park, Y. Baek, and H. Lee, "Re-ranking algorithm using post-retrieval clustering for content-based image retrieval," *Inf. Process. Manage.*, vol.41, no.2, pp.177–194, March 2005.
 - [4] Y. Jing and S. Baluja, "Pagerank for product image search," *Proc. 17th international conference on World Wide Web*, pp.307–316, Beijing, China, April 2008.
 - [5] R. Yan, A. Hauptmann, and R. Jin, "Multimedia search with pseudo-relevance feedback," *The 2nd International Conference Image and video retrieval*, pp.649–654, Urbana-Champaign, IL, USA, July 2003.
 - [6] J. Zhang and L. Ye, "Content based image retrieval using unclean positive examples," *IEEE Trans. Image Process.*, vol.18, no.10, pp.2370–2375, Oct. 2009.
 - [7] D.D. Le and S. Satoh, "Unsupervised face annotation by mining the web," *8th IEEE International Conference on Data Mining*, pp.383–392, Pisa, Italy, Dec. 2008.
 - [8] T. Lehmann, H. Schubert, B. Ott, and M. Leisten, "Image retrieval in medical applications," RWTH Aachen University, <http://ganymed.imib.rwth-aachen.de/irma/datasets>, accessed Sept. 3, 2011.
 - [9] C.F. Lin and S.D. Wang, "Fuzzy support vector machines," *IEEE Trans. Neural Netw.*, vol.13, no.2, pp.464–471, March 2002.
 - [10] C. Chang and C. Lin, "Libsvm: A library for support vector machines," National Taiwan University, <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>, accessed Sept. 3, 2011.
-