

UESTC at ImageCLEF 2012 Medical Tasks

Hong Wu, Kuangkai Sun, Xianzhi Deng, Yi Zhang, Bili Che

School of Computer Science and Engineering,
University of Electronic Science and Technology of China,
611731 Chengdu, P. R. China
hww@uestc.edu.cn, sunkuangkai@163.com, dengxian.zhi@gmail.com,
zhangyi308@163.com, blche@sina.cn

Abstract. This paper describes the methods used and results archived by our research group in the ImageCLEF 2012 medical retrieval and classification tasks. We performed three sub-tasks, ad-hoc retrieval, case-based retrieval, and modality classification. For the retrieval tasks, we combined semantic-based retrieval with traditional text-based retrieval. The semantic-based retrieval was conducted by comparing query concepts and document concepts with semantic similarity measure, and asymmetric similarity measures were also proposed by modifying the existing symmetric measures. For the modality classification task, we used multiple kernel learning to combine various visual features.

Keywords: Medical Retrieval, Semantic Similarity, Lemur, MeSH, Modality Classification, MKL

1 Introduction

This paper describes the third participation of the UESTC group in the ImageCLEF medical retrieval and classification tasks [1]. For the retrieval tasks, we adopted our retrieval method [2] which combining semantic-based retrieval with text-based retrieval. In our approach, each query or document has two representations, a text-based one and a concept-based one. The concept-based representations are constructed by mapping the text of a query or document to MeSH [3] descriptors (concepts) by MeSHUP [4], a MeSH classification system. In the semantic-based retrieval model, the query concepts and document concepts are compared with semantic similarity measure, and asymmetrical similarity measures are also proposed. Then the inter-concept similarities are aggregated to compute the relevance score of a document. For the text-based retrieval, any state of the art information retrieval models can be used. Finally, the semantic-based and text-based retrievals are combined. Our retrieval method was implemented on Lemur toolkit [5]. Based on our previous study [6], we used multiple kernel learning (MKL) to combine various visual features for modality classification.

The remainder of this paper is organized as follows. Our retrieval method used in both ad-hoc retrieval and case-based retrieval is described in section 2, and our MKL-

based modality classification method is given in section 3. Finally, our submitted runs and results are presented in section 4, followed by the conclusions in section 5.

2 Improving Medical Retrieval by Incorporating Semantic-Based Retrieval

In our approach, each document or query has two representations, a text-based representation and a concept-based representation, which are indexed and searched separately. The concept-based representation is constructed by mapping the text of a document or query to MeSH descriptors by MeSHUP [4], and each MeSH descriptor is used as index term as a whole. For the semantic-based retrieval, the query concepts and document concepts are compared with semantic similarity measures. And we also proposed asymmetrical semantic similarity measures which can be built by modifying existing symmetric measures. Then the inter-concept similarities are aggregated to compute the relevance score of a document. In the text-based representation, single words or word stems are used as index term. Any state of the art retrieval models, such as TF-IDF, BM25, etc. can be used for the text-based retrieval. Finally, the semantic-based and text-based searches are combined. Following section will give more details of our approach.

2.1 Semantic Similarity Measures

Ontology-based semantic similarity measures use ontology as the primary information source. They can be roughly grouped into three categories: path-based, information-based, and feature-based measures [7], and all of them are symmetric measure. Path based similarity measure is a straightforward and efficient approach. It usually utilizes the information of the shortest path between two concepts, of the generality or specificity of both concepts in ontology hierarchy. The information-based approaches are based on the information theory which use text corpus as secondary information source. They all use information content (IC) of concept nodes derived from the IS-A relations and corpus statistics. Feature based measure assumes that each concept is described by a set of terms indicating its properties or features. Then, the more common characteristics and the less non-common characteristics two concepts have, the more similar they are.

In this paper, we use a path-base measure, Li's measure [8], which combines the shortest path and the depth of ontology information in a non-linear function:

$$S_{Li}(c_1, c_2) = e^{-\alpha L} \frac{e^{\beta H} - e^{-\beta H}}{e^{\beta H} + e^{-\beta H}} \quad (1)$$

where L stands for the shortest path between two concepts, α and β are parameters scaling the contribution of shortest path length and depth respectively. In our experiment, we set α and β to 0.2 and 0.6 respectively.

2.2 Asymmetrical Semantic Similarity

Contrary to document clustering and classification, document retrieval is an asymmetric problem. For example, for a query with term *Brain Diseases*, the document containing term *Alzheimer Disease* has a high probability of being relevant. But, for a query with term *Alzheimer Disease*, the document containing term *Brain Diseases* would not necessarily be relevant. Based on this observation, we propose asymmetrical semantic similarity, with which the similarity of the former pair in the example is greater than that of the later one. The asymmetric measure is given as

$$S_1^A(c_1, c_2) = \begin{cases} S(c_1, c_2), & c_1 \text{ is an ancestor of, or equal to } c_2 \\ \gamma \times S(c_1, c_2), & \text{else} \end{cases} \quad (2)$$

where c_1, c_2 are two concepts from a query and a document respectively. $S(c_1, c_2)$ can be any semantic similarity measures, such as edge-based, information-based measures etc.. $\gamma \in [0,1)$ is a punishment factor to reduce the similarity value if c_1 is neither an ancestor of, nor equal to c_2 . When γ is set to 0, only the document concept, which is a child of, or equal to a query concept, will contribute to the relevance score of that document.

If we do not distinguish the document concepts, which are children of, or equal to a query concept, we can get following similarity measure,

$$S_2^A(c_1, c_2) = \begin{cases} 1, & c_1 \text{ is an ancestor of, or equal to } c_2 \\ \gamma \times S(c_1, c_2), & \text{else} \end{cases} \quad (3)$$

when setting γ to 0, this formula can simulate query expansion which expand query with all descendants of the query concepts.

The above measures can be used to compare two concept nodes (tree numbers). However, each MeSH descriptor corresponds to one or several nodes in the MeSH trees. When two descriptors are compared, there exist many similarities between the two sets of concept nodes. Therefore, these similarities should be aggregated to get the similarity between descriptors. An easy solution is to choose the maximum similarity among these similarities

$$S(m_1, m_2) = \max_{c \in T(m_1), c' \in T(m_2)} S^A(c, c') \quad (4)$$

where m_1, m_2 are descriptors from a query and a document respectively. $T(m_1)$ and $T(m_2)$ are the sets of corresponding tree numbers of m_1 and m_2 , and S^A is the asymmetrical semantic similarity between the two concept nodes. Following Azuaje's work [9], an alternative measure for two descriptors is defined as

$$S(m_1, m_2) = \frac{\sum_{c \in T(m_1)} \max_{c' \in T(m_2)} S^A(c, c') + \sum_{c' \in T(m_2)} \max_{c \in T(m_1)} S^A(c, c')}{|T(m_1)| + |T(m_2)|} \quad (5)$$

where $|\cdot|$ is the cardinality of a set. Our previous study indicates this measure can achieve better performance, and will be used in following experiments. Both measures have asymmetric property due to the use of S^A .

2.3 Semantic Similarity between Query and Document

In semantic-based retrieval, a query is defined as a set of descriptors, $Q = \{m\}$, and a document is given as $D = \{m'\}$. The similarity between a query and a document, and also the retrieval score can be the average of all the inter-descriptor similarities:

$$rsv_{MeSH}(Q, D) = S_{MeSH}(Q, D) = \frac{\sum_{m \in Q, m' \in D} S(m, m')}{|Q| \times |D|} \quad (6)$$

This measure tends to give small results. Alternatively, we can build measure based on the best conceptual matches between the two groups of concepts. Following Azuaje's measure [9], the similarity is defined as

$$rsv_{MeSH}(Q, D) = S_{MeSH}(Q, D) = \frac{\sum_{m \in Q} \max_{m' \in D} S(m, m') + \sum_{m' \in D} \max_{m \in Q} S(m, m')}{|Q| + |D|} \quad (7)$$

Considering the application in retrieval, the relevance score can be further simplified, by ignoring normalization and the comparison from document side, to

$$rsv_{MeSH}(Q, D) = \sum_{m \in Q} \max_{m' \in D} S(m, m') \quad (8)$$

2.4 Combination of Semantic-based and Text-based Search

To combine semantic-based retrieval and text-based retrieval, the score of each ranking should be normalized. Given a ranking, the normalized retrieval score of document D is given by

$$rsv'(Q, D) = \frac{rsv(Q, D) - min}{max - min} \quad (9)$$

where max and min are the maximum and minimum scores in this ranking. Then the normalized scores of text-based ranking and semantic-based ranking are combined to get the score as

$$rsv_{combined}(Q, D) = \omega \times rsv'_{text}(Q, D) + (1 - \omega) \times rsv'_{MeSH}(Q, D) \quad (10)$$

where ω is between 0 and 1, and determined by experiment.

3 Modality Classification with Multiple Kernel Learning

3.1 Multiple Kernel Learning

A normal SVM classifier is designed for two-class problem, and can treat with only a single kernel. Given n training samples $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$, where $\mathbf{x}_i \in \mathcal{X}$ is the input vector and label $y_i \in \{+1, -1\}$.

Support vector machines originate from linear classifiers and maximize the margin between samples of both classes. Introducing a feature mapping ϕ from the input

space \mathcal{X} to a reproducing kernel Hilbert space (RKHS) \mathcal{H} , linear classifiers in \mathcal{H} of the form

$$f(\mathbf{x}) = \mathbf{w}^\top \phi(\mathbf{x}) + b \quad (11)$$

which provides a rich set of flexible classifiers in \mathcal{X} . The parameters (\mathbf{w} , b) are determined by solving an equivalent dual optimization. The dual optimization depends only on inner products (similarities) of inputs which can be alternatively computed by means of kernel functions k , given by

$$k(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle_{\mathcal{H}} \quad (12)$$

And the final decision function can be written as

$$f(\mathbf{x}) = \sum_{i=1}^n \alpha_i k(\mathbf{x}, \mathbf{x}_i) + b \quad (13)$$

The multiple kernel learning framework extends the regular SVM formulation by an adaptively-weighted combined kernel which fuses different kinds of features. The combined kernel is as follows:

$$k(\mathbf{x}, \mathbf{x}') = \sum_{j=1}^M \beta_j k_j(\mathbf{x}, \mathbf{x}') \quad \text{with } \beta_j \geq 0, \sum_{j=1}^M \beta_j = 1 \quad (14)$$

where β_j is weight to combine M sub-kernels $k_j(\mathbf{x}, \mathbf{x}')$. MKL can estimate optimal weights from training data. Sonnenburg et al. [10] proposed an efficient algorithm of MKL to estimate optimal weights and SVM parameters simultaneously by iterating training steps of a normal SVM. This implementation is available as the SHOGUN machine learning toolbox. For medical modality classification with image features, firstly, one sub-kernel for each image features are prepared, then the weights are estimated by the MKL method, finally, the optimal combined kernel is obtained. In the experiment, we used the MKL library included in the SHOGUN toolbox as the implementation of MKL.

3.2 Visual Features for Modality Classification

In our study, we tested five image features, which are described as following.

Gray Value Histogram (GH): The gray value histogram of an image represents the distribution of the pixels in the image over the gray-level scale. 32 bins are used in our study.

Edge Histogram Descriptor (EHD): The edge histogram descriptor [11] represents the spatial distribution of five types of edges, namely vertical, horizontal, 45-degree diagonal, 135-degree diagonal, and non-edge types.

Tamura Texture Feature (Tamura): Based on the research of textural features corresponding to human visual perception, Tamura et al.[12] proposed six basic textural features, namely, coarseness, contrast, directionality, line likeness, regularity, and roughness. The coarseness, contrast and directionality features are used in this study.

Gabor Texture Feature (Gabor): It has been proposed that Gabor filters can be used to model the responses of the human visual system, and Gabor filter based approaches are popular for texture feature extraction. We use an implementation

proposed by Manjunath et al [13]. The feature is built by filtering the image with a bank of orientation and scale sensitive filters and computing statistic measures of the output in the frequency domain.

SIFT: SIFTs [14] are local features and designed to describe an area of an image so to be robust to noise, illumination, scale, translation and rotation changes. For medical image classification, the SIFT rotation-invariance is not relevant, as the various structures in the radiographs are likely to appear always with the same orientation. Moreover, the scale is not likely to change too much between images of the same class. So, we ignore the scale- and rotation-invariance in our study.

4 Experiments and Results

Our previous study on OHSUMED data indicated that the asymmetric measure (equation 2) with the punishment factor γ set to 0.5 gets the best results, and it was used in our submitted runs. Since we hadn't study our method on previous ImageCLEF data when submitting the runs, we set the other parameters by guess. For ad-hoc retrieval, we constructed MeSH query representation with the top 5 descriptors returned by MeSHUp, and MeSH document with the top 10 descriptors, and set the combining parameter ω to 0.6. For case-based retrieval, if the article had MeSH terms assigned to it, we used them as the MeSH document. Otherwise, we constructed both MeSH query and document with the top 10 returned descriptors, and set the combining parameter ω to 0.8. These guessed setting resulted in unpromising performances in this campaign. When preparing this report, we trained our method on ImageCLEF 2010 ad-hoc retrieval data [2], and the best performance was achieved by the asymmetric measure (equation 2) while using the top 10 descriptors for both query and document and setting ω to 0.15. We then tested our method with the best settings on both ad-hoc and case-based retrieval. All our methods are implemented on Lemur toolkit [15].

4.1 Ad-hoc Retrieval

For ad-hoc image retrieval, we test two document representations 'c' and 'tc', which respectively uses only the caption of each image, and it plus the title of the article. When combining text-based retrieval with MeSH-based retrieval, the run name is also suffixed with 'm'. We also considered use modality classification information to enhance retrieval, and such runs have the name suffixed of 'mc'. All six submitted runs are prefixed with 'UESTC'. Besides the submitted runs, we also report two runs with the parameters trained on ImageCLEF 2010 ad-hoc retrieval data. One is with the asymmetric similarity measure (asym), and the other is with symmetric measure (symm). The text-based retrievals in all runs are based on TF-IDF model. The results of all runs are listed in the Table 1. It indicates that our methods with trained parameters can achieve promising results, and the one with asymmetric measure achieves the best performance, which can rank sixth among all textual runs of this year.

Table 1. Retrieval performance of ad-hoc retrieval runs

Runs	MAP	bpref	P@10
ad-tcm-asym	0.1939	0.1761	0.3182
ad-tcm-symm	0.1817	0.1628	0.3125
UESTC-ad-tc	0.1769	0.1584	0.3
UESTC-ad-c	0.1443	0.1446	0.2409
UESTC-ad-tcm	0.1434	0.1397	0.2182
UESTC-ad-cm	0.106	0.1154	0.2091
UESTC-ad-tcm-mc	0.101	0.1223	0.2
UESTC-ad-cm-mc	0.0653	0.0846	0.1727

Table 2. Retrieval performance of case-based retrieval runs

Runs	MAP	bpref	P@10
case-fm-asym	0.1437	0.1228	0.1411
case-fm-symm	0.1368	0.1164	0.1356
UESTC_case_f	0.1288	0.1092	0.1231
UESTC-case-fm	0.1269	0.1117	0.1231

Table 3. Features used for different runs

Runs	Features
UESTC-SIFT.txt	SIFT
UESTC-MKL2.txt	SIFT+Gabor
UESTC-MKL3.txt	SIFT+Gabor+EHD
UESTC-MKL5.txt	SIFT+Gabor+EHD+GH
UESTC-MKL6.txt	SIFT+Gabor+EHD+GH+Tumura

Table 4. Accuracies of modality classification runs

Runs	Accuracy (%)
UESTC-MKL3.txt	57.9
UESTC-MKL2.txt	56.7
UESTC-MKL5.txt	56.0
UESTC-MKL6.txt	56.0
UESTC-SIFT.txt	52.8

4.2 Case-Based Retrieval

For case-based retrieval, the full text of an article is used for the text-based document representation, and denoted by ‘f’. Similar to ad-hoc retrieval, we also report two runs with the trained parameters. And the text-based retrievals in all runs are also based on TF-IDF model. All results are listed in the Table 2. It indicates that

our methods with trained parameters can achieve promising results, and the one with asymmetric measure achieves the best, which can rank 3rd among all case-based runs of this year.

4.3 Modality Classification

Table 3. describes the features used for different runs, and Tables 4. lists all accuracies of different runs. It indicates combining SIFT, Gabor and EHD features achieves the best accuracy among our submitted runs, which ranks equal 3rd among all visual methods of this year.

5 Conclusions

This paper describes our contribution to the ImageCLEF 2012 medical retrieval task. We combine semantic-based retrieval with text-based retrieval for medical retrieval. For semantic-based retrieval, asymmetric similarity measures are also proposed to compare query concepts and document concepts. Our submitted runs for ad-hoc and case-based retrieval are based on guessed parameters and perform unpromisingly. But our new runs with the trained parameters perform well. This validates our methods. For modality classification, we use MKL to combine different visual features and achieve good results.

Acknowledgments. This research is partly supported by the National Science Foundation of China under grants 60873185.

Reference

1. Henning Müller, Alba Garcia Seco de Herrera, Jayashree Kalpathy-Cramer, Dina Demner Fushman, Sameer Antani, Ivan Eggel, Overview of the ImageCLEF 2012 medical image retrieval and classification tasks, CLEF 2012 working notes, Rome, Italy, 2012.
2. Wu H., Sun K. Improving Context-Based Medical Image Retrieval by Incorporating Semantic-Based Retrieval. In International Conference on Internet Multimedia Computing and Service, ICIMCS 2012, (2012). to appear
3. MeSH: <http://www.nlm.nih.gov/mesh/meshhome.html>
4. Lemur: <http://www.lemurproject.org>
5. Trieschnigg D., Pezik P., Lee V., de Jong F., Kraaij W., Rebholz-Schuhmann D. MeSH Up: effective MeSH text classification for improved document retrieval. *Bioinformatics*. 25, 11 (Jun. 2009) 1412-1418.
6. Wu H., Zhang H., Li C. Medical Image Classification with Multiple Kernel Learning. In the 2nd International Conference on Internet Multimedia Computing and Service, ICIMCS 2010, (2010)
7. Zhang, X., Jing L., Hu X., Ng K.M. A comparative study of ontology based term similarity measures on PubMed document clustering. In DASFAA'07, (2007) 115–126.

8. Li, Y., Bandar, Z. A. and McLean D., An Approach for Measuring Semantic Similarity between Words Using Multiple Information Sources. *IEEE Trans. Knowl. Data Eng.*, 15, 4 (2003), 871-882.
9. Azuaje, F., Wang, H., and Bodenreider, O. (2005). Ontology-driven similarity approaches to supporting gene functional assessment. In *ISMB'2005 SIG meeting on Bio-ontologies (2005)* 9-10.
10. Sonnenburg S., Rätsch G., Schäfer C., Schölkopf B. 2006. Large scale multiple kernel learning. *Journal of Machine Learning Research*, 7 (July 2006) 1531–1565.
11. Won C.S., Park D.K., and Park S.-J., Efficient use of MPEG-7 edge histogram descriptor, *ETRI Journal*, 24, (2002) 23-30.
12. Tamura H., Mori S., and Yamawaki T. 1978. Texture features corresponding to visual perception. *IEEE Trans. On Systems, Man, and Cybernetics*, 8, 6 (June, 1978), 460-473.
13. Manjunath B. and Ma W. 1996. Textures for browsing and retrieval of image data. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 18, 8 (Aug. 1996), 837-842.
14. Lowe, D.G. 1999. Object recognition from local scale-invariant features. In *Proceedings of International Conference on Computer Vision, ICCV'99*. 2, 1150–1157.