Improving Context-Based Medical Image Retrieval by Incorporating Semantic-Based Retrieval

Hong Wu
School of Computer Sci. & Eng.,
University of Electronic Science & Technology of China
Chengdu 611731, P.R.China
hwu@uestc.edu.cn

Kuangkai Sun
School of Computer Sci. & Eng.,
University of Electronic Science & Technology of China
Chengdu 611731, P.R.China
sunkuangkai@163.com

Zhongliu Zhuo
School of Computer Sci. & Eng.,
University of Electronic Science & Technology of China
Chengdu 611731, P.R.China
zhuozhongliu@126.com

ABSTRACT

The rapid increasing amount of medical images has led to demands for more effective and efficient medical image retrieval technology. In recent years, context-based retrieval has begun to attract more research interest due to the limitation of current content-based technology. Traditional text retrieval methods determine relevance of documents based on term matching, and suffer from the severe ambiguity problem existed in biomedical domain. This paper proposes to combine semantic-based retrieval with traditional text-based retrieval for context-based medical image retrieval. In our approach, each query or document has a text-based representation and a concept-based representation. For semantic-based retrieval, semantic similarity measure is used to comparing query concepts and document concepts, and asymmetric similarity measures are also proposed by modifying the existing symmetric measures. Then the inter-concept similarities are aggregated to compute the relevance score of a document. Finally, this semantic-based retrieval is combined with text-based retrieval. Our approach is evaluated on ImageCLEFmed 2010 dataset, which contains more than 77,000 images and their captions from online medical journals. The experimental results indicate incorporating semantic-based retrieval can improve the performance of context-based medical image retrieval, and our asymmetric semantic similarity measures can achieve better MAP than symmetric ones.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Retrieval models; J.3

General Terms

Algorithms, Measurement, Performance, Experimentation.

Keywords

Medical Image Retrieval; Context-Based Retrieval; Asymmetric Semantic Similarity; MeSH.

1. INTRODUCTION

In the last decades, digital medical images have played an increasingly important role in diagnosis, treatment planning and education [1], and the rapid increasing amount of data is stored in PACS, published as online collections or in the online content of journals. This large volume of image data produces a strong need for effective medical image retrieval technology.

ICIMCS'12, September 9–11, 2012, Wuhan, Hubei, China.

Traditional medical image retrieval relies on manual image annotation, which is an expensive and labor intensive procedure. And due to the semantic gap problem, content-based retrieval has limited retrieval performance for heterogeneous dataset. Consequently, context-based image retrieval has begun to attract more research interest in recent years. In context-based medical image retrieval, the text information that describes the images is automatically extracted from medical reports or articles, and used to retrieve images. Most traditional text retrieval methods depend on term matching between query and documents, and will suffer from the severe ambiguity problems in medical domain. Much effort has been made to integrate controlled vocabularies or ontologies, such as MeSH [2], UMLS [3] etc., to improve the performance of biomedical text retrieval. In medical image retrieval applications, the main effort has been taken on query expansion with biomedical ontologies for context-based retrieval or multi-model retrieval, mainly with ImageCLEF track data. With conceptual indexing model, Diem et al. [4] expanded both queries and documents based on *is-a* semantic relation in UMLS, and achieved notable improvement. Díaz-Galiano et al. [5] proposed to expand query with synonyms in MeSH, and improved not only the performances of a text-based retrieval system but also that of a textual and visual combined retrieval system. In the medical image retrieval task of ImageCLEF 2008, they compared query expansion with MeSH and UMLS [6]. The query expansion with MeSH got good results, but that with UMLS obtained worse results than the baseline. But in ImageCLEF 2010, their method with MeSH also got a decline in retrieval performance [7].

All the above mentioned methods use term-matching retrieval mechanism, and do not consider the meaning of the words or the semantic relationship between concepts. In this paper, a semantic-based retrieval model is proposed and integrated with text-based model for context-based medical image retrieval. In our approach, earch query or document has two representations, a word-based representation and a concept-based representation, which are indexed separately. The concept-based representations are constructed by mapping text of a query or document to MeSH descriptors (concepts) with MeSHUP [8], a MeSH classification system. In the semantic-based retrieval model, the query concepts and document concepts are compared with semantic similarity measure. And we also proposed asymmetrical similarity measures which can be designed by modifying existing symmetric measures. 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 retrieval models are combined to retrieve images. Our retrieval method is implemented on Lemur toolkit [9], and evaluated on ImageCLEFmed 2010 dataset [10].

This paper is structured as follows. In section 2, some related works are introduced. Our context-based medical image retrieval method is presented in section 3, and experiments are given in section 4, followed by conclusions in section 5.

2. RELATED WORKS

2.1 MeSH

MeSH [2], developed by the NLM, mainly consists of the controlled vocabulary and MeSH Trees. Its primary purpose is to support indexing, cataloguing, and retrieval of medical literature articles stored in MEDLINE database. The controlled vocabulary contains several types of terms such as Descriptor (Main Heading), Qualifiers (Subheading), Publication Types, Geographics, and Entry terms. Descriptor terms are main concepts or main headings. Entry terms are the synonyms or the related terms to descriptors. There are 26,142 descriptors, and over 177,000 entry terms in MeSH 2011. MeSH descriptors are organized in 16 categories: category A for anatomic terms, category B for organisms, C for diseases, D for drugs and chemicals, etc. Each category is further divided into subcategories. Within each subcategory, descriptors are arranged hierarchically from most general to most specific in up to 12 hierarchical levels. Each MeSH descriptor appears in one or more places in these trees, and is correspondingly assigned with one or more tree numbers.

2.2 Mapping Free Text to MeSH Descriptor

The mapping of text to MeSH descriptors is a large multi-class and multi-label text classification problem: one or more of MeSH descriptors can be assigned to a piece of text. Many researchers have developed MeSH classification techniques. Some research works adopted sophisticated techniques, but were limited to a subset of MeSH descriptors. Here, the works which offer a complete solution is our concern. In MetaMap [11], thesaurus concepts are found by first parsing the text into simple noun phrases and then by matching a large number of generated variants to the entries in the UMLS metathesaurus. Ruch et al. [12] propose a classifier, EAGL, based on a retrieval system. For each MeSH term, its synonyms and description are indexed as a document in a retrieval index. A piece of text is classified with the top-1 MeSH 'documents' returned by the retrieval system. MeSHUp [8] uses the available manual assignments of MeSH terms to citations as training data to build KNN classifiers. And MTI [13], provided to registered users by the NLM, incorporates different classifiers. In this work, we use MeSHUp for its superior performance.

2.3 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 [14], 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 [15], which combines the shortest path and the depth of ontology information in a non-linear function:

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

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.

3. OUR APPROACH

3.1 Overview

In our approach, each document or query has two representations, a word-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 with MeSHUP [8], 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 word-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 to retrieve medical images. Following section will give more details of our approach.

3.2 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

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

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,

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

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 exit 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 [16], 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.

3.3 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 [16], 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 simplied, by ignoring normalization and the comparison from document side, to

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

3.4 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

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

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

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

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

4. EXPERIMENT

ImageCLEFmed is a medical image retrieval task within CLEF campaign. The dataset of image retrieval task in ImageCLEFmed 2010 [3] was used for our experiments. It contains more than 77,000 images from articles published in *Radiology* and *Radiographics*, together with the text of the captions and a link to the html of the full text articles. The image captions were used as the context information in this study. Our experiments include two phases. The first one is to conduct text-based retrieval and semantic-based retrieval separately, and the second one is to test the methods which combining the two retrievals. The experiments were based on the 16 short queries and the ground-truth of the retrieval task. The retrieval performance was measured using Mean Average Precision and Precision at 10.

Our retrieval system is implemented with Lemur toolkit. Vector space model and TF-IDF weighting are used for text-based retrieval. For semantic-based retrieval, Li's measure (equation(1)) is used as the base semantic similarity measure, and equation (5) is used to compute similarity between descriptors. And three values (1, 0.5 and 0) are tested for the punishment factor γ. In addition, we compared constructing concept-based representation of query and document with the top 5, 10, 15 descriptors returned by MeSHUp respectively, and found that using top 10 descriptors for both query and documents gave the best MAP on ImageCLEFmed 2010 dataset. We also compare two methods to computed relevance score of a document, equation (7) and (8), and found the un-normalized one (equation (8)) gave the best results in all setting. Due to the page limitation, we do not give the results of these comparisons.

4.1 Text-Only and Semantic-Only Retrieval

We first conduct text-based retrieval and semantic-based retrieval separately. The text-based search which only uses the word-based representation serves as baseline. The MAP and P@10 of this run are 0.2863 and 0.3500 respectively. In semantic-based retrieval experiments, we compare the symmetric similarity (Symm) and the asymmetric similarity (Asym1 for equation (2), Asym2 for equation (3)) for semantic-based retrieval.

The experimental results are listed in table 1. They indicate that all semantic-based retrieval runs perform worse than text-only retrieval. This phenomenon implies that automatically generated MeSH descriptors alone can not convey the most meaning of short texts. For the asymmetric measures, the best MAP is 0.0765, and best P@10 is 0.1878, both of which are achieved by Asym1 with γ=0.5. These results are better than those of the symmetric measure, which achieves MAP of 0.0712, and P@10 of 0.1688. But the asymmetric measures with other punishment values get worse performances than the symmetric measure.

4.2 Fusing Semantic-Based and Text-based Retrieval

Then we compare different methods which using both text information and MeSH descriptors. The test runs include the methods introduced in section 3.4 with different semantic similarity measures and an ad-hoc method, QEDE, which expends query and document with MeSH descriptors and use text-based retrieval. For methods which combine text-based retrieval and semantic-based retrieval (in section 3.4), the combining parameter ω is tuned by experiments. The experimental results are listed in table 2. The results show that, all our tested methods outperform the baseline, and all methods which incorporating semantic-based retrieval outperform the ad-hoc method not only on MAP but also P@10. Among methods incorporating semantic-based retrieval,

Table 1. Comparison of Semantic-Based Retrievals

Methods	MAP	P@10
Symm	0.0712	0.1688
Asym1($\gamma = 0.5$)	**0.0765**	**0.1878**
Asym1($\gamma = 0$)	0.0613	0.1618
Asym2($\gamma = 1$)	0.0656	0.1625
Asym2($\gamma = 0.5$)	0.0683	0.1625
Asym2($\gamma = 0$)	0.0597	0.1603

Table 2. Comparison of Combined Methods

Methods	MAP	P@10
Baseline	0.2863	0.3500
QEDE	0.2973	0.3615
Symm	0.3217	**0.3875**
Asym1($\gamma = 0.5$)	**0.3305**	0.3675
Asym1($\gamma = 0$)	0.3068	0.3454
Asym2($\gamma = 1$)	0.3125	0.3525
Asym2($\gamma = 0.5$)	0.3198	0.3620
Asym2($\gamma = 0$)	0.3105	0.3458

Asym1 with γ=0.5 achieves the best MAP at 0.3305, and Symm achieves the best P@10 at 0.3875. Among all asymmetric measures, Asym1 with γ=0.5 achieves both the best MAP and P@10. We do not list the result of Díaz-Galiano's expansion method [7], because their result is worse than their baseline, and they use caption and title as the document.

5. CONCLUSIONS

We propose a semantic-based retrieval model for improving context-based medical image retrieval. In this semantic retrieval model, the text of query and documents are mapped to MeSH descriptors to construct concept-based representations. The concepts from query and document are compared with semantic similarity measure, and asymmetric measures are also proposed by modifying the existing symmetric measures. Then the inter-concept similarities are aggregated to compute the relevance score of a document. Finally, this semantic-based retrieval is combined with traditional text-based retrieval. Our approach is evaluated on ImageCLEFmed 2010 dataset, which contains more than 77,000 images and their captions from online medical journals. The experimental results indicate incorporating semantic-based retrieval can improve the performance of context-based medical image retrieval, and our asymmetric semantic similarity measures can achieve better MAP.

6. ACKNOWLEDGMENTS

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

7. REFERENCES

[1] Müller, H., Michoux, N., Bandon, D., Geissbuhler, A.: A Review of Content-Based Image Retrieval Systems in Medical Applications-Clinical Benefits and Future Directions. Int J Med Inform. 73, 1 (Feb. 2004), 1-23.

[2] MeSH: http://www.nlm.nih.gov/mesh/meshhome.html

[3] UMLS: http://www.nlm.nih.gov/research/umls

[4] Le, D.T.H., Chevallet, J., Thuy, D.T.B.: Thesaurus-Based Query and Document Expansion in Conceptual Indexing with UMLS: Application in Medical Information Retrieval. In *Proceedings of IEEE International Conference on Research, Innovation and Vision for the Future* (2007) RIVF'07, 242-246

[5] Díaz-Galiano, M.C., Martín-Valdivia, M.T., Ureña-López, L.A.: Query Expansion with a Medical Ontology to Improve a Multimodal Information Retrieval System. Computers in Biology and Medicine, 39, 4 (April 2009), 396-403

[6] Díaz-Galiano, M.C., García-Cumbreras, M.A., Martín-Valdivia, M.T., Ureña-López, L.A., Montejo-Ráez, A.: SINAI at ImageCLEFmed 2008. In *On-lineWorking Notes, CLEF 2008* (2008)

[7] Díaz-Galiano, M.C., Martín-Valdivia, M.T., Montejo-Ráez, A. , García-Cumbreras, M.A.: SINAI at ImageCLEFmed 2010. In *On-lineWorking Notes, CLEF 2010* (2010)

[8] 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.

[9] Lemur: http://www.lemurproject.org

[10] Müller, H., Kalpathy-Cramer, J., Eggel, I., Bedrick, S., Reisetter, J., Kahn Jr., C., and Hersh, W. Overview of the CLEF 2010 medical image retrieval track. In *Working Notes of CLEF 2010* (Padova, Italy, 2010).

[11] Aronson,A.R. Effective mapping of biomedical text to the UMLS metathesaurus: the MetaMap program. In *Proceedings of AMIA Symp* (Washington DC, USA, 2001) 17–21.

[12] Ruch,P. Automatic assignment of biomedical categories: toward a generic approach. Bioinformatics, 22, 6 (Mar. 2006) 658-664.

[13] MTI: http://skr.nlm.nih.gov/interactive/mti.shtml

[14] 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.

[15] 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.

[16] Azuaje, F., Wang, H., and Bodenreider, 0. (2005). Ontology-driven similarity approaches to supporting gene functional assessment. In ISMB'2005 SIG meeting on Bio-ontologies (2005) 9-10.