Medical Case Retrieval

Mario Taschwer
Institute of Information Technology (ITEC)
Alpen-Adria-Universität Klagenfurt, Austria
mario.taschwer@aau.at

ABSTRACT

The proposed PhD project addresses the problem of finding descriptions of diseases or patients' health records that are relevant for a given description of patient's symptoms, also known as medical case retrieval (MCR). Designing an automatic multimodal MCR system applicable to general medical data sets still presents an open research problem, as indicated by the ImageCLEF 2013 MCR challenge, where the best submitted runs achieved only moderate retrieval performance and used purely textual techniques. This project therefore aims at designing a multimodal MCR model that is capable of achieving a substantially better retrieval performance on the ImageCLEF data set than state-of-the-art techniques. Moreover, the potential of further improvement by leveraging relevance feedback of medical expert users for long-term learning will be investigated.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval; I.5.2 [**Pattern Recognition**]: Design Methodology; H.4.2 [**Information Systems Applications**]: Types of Systems—*decision support*; J.3 [**Life and Medical Sciences**]: medical information systems; I.2.6 [**Artificial Intelligence**]: Learning

General Terms

Algorithms; Design; Experimentation; Measurement; Performance

Keywords

biomedical information retrieval; multimodal information retrieval; ImageCLEF medical tasks

1. PROBLEM DESCRIPTION

Medical case retrieval (MCR) is the problem of finding descriptions of diseases or patients' health records (document corpus) that are relevant for a given description of patient's symptoms (query), as decided by medical experts. MCR is a major building block of clinical decision support systems [15] employing the paradigm of case-based reasoning [3], where the "most similar" medical cases need to be retrieved for a given symptom description before diagnosis and treatment can be proposed by the system. Moreover, MCR is also a relevant problem in medical education and research, because it allows to select interesting cases for students and to generate data sets for studies meeting case-based criteria.

Case and symptom descriptions are multimedia documents, typically consisting of structured text and medical images. Designing an automatic MCR system applicable to general medical data sets (as opposed to data sets in narrow medical domains, see [3]) still presents an open research problem. The ImageCLEF evaluation campaign[1] [19] issued a yearly MCR challenge between 2009 and 2013, leading to a general biomedical data set of about 75,000 documents (case descriptions) and 35 queries (symptom descriptions) in 2013. The moderate retrieval performance of the best MCR runs submitted to ImageCLEF 2013 and their purely textual techniques emphasize the need for further research regarding multimodal MCR techniques [9].

This PhD project aims at developing an MCR model that is able to substantially improve retrieval performance on the ImageCLEF MCR data set upon state-of-the-art retrieval systems. Moreover, in the context of case-based reasoning, the proposed MCR model should support long-term learning from the relevance feedback of medical expert users.

2. STATE OF THE ART

The narrow research field of medical case retrieval (MCR) can be positioned at the intersection of three larger areas of artificial intelligence research: *multimedia information retrieval* provides techniques to index and retrieve multimedia documents; *knowledge representation* research produces models to capture medical expert knowledge; and *computer vision* methods specialized to medical images are needed to extract discriminative features and detect semantic concepts. Corresponding domain-specific research fields are predominant subjects of medical informatics [14].

From these research areas, multimedia IR (information retrieval) is the most relevant one for the proposed PhD topic. We categorize its relevant techniques into four groups: *text retrieval* (classical IR), *visual retrieval* (content-based image and video retrieval), *data fusion* (combining several

[1] `http://www.imageclef.org/`

MM'14, November 3–7, 2014, Orlando, FL, USA.

http://dx.doi.org/10.1145/2647868.2654856.

IR systems or information sources), and *relevance feedback* (learning from users to improve IR).

Text retrieval research has developed various IR models during the last four decades. The most prominent ones are the *vector space model* [26], the *probabilistic model* [23], language models [18], and divergence from randomness [2]. The latter has been found to be the most effective model on a biomedical data set [1].

A fundamental limitation of text retrieval systems is the mismatch of words used to express the same concepts in the query and in the document collection, known as the *vocabulary problem* in IR. It has been addressed by several techniques in IR research. For MCR, query expansion [5] and relevance feedback [25] are most relevant, as they allow to inject medical expert knowledge into the retrieval process.

In the context of this work, **visual retrieval** is mainly concerned with content-based image retrieval (CBIR) [8]. Typically, the user of a CBIR system expects semantic similarity of images – depending on the user's context and application domain –, which is hard to express in terms of digital image properties. This discrepancy is known as the *semantic gap* [27]. It is still an open problem in many application domains of CBIR, including the medical domain [34]. A review of CBIR in medical applications and its clinical benefits is given by Müller et al. [20].

The core techniques of the CBIR process include: feature extraction, visual signature building, and applying similarity measures to visual signatures. Due to the wealth of *image features* and *similarity measures* proposed in the literature [8], extensive experimental comparisons on different data sets, including medical ones, are valuable [10, 12]. Whereas global features (describing an entire image) are often used directly as *visual signatures*, pattern recognition techniques are applied to local features (describing an image region) to build the signature. In an attempt to reduce the semantic gap, visual signatures can be represented in terms of *semantic concepts* detected in the image or video [21, 28]. The IRMA code [17] provides a visual signature for diagnostic images based on coarse semantic categories.

Combining text and visual retrieval leads to the more general problem of **data fusion** [31]. The objective is to combine several information sources to improve retrieval effectiveness, either at the feature level (early fusion) or at the decision level (late fusion). An overview of late fusion techniques based on ranks or scores of individual result lists as well as proposals for linear score combination and score normalization are given by Wu [32]. A recent comparison of late fusion techniques on the ImageCLEF 2013 MCR data set showed that linear score combination of text and image retrieval is superior over individual retrieval [13]. Rahman et al. [22] successfully applied fusion techniques to multimodal biomedical image retrieval.

Relevance feedback (RF) is not only a technique to overcome the vocabulary problem in text retrieval, but it also addresses the semantic gap problem in visual retrieval [8, 33]. As it works by incorporating interactive relevance judgments from users during the search process, it may provide a useful approach to learn from medical expert users of an MCR system. Whereas *short-term learning* affects the current query only, *long-term learning* aims at improving retrieval for future queries [7] and provides a more promising avenue for MCR. Because expert users typically label only a few results suggested by the system, advanced RF approaches make use of *semi-supervised learning* [6].

Medical **knowledge representations**, particularly ontologies or controlled vocabularies [24] (e.g. UMLS[2], MeSH[3]), can help improve medical multimedia IR. Two common techniques are *query expansion* [4, 11] and *semantic annotation* of images [16].

3. OBJECTIVES AND CONTRIBUTION

As explained in previous sections, *medical case retrieval* (MCR) is a relevant problem in computer science whose known solutions for large and heterogeneous data sets are too ineffective to be of practical value. This PhD project therefore aims at designing an MCR model that is able to deliver a substantially better retrieval performance on such data sets than known solutions. Moreover, the proposed techniques need to be both efficient and robust to be applicable to large medical data sets of diverse content.

Considering the retrieval methods used for the ImageCLEF MCR challenge in 2013 (see Section 1), the most obvious potential for improvement seems to be in data fusion methods combining different modalities of case representation (text and images) and external knowledge. The focus of this work will therefore be on *multimodal* approaches to MCR. However, in order to achieve successful data fusion with visual modalities, improvement of visual-only retrieval performance is a necessary goal, too.

More precisely, this PhD project will address the following research objectives constituting its scientific contribution: (**O1**) Determine the reasons for the moderate retrieval performance of current multimodal techniques on the ImageCLEF MCR data set. (**O2**) Design a novel MCR model combining different modalities of case representations and information sources (without relevance feedback) to enable a substantial improvement of retrieval performance on the ImageCLEF 2013 MCR data set, achieving at least 30% MAP. (**O3**) Using the system resulting from O2, investigate the potential of further improvement of retrieval performance by long-term learning from medical expert users.

4. APPROACH

As there is no other recent comparative study of multimodal MCR techniques than that of ad-hoc solutions to the ImageCLEF MCR challenge [9], pursuing objective **O1** requires selecting, implementing, evaluating, and analyzing some of the most promising known approaches to MCR. Moreover, to understand the reasons for their retrieval performance, statistical properties of features extracted from the ImageCLEF MCR data set need to be investigated.

Designing a better MCR model according to objective **O2** includes the sub-problems of (1) choosing well-performing text retrieval techniques, (2) improving visual retrieval performance, (3) utilizing medical knowledge for text and visual retrieval, and (4) combining text and visual retrieval by data fusion methods. The semantic gap problem can be addressed by detecting semantically meaningful *concepts* in multimedia documents using pattern recognition techniques. It is hoped that semantic case similarity can be expressed in a simpler and more robust way in terms of concepts than in terms of document features. Meaningful concepts can

[2]`http://www.nlm.nih.gov/research/umls/`
[3]`http://www.nlm.nih.gov/mesh/`

be taken from medical thesauri (e.g. MeSH), ontologies, or medical image categorization systems (e.g. IRMA).

Proper evaluation of a long-term learning system according to **O3** would require a user study with several (e.g. 20) representative medical experts and an appropriate experimental design to derive statistically significant results. Due to the difficulty and costs of enlisting so many medical expert users, evaluation of the system according to **O3** will simulate experts by using part of the relevance judgments (ground truth) provided with the ImageCLEF MCR data set.

5. PREVIOUS AND FURTHER WORK

The PhD proposal has been accepted by AAU in March 2013[4]. We participated in the ImageCLEF 2013 MCR task using a simple text retrieval approach utilizing MeSH terms for query and document expansion [29]. We then extended the approach to include pseudo-relevance feedback, and recently we completed a technical report systematically evaluating more than 500 combinations of these textual MCR methods [30]. The best method combinations achieve state-of-the-art retrieval performance on the ImageCLEF 2013 MCR data set, as depicted in Figure 1.

Previous work established a basis for research objectives **O1** and **O2**, restricted to text retrieval and the utilization of a medical thesaurus (MeSH). The next project phase will extend work on these objectives to visual and concept-based retrieval. The incorporation of data fusion methods should provide results for **O1** and **O2** in fall 2015. The investigation of relevance feedback for long-term learning (**O3**) is planned to be finished in summer 2016, and the PhD thesis should be completed until end of 2016.

6. SUPERVISORS

Laszlo Böszörmenyi, Alpen-Adria-Universität Klagenfurt, Austria.
Oge Marques, Florida Atlantic University, Boca Raton, FL, USA.

7. REFERENCES

[1] S. Abdou and J. Savoy. Searching in Medline: Query expansion and manual indexing evaluation. *Inf. Process. Manage.*, 44(2):781–789, Mar. 2008.

[2] G. Amati and C. J. Van Rijsbergen. Probabilistic models of information retrieval based on measuring the divergence from randomness. *ACM Trans. Inf. Syst.*, 20(4):357–389, Oct. 2002.

[3] S. Begum, M. Ahmed, P. Funk, N. Xiong, and M. Folke. Case-based reasoning systems in the health sciences: A survey of recent trends and developments. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 41(4):421–434, July 2011.

[4] J. Bhogal, A. Macfarlane, and P. Smith. A review of ontology based query expansion. *Inf. Process. Manag.*, 43(4):866–886, July 2007.

[5] C. Carpineto and G. Romano. A survey of automatic query expansion in information retrieval. *ACM Comput. Surv.*, 44(1):1:1–1:50, Jan. 2012.

[4] `https://www-itec.uni-klu.ac.at/~mt/2013/03/phd-proposal/`

[6] O. Chapelle, B. Schölkopf, and A. Zien. *Semi-Supervised Learning*. The MIT Press, 1st edition, 2010.

[7] M. Cord and P. H. Gosselin. Image retrieval using long-term semantic learning. In *Image Processing, 2006 IEEE International Conference on*, pages 2909–2912. IEEE, 2006.

[8] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv.*, 40(2):5:1–5:60, May 2008.

[9] A. G. S. de Herrera, J. Kalpathy-Cramer, D. Demner-Fushman, S. Antani, and H. Müller. Overview of the ImageCLEF 2013 medical tasks. In *Working notes of CLEF 2013*, Valencia, Spain, 2013.

[10] T. Deselaers, D. Keysers, and H. Ney. Features for image retrieval: an experimental comparison. *Inf. Retr.*, 11(2):77–107, Apr. 2008.

[11] M. C. Díaz-Galiano, M. Martín-Valdivia, and L. A. Ureña López. Query expansion with a medical ontology to improve a multimodal information retrieval system. *Comput. Biol. Med.*, 39(4):396–403, Apr. 2009.

[12] H. Eidenberger. Evaluation and analysis of similarity measures for content-based visual information retrieval. *Multimedia Systems*, 12(2):71–87, 2006.

[13] A. García Seco de Herrera and H. Müller. Fusion techniques in biomedical information retrieval. In B. Ionescu, J. Benois-Pineau, T. Piatrik, and G. Quénot, editors, *Fusion in Computer Vision*, Advances in Computer Vision and Pattern Recognition, pages 209–228. Springer International Publishing, 2014.

[14] R. Haux. Medical informatics: Past, present, future. *International Journal of Medical Informatics*, 79(9):599–610, 2010.

[15] K. Kawamoto, C. A. Houlihan, E. A. Balas, and D. F. Lobach. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *British Medical Journal*, 330(7494):765, 2005.

[16] C. Kurtz, A. Depeursinge, S. Napel, C. F. Beaulieu, and D. L. Rubin. On combining image-based and ontological semantic dissimilarities for medical image retrieval applications. *Medical Image Analysis*, 18(7):1082–1100, 2014.

[17] T. M. Lehmann, H. Schubert, D. Keysers, M. Kohnen, and B. B. Wein. The IRMA code for unique classification of medical images. In *Medical Imaging 2003*, pages 440–451. International Society for Optics and Photonics, 2003.

[18] X. Liu and W. B. Croft. Statistical language modeling for information retrieval. *Annual Review of Information Science and Technology*, 39(1):1–31, 2005.

[19] H. Müller, P. Clough, T. Deselaers, and B. Caputo, editors. *ImageCLEF – Experimental Evaluation in Visual Information Retrieval*, volume 32 of *The Information Retrieval Series*. Springer Berlin Heidelberg, 2010.

[20] H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler. A review of content-based image retrieval systems in medical applications – clinical

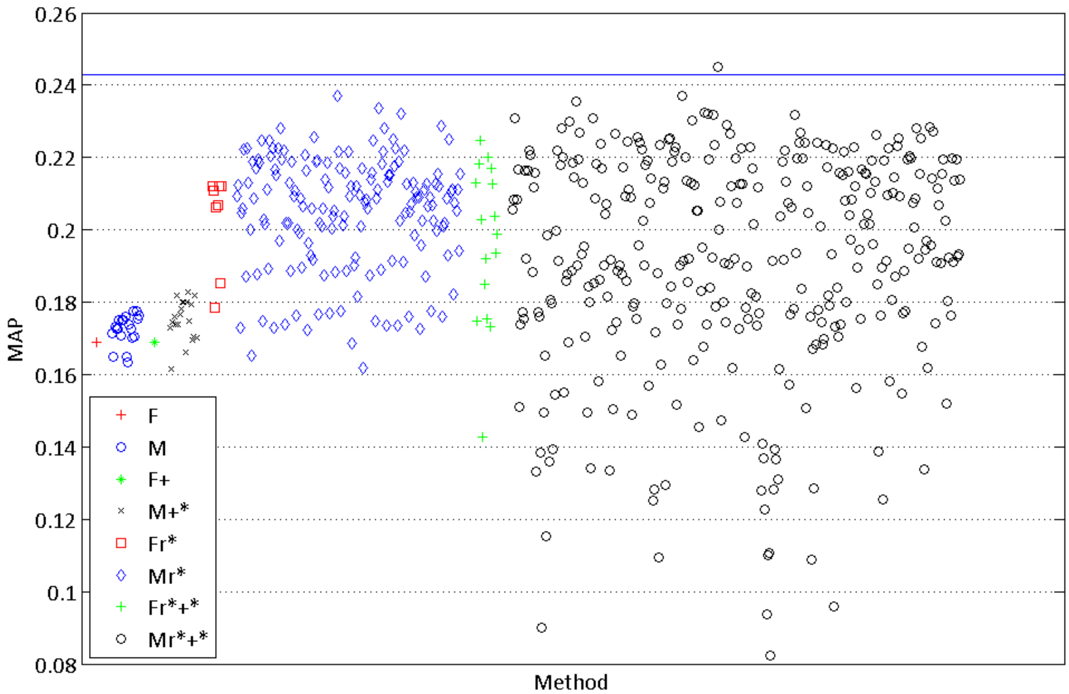

Figure 1: Scatter plot of 546 combinations of query and document expansion methods with parameters optimized on ImageCLEF 2012 dataset and evaluated on ImageCLEF 2013 MCR dataset. Method combinations are grouped according to utilized techniques: fulltext-only (F), MeSH query expansion (M), local feedback (r*), document expansion (+, +*). The horizontal line at MAP 0.2429 corresponds to the best run submitted to ImageCLEF 2013 [9].

benefits and future directions. *International Journal of Medical Informatics*, 73(1):1 – 23, 2004.

[21] M. Rahman, S. Antani, and G. Thoma. A learning-based similarity fusion and filtering approach for biomedical image retrieval using SVM classification and relevance feedback. *IEEE Trans. Inf. Tech. Biomedicine*, 15(4):640–646, July 2011.

[22] M. M. Rahman, D. You, M. S. Simpson, S. K. Antani, D. Demner-Fushman, and G. R. Thoma. Multimodal biomedical image retrieval using hierarchical classification and modality fusion. *International Journal of Multimedia Information Retrieval*, 2(3):159–173, 2013.

[23] S. E. Robertson and K. Sparck Jones. Relevance weighting of search terms. *Journal of the American Society for Information Science*, 27(3):129–146, 1976.

[24] D. L. Rubin, N. H. Shah, and N. F. Noy. Biomedical ontologies: a functional perspective. *Briefings in Bioinformatics*, 9(1):75–90, 2008.

[25] I. Ruthven and M. Lalmas. A survey on the use of relevance feedback for information access systems. *The Knowledge Engineering Review*, 18(02):95–145, 2003.

[26] G. Salton, A. Wong, and C. S. Yang. A vector space model for automatic indexing. *Commun. ACM*, 18(11):613–620, Nov. 1975.

[27] 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.*, 22(12):1349–1380, Dec. 2000.

[28] C. G. M. Snoek and M. Worring. Concept-based video retrieval. *Found. Trends Inf. Retr.*, 2(4):215–322, Apr. 2009.

[29] M. Taschwer. Text-based medical case retrieval using MeSH ontology. In P. Forner, R. Navigli, and D. Tufis, editors, *CLEF 2013 Evaluation Labs and Workshop, Online Working Notes*, 2013.

[30] M. Taschwer. Textual methods for medical case retrieval. Technical Report TR/ITEC/14/2.01, Institute of Information Technology (ITEC), Alpen-Adria-Universität Klagenfurt, Austria, May 2014. `http://www-itec.uni-klu.ac.at/bib/files/textual-mcr.pdf`.

[31] L. Valet, G. Mauris, and P. Bolon. A statistical overview of recent literature in information fusion. *IEEE Aerospace and Electronic Systems Magazine*, 16(3):7–14, 2001.

[32] S. Wu. Linear combination of component results in information retrieval. *Data Knowl. Eng.*, 71(1):114–126, Jan. 2012.

[33] X. S. Zhou and T. S. Huang. Relevance feedback in image retrieval: A comprehensive review. *Multimedia systems*, 8(6):536–544, 2003.

[34] X. S. Zhou, S. Zillner, M. Moeller, M. Sintek, Y. Zhan, A. Krishnan, and A. Gupta. Semantics and CBIR: a medical imaging perspective. In *Proceedings of the 2008 International Conference on Content-based Image and Video Retrieval*, CIVR '08, pages 571–580, New York, NY, USA, 2008. ACM.