

UNT at ImageCLEFmed 2009

Miguel E Ruiz

College of Information
Department of Library and Information Sciences
University of North Texas
1155 Union Circle 311068
Denton, Texas 76203-1068
USA

Abstract:

For this year our team participated in the medical image retrieval task. Most of our effort was invested in processing the collection using Metamap to assign Unified Medical Language System (UMLS) concepts to each of the images that included some associated text. This process generated metadata that was added to each image and included the UMLS concept number as well as the primary terms associated to each concept. Queries were also processed using Metamap to generate the corresponding UMLS concepts and terms associated with each query request. The SMART system was used to perform retrieval using a generalized vector space model that included the original text, the automatically assigned UMLS concepts, and the UMLS terms. We use a simple weighting scheme (tf-idf) to perform retrieval. Our text based runs included a simple run and a retrieval feedback run. The parameters for retrieval feedback and for the linear combination of the generalized vector space model were tuned using queries for the 2008 CLEF medical image retrieval task (imageCLEFmed). We also worked on using the results from the open source content-based image retrieval (CBIR) system GIFT but ran into some technical problems that prevented us to generate the retrieval results on time for the deadline. However, the University of Geneva (UG) team allowed us to use one of their Image results. The mixed results were generated using the GIFT run provided by UG and used a standard fusion mechanism by combining the text and CBIR results into a single list. To tune the parameters for the combination we used the results from the imageclefmed 2008 queries. Our results indicate that the pseudo relevance feedback mechanism yields only small improvements. The Combination of image features and text gave mixed results. While the combination of standard retrieval and CBIR yields small improvements, the combination of retrieval feedback and CBIR resulted in results significantly below using only text. At this point we still are investigating the reasons for this unexpected result.

Introduction

Retrieval of medical images from medical related literature is been used more and more frequently by physicians and other health related professions and seems to be a promising application for content based image retrieval systems. Our research for the CLEF medical image retrieval track consisted in evaluating the contribution of automatically generated UMLS terms and its impact in retrieval performance of images with caption or other associated text. We also explored the use of automatically extracted image features such as the ones generated by content-based image retrieval systems. We used the SMART retrieval system (Salton, 1971) for handling the text retrieval part of this task and the GNU Image Finding Tool (GIFT) for performing content-based image retrieval . We also use Metamap (Aronson, 2001), which the an open source tool developed by the National Library of

medicine, for automatic identification of medical terms compiled in the Unified Medical Language System (UMLS).

Section 1 of this paper presents a brief description of the systems used. Section 2 describes the way in which the text, images and queries were processed. Section 3 describes the runs that were submitted and section 4 presents our results and analysis. Finally section 5 presents our conclusions and future work.

1 System Description

The system used in these experiments combines three independent open source tools that have been extensively used in the literature. The SMART system, which was developed by G. Salton and his collaborators at Cornell University, is an information retrieval system that uses the vector space model to represent queries and documents and perform retrieval using a variety of weighting schemes (Salton, 1971). The version that we use was modified by Miguel Ruiz and currently supports 13 European languages using the ISO-Latin-I encoding. Smart also allows the use of multiple indexes that can be combined using a linear model into what can be considered a generalized vector space model (Ruiz & Southwick, 2006). As will be explained in more detail in the next section, documents and queries are represented using three different types of terms: Free text, UMLS concepts, and the terms from the UMLS vocabulary. Since each of these three types of terms have very different origin (one is author generated while the other two are automatically assigned), we decided to keep them in four separate indexes (called ctypes in SMART) and use the following linear model to compute similarity:

$$sim(Q, D) = \lambda \times sim_{title}(Q, D) + \delta \times sim_{caption}(Q, D) + \eta \times sim_{concepts}(Q, D) + \gamma \times sim_{terms}(Q, D)$$

where λ , δ , η , and γ are coefficients that control the contribution of the similarity of the text from the title, text from the captions, UMLS concepts and UMLS terms respectively. *sim* is the similarity score which is computed using the standard dot product of the query Q and document D vectors.

For processing the images and extracting features from the images we used the GNU Image Retrieval Tool (GIFT), which is an open source CBIR system created by the University of Geneva (GNU Image Finding Tool (GIFT), 2004). GIFT works also with a vector space model that uses automatically extracted features such as color, texture and shape. GIFT uses a tf-idf similarity to rank images according to the similarity with a single image. For our mixed runs we combined the image and text results using a linear combination with two parameters that control the contribution of the results from each system.

Metamap is an open source tool developed by Alan Aronson at the National Library of Medicine (NLM). Metamap takes free text and maps it to UMLS concepts using natural language processing and information extraction techniques (Aronson, 2001). This tool is offered to institutions that subscribe to UMLS Knowledge Sources which is provided by the NLM.

2. Document Processing

The data collection is described in more detail in the ImageCLEFmed overview paper (Müller, et al., 2009). We processed all 74,901 documents associated to the images by using the UMLS knowledge sources and then adding two new fields (UMLS concepts, and UMLS terms) to each document. For each term we also added the corresponding semantic type generated by Metamap. The following is an

example that shows a document with the added fields using the standard simple document format of SMART.

.I 28652

.C

C1281580; C0024487; C0205123; C0456079; C0066563; C0043100; C0005910;
C0230442; C0016068; C0025086; C1305866

.M

Fibula (Entire fibula); body_Part__Organ__or_Organ_Component

MRS (Magnetic Resonance Spectroscopy); Diagnostic_Procedure

Frontal (Coronal); Spatial_Concept

Level (Disease classification level); Classification

MR (Mineralocorticoid Receptor); Amino_Acid__Peptide__or_Protein_Receptor

Weight; Clinical_Attribute_Quantitative_Concept

WEIGHT (Body Weight); Organism_Attribute

Right leg (Structure of right lower leg); Body_Part__Organ__or_Organ_Component

Fibula (Bone structure of fibula); Body_Part__Organ__or_Organ_Component

Image (Medical Imaging); Diagnostic_Procedure

Weight (Weighing patient); Diagnostic_Procedure

.W

Figure 1b. (Figure continued.) (b) Coronal precontrast T1-weighted (700/20) spin-echo MR image of the right lower leg at the level of the fibula. (c) Coronal postcontrast T1-weighted (700/20) spin-echo MR image of the right lower leg at the same position as in b

.T

Case 12

The queries were also processed using Metamap and generating a similar representation as for the documents. The following is an example of one of the official queries after adding the UMLS concepts, and terms:

```
.I      1

.C

      C0041834; C0332476

.M

      Erythema; Sign_or_Symptom

      Erythematous; Functional_Concept

.W

      Photos of erythema

.T

      Photos of erythema
```

The images were processed using GIFT standard add-image routine. Unfortunately we had some technical problems before the deadline for submitting official results which did not allow us to create our own image runs. However, Henning Müller and the University of Geneva team supplied us with a visual run (also generated with GIFT) that we were able to use in our mixed runs. This was a simple run using 8 gray levels which is a type of run that has performed reasonably well in previous years.

3. Description of Runs and Parameters

Parameter tuning was done using the 2008 queries that use the same collection of this year. For text based runs we decided to generate two types of runs:

- a) Simple retrieval runs: using the free text, concepts and UMLS terms.
- b) Relevance feedback runs: automatic retrieval feedback using Rocchio's formula and assuming that the top n documents are relevant and expanding the queries with the top m terms.

For the relevance feedback runs we used the smart implementation of the Rocchio algorithm which applies query expansion on each of the indexes. In other words, our relevance feedback runs will generate expanded terms for each of the fields that are used in the retrieval model (title, captions, UMLS concepts and UMLS terms). The Rocchio formula requires several parameters such as the number of documents to be considered relevant in the automatic feedback (n), the number of terms to be added to each field (m) and the coefficients that control the contribution of the original query terms (α), the terms found in the relevant document (β) and the terms found in the non-relevant documents (δ). All these parameters were set by optimizing the retrieval performance on the 2008 queries.

The official text runs that we submitted are: one with retrieval feedback and one without retrieval feedback. We also submitted two mixed runs that are basically the combination of the previously described text based runs and the results from the CBIR system. As explained previously we used a linear combination of the text based and CBIR results are combined using a linear function that includes two coefficients that control the contribution of each run to the final results. These coefficients were also tuned using the 2008 queries which yield a 3:1 ratio for combining the text and CBIR results (the text based results contributed 3 times more than the CBIR results). This is consistent with previous setting that we have used in CLEF 2005, 2006 and 2007.

4. Results and Analysis

Table 1 shows the official results of our four runs. Our highest scoring run was the one that used retrieval feedback with text only which improved the textual runs without feedback by 7%. Although noticeable this is not a significant difference. When we combined these textual runs with the CBIR results, which has a much lower score (MAP=0.0153). Our combination with the simple text run yields very small improvement of results. However, the combination with the retrieval feedback run yields a significant drop in performance. We still are not sure of the reason for this drop in performance but will continue exploring to confirm whether it was an error in one of our procedures or if this is in fact a problem that needs to be examined in more detail.

Table 1 Official Runs

Run name	Type	MAP	Bpref	P@10
UNTtextb1	Text	0.2416	0.2784	0.404
UNTtextrf	Text with Retrieval feedback	0.2585	0.2826	0.436
UNTmixed1	Mixed	0.2447	0.2796	0.404
UNTmixedrf1	Mixed with retrieval feedback	0.1924	0.2358	0.42

5. Conclusions

Our experiments show that retrieval feedback does improve performance on the text based runs, although this improvement is relatively small (7%). Processing the entire collection using Metamap was time consuming but we think that the results justify the usage of the automatically assigned UMLS concepts and the corresponding terms. Parameter tuning of this generalized vector space model with 4 different spaces proved to be tricky and time consuming. We plan to work more on the query by query analysis so that we can get a better sense of what works in this approach and what needs to be improved or changed. One possibility that we will explore is the use a faceted classification that could be more appropriate for tagging images along the main aspects of interest for medical image retrieval such as image modality, orientation, etc.

Bibliography

Aronson, A. R. (2001). Effective Mapping of Biomedical Text to the UMLS Metathesaurus: The MetaMap Program. Aronson A. *Effective Mapping of Biomedical Text to the UMLS Metathesaurus* American Medical Informatics Association Annual Symposium.

GNU Image Finding Tool (GIFT). (2004). Retrieved August 22, 2009, from <http://www.gnu.org/software/gif>

Müller, H., Kalpathy-Crume, J., Eggel, I., Bedrick, S., Radhouani, S., Bakke, B., et al. (2009). Overview of the CLEF 2009 medical image retrieval track, CLEF working notes 2009. Corfu, Greece.

Ruiz, M. E., & Southwick, S. (2006). UB at CLEF 2005: Bilingual Portuguese and Medical Image retrieval tasks. *In Accessing Multilingual Information Repositories: 6th Workshop of the Cross-Language Evaluation Forum, CLEF 2005, , Revised Selected Pape*. Vienna, Austria: Springer.

Salton, G. (1971). *The SMART Retrieval System: Experiments in Automatic Document Processing*. Englewood Cliff, NJ: Prentice Hall.