A Task-Specific Query and Document Representation for Medical Records Search

Nut Limsopatham[1], Craig Macdonald[2], and Iadh Ounis[2]

School of Computing Science
University of Glasgow
G12 8QQ, Glasgow, UK
nutli@dcs.gla.ac.uk[1], firstname.lastname@glasgow.ac.uk[2]

Abstract. One of the challenges of searching in the medical domain is to deal with the complexity and ambiguity of medical terminology. Concept-based representation approaches using terminology from domain-specific resources have been developed to handle such a challenge. However, it has been shown that these techniques are effective only when combined with a traditional term-based representation approach. In this paper, we propose a novel technique to represent medical records and queries by focusing only on medical concepts essential for the information need of a medical search task. Such a representation could enhance retrieval effectiveness since only the medical concepts crucial to the information need are taken into account. We evaluate the retrieval effectiveness of our proposed approach in the context of the TREC 2011 Medical Records track. The results demonstrate the effectiveness of our approach, as it significantly outperforms a baseline where all concepts are represented, and markedly outperforms a traditional term-based representation baseline. Moreover, when combining the relevance scores obtained from our technique and a term-based representation approach, the achieved performance is comparable to the best TREC 2011 systems.

1 Introduction

Searching in the medical domain is challenging due to the complexity, inconsistency and ambiguity of the terminology [1, 2]. For example, some practitioners may refer to *cancer* as *carcinoma* rather than a *malignant tumour*. Prior works (e.g. [2–4]) have resorted to domain-specific resources to represent medical documents and queries in terms of controlled-vocabulary concepts to cope with such a challenge. For instance, *cancer*, *carcinoma*, and *malignant tumour* share similar meanings; hence, they are represented with the same medical concept [2, 4]. Intuitively, this could alleviate the mismatch of synonymous terms in a document and a query. However, it has been shown that a concept-based representation is effective only when combined with a term-based representation [2, 4]. Hersh et al. [3] reported that using only a concept-based representation was not effective. Later, Srinivasan [4] and Trieschnigg et al. [2] showed that a combination of term- and concept-based representation could be effective.

In this work, we deal with such a challenge in the context of medical records search, which focuses on finding patients having a medical history relevant to the query based on their medical records [5]. Medical records search systems aid healthcare practitioners in identifying effective procedures (e.g. diagnostic tests and treatments) for patients showing particular symptoms or diseases [6]. For example, it could be advantageous to

be able to search for patients who were previously admitted to a hospital with a heart disease, when a doctor compiles a list of possible effective procedures for dealing with a heart disease patient. In this paper, we hypothesise that representing medical records and queries by focusing on essential information for the medical records search could improve the retrieval effectiveness of the search system. Hence, we propose to use only the medical concepts related to four aspects of the medical decision criteria [7] to represent medical records and queries. We compare our proposed task-specific representation approach with the traditional term- and concept-based representation baselines. Our results show a marked improvement in the retrieval effectiveness.

2 A Task-Specific Query and Document Representation

We propose our task-specific representation approach to represent medical records and queries by focusing on medical concepts crucial for the medical records search task, which is to find patients having medical records relevant to the query [5]. Initially, we deploy Metamap [8] – a medical concept recognition tool based on the UMLS Metathesaurus that is widely used in previous works [8, 9] – to identify medical concepts, in medical records and queries, and represent them in the form of the UMLS Concept Unique Identifier (CUI). However, while traditional concept-based representation approaches use all identified medical concepts [3, 4], our proposed approach represents only the medical concepts related to criteria that are typically considered by healthcare practitioners when dealing with patients. In particular, we consider only the medical concepts directly relating to four aspects of the *medical decision criteria* [7], namely, symptom, diagnostic test, diagnosis, and treatment. We identify the concepts related to these four aspects based on the Metamap's semantic type field[1] - as listed in Table 1. For example, Table 2 shows the medical concepts obtained from the query *Patients with complicated GERD who receive endoscopy*, using our proposed task-specific representation approach. Some medical concepts, such as *receive* associated to the semantic type *Qualitative Concept*, are discarded by our system, since their semantic types are not related to the four aspects of the medical decision criteria.

3 Experiments and Results

We evaluate our proposed task-specific representation approach using the 34 query topics from the TREC 2011 Medical Records track [5], where the task is to identify relevant patient *visits* for each query topic. A visit, which contains all of the medical records associated with a patient's visit to a hospital, is used to represent a *patient* as a unit of retrieval, since relating multiple visits to a particular patient is made impossible because of privacy concerns [5]. For indexing and retrieval, we use the Terrier retrieval platform[2]. In the term-based representation, we apply Porter's English stemmer and remove stopwords. The parameter-free DPH term weighting model from the Divergence from Randomness framework (DFR) is used to rank medical records as it was shown to be effective in prior work [10]. The expCombSUM voting technique [11] is then used to rank visits based on the scores of their associated medical records [10]. The number of medical records that vote for the relevance of patient visits is limited to 5,000, as suggested in [10]. Moreover, since query expansion (QE) has been shown to be effective on the

[1] `http://metamap.nlm.nih.gov/SemanticTypeMappings_2011AA.txt`
[2] `http://terrier.org`

Table 1. List of 16 of the Metamap's 133 semantic types that we consider for our proposed approach, based on the four aspects of the medical decision criteria.

MetaMap's Semantic Type	Aspects of the Medical Decision Criteria			
	Symptom	Diagnostic test	Diagnosis	Treatment
Body Location or Region	✔	✔	✔	✔
Body Part, Organ, or Organ Component	✔	✔	✔	✔
Clinical Drug	–	–	–	✔
Diagnostic Procedure	–	✔	–	–
Disease or Syndrome	–	–	✔	–
Finding	✔	–	–	–
Health Care Activity	–	✔	–	✔
Injury or Poisoning	✔	–	–	–
Intellectual Product	–	✔	–	✔
Medical Device	–	✔	–	✔
Mental or Behavioral Dysfunction	✔	–	✔	–
Neoplastic Process	✔	✔	✔	✔
Pathologic Function	✔	–	–	–
Pharmacologic Substance	–	–	–	✔
Sign or Symptom	✔	–	–	–
Therapeutic or Preventive Procedure	–	–	–	✔

Table 2. Example of medical concepts obtained from the query *patients with complicated GERD who receive endoscopy* using our task-specific representation approach.

Concept (CUI)	Metamap's Definition	Related Aspects
C0017168	GERD (Gastroesophageal reflux disease)	Diagnosis
C0014245	Endoscopy (Endoscopy (procedure))	Diagnostic test

task [5, 10], we also evaluate the effectiveness of our proposed approach when QE is applied. In particular, we deploy the default DFR Bose-Einstein statistics-based (Bo1) QE model from Terrier to expand queries when calculating scores for medical records. Finally, to verify that our proposed approach could bring novel evidence for inferring relevance, which differs from that of the term-based representation, we follow the approach by Srinivasan [4], which we refer to as the *score combination* approach, to linearly combine the relevance scores of a medical record d towards query Q, calculated using both the term-based and our proposed task-specific representations, as followings:

$$score(d, Q) = \delta \cdot score_{term-based}(d, Q) + score_{task-specific}(d, Q) \qquad (1)$$

where δ is a parameter to emphasise the relevance score computed using the term-based representation, which is set to 2.00, as suggested in [4].

Table 3 compares the bpref retrieval performance of our proposed task-specific representation approach with the baselines where all medical concepts and terms, respectively, are used to represent medical records and queries. In particular, we show the retrieval effectiveness both with and without applying QE. Moreover, the retrieval performance of the score combination approach and the TREC 2011 best systems are also reported. From Table 3, when QE is not applied, our approach outperforms both baselines markedly. Indeed, our approach performs significantly better than the concept-based representation baseline (paired t-test, $p < 0.01$). In addition, we find that QE improves the retrieval effectiveness of all approaches. However, we observe a difficulty for Bo1 to improve the retrieval effectiveness of our approach and the concept-based representation approach. Furthermore, when applying QE, the score combination approach (bpref 0.5510) markedly outperforms either constituent approach (bpref 0.5264

Table 3. Bpref performance of different representation approaches and the TREC 2011 best system. Statistical significance (paired t-test) at $p < 0.05$ and at $p < 0.01$ over the corresponding concept-based representation baseline is denoted * and **, respectively.

Approach	bpref retrieval performance	
	without QE	with QE
Traditional concept-based representation	0.4485	0.4502
Traditional term-based representation	0.4871	0.5264*
Our proposed task-specific representation	0.5149**	0.5198*
Score combination ($\delta = 2.00$, as suggested in [4])	0.5214	0.5510
Best TREC 2011	0.5520	

and 0.5198). Finally, we find that when applying QE, the performance of the score combination approach (bpref 0.5510) is comparable to the TREC 2011 best systems (bpref 0.5520) [5], which deployed more sophisticated techniques, such as negation handling and ontology-based QE.

4 Conclusions

We have proposed a novel approach to represent medical records and queries by focusing only on the medical concepts from the four aspects of medical decision criteria. Our approach is shown to be effective on the Medical Records track 2011 test collection. Moreover, we find that our task-specific representation could provide new evidence to infer relevance in medical records search, as the retrieval performance is markedly improved, when combining the relevance scores computed using our proposed task-specific representation and the term-based representation. For future work, we plan to make inferences on the relationships of medical concepts within these four aspects to further improve the query representation in a medical records search system.

References

1. N. Limsopatham, R.L.T. Santos, C. Macdonald, and I. Ounis. Disambiguating biomedical acronyms using EMIM. In *SIGIR*'11.
2. D. Trieschnigg, D. Hiemstra, F. de Jong, and W. Kraaij. A cross-lingual framework for monolingual biomedical information retrieval. In *CIKM*'10.
3. W. Hersh, D. Hickam, R. Haynes, and K. McKibbon. A performance and failure analysis of SAPHIRE with a MEDLINE test collection. *JAMIA*, 1(1):51–60, 1994.
4. P. Srinivasan. Optimal document-indexing vocabulary for MEDLINE. *Inf. Process. Manage.*, 32(5):503–514, 1996.
5. E. Voorhees and R. Tong. Overview of the TREC 2011 Medical Records track. In *TREC*'11.
6. W. Hersh. *Information Retrieval: A Health and Biomedical Perspective (Health Informatics)*. Springer, 3rd edition, 2008.
7. E. Silfen. Documentation and coding of ED patient encounters: an evaluation of the accuracy of an electronic medical record. *AJEM*, 24(6):664 – 678, 2006.
8. A. R. Aronson and F. Lang. An overview of MetaMap: historical perspective and recent advances. *JAMIA*, 17(3):229–236, 2010.
9. A. R. Aronson. Exploiting a large thesaurus for information retrieval. In *RIAO*'94.
10. N. Limsopatham, C. Macdonald, R. McCreadie, and I. Ounis. Exploiting term dependence while handling negation in medical search. In *SIGIR*'12.
11. C. Macdonald and I. Ounis. Voting for candidates: adapting data fusion techniques for an expert search task. In *CIKM*'06.