Augmenting Medical Queries with UMLS Concepts via MetaMap

Christopher Agrafiotes and Avi Arampatzis

Database & Information Retrieval research unit,
Department of Electrical & Computer Engineering,
Democritus University of Thrace, Xanthi 67100, Greece

cagrafiotes@outlook.com, avi@ee.duth.gr

In: *Proceedings of The Twenty-Fifth Text REtrieval Conference, TREC 2016, Gaithersburg, Maryland, USA, November 15–18, 2016. National Institute of Standards and Technology (NIST) Special Publication.*

Abstract

This report is an overview of the methods used by the Democritus University of Thrace team (DUTH) at the Clinical Decision Support (CDS) Track of the 25th Text REtrieval Conference (TREC). We investigated the concept of atoms from the United Medical Language System (UMLS) as a method to augment medical queries. We ultimately found that massively augmenting queries with too many words leads to a decreased overall performance, whereas adding a few specific words works reasonably well boosting the baseline effectiveness.

I. Introduction

TREC 2016 is the third year the Clinical Decision Support (CDS) track is running. The track's goal is to investigate techniques for linking medical cases to biomedical articles relevant for patient care. In other words, the track's challenge is to retrieve, for a given case report, full-text biomedical articles that answer questions related to several types of clinical information needs.

This year's specific task was mostly the same with those of the previous' years. The main difference is that actual electronic health records (EHR) of patients were used as topics rather than synthetic cases specifically made for the track. Each of the 30 topics, besides the note regarding the medical record, was accompanied by a generic clinical question type, such as "What is the patient's diagnosis?", that should determine which biomedical articles are relevant or not.

II. **Approach**

i. Preprocessing and Indexing

This year, the number of the biomedical articles that comprise the dataset has grown to include 1.25M items, a 70% increase from last year's number. These articles are sourced from the Open Access Subset of PubMed Central, an online digital database of freely available full-text biomedical literature.

To build an index for these articles, we followed no special technique or preprocessing, in order to preserve the integrity of the dataset. Indri's (v5.8) indexer[1] was used with the Krovetz stemmer enabled.

ii. Method

During the previous runs of the CDS track, it has been observed that experimentations using MetaMap, a tool by UMLS to recognise and extract medical concepts in text, yielded acceptable results [1,2]. However, MetaMap is only a tool for recognising concepts and, usually, it returns entries from the UMLS Metathesaurus that are identical or almost identical to the concepts extracted [3]. Starting from this observation, it was decided to use the concepts detected using MetaMap, find the atoms of each such concept directly from the UMLS Metathesaurus, and augment with them the summary of each topic of the track.

These atoms are the basic building blocks of the Metathesaurus: if a concept appears in multiple source vocabularies, a unique atom identifier (AUI, atom for short) is assigned for each occurrence. Because of the extreme specialty of the medical language, it has been observed that the same concept may be expressed with completely different, linguistically irrelevant words. The usage of atoms will link those different words to the same concept; adding those to the query may be able to yield higher recall.

iii. Example

The following example will help illustrate more clearly the approach followed. This is a topic summary (topic #22) from the 2016 topicset:

```
94 M with CAD s/p 4v-CABG, CHF, CRI presented with vfib arrest.
```

Sumbitting the above sentence to MetaMap 2016 will produce the following:

```
1. Coronary Artery Bypass Surgery|C0010055|[topp]

2. Costa Rica|C0010182|[geoa]

3. Ventricular Fibrillation|C0042510|[dsyn]
```

[1] www.lemurproject.org

4. Complete Response
   with Incomplete Bone Marrow Recovery|C3538993|[fndg]
5. Congestive heart failure|C0018802|[dsyn]
6. Cranial Rhythmic Impulse|C1704519|[phsf]
7. RNH1 gene|C1826843|[gngm]
8. Arrested progression|C0237477|[tmco]
9. Law enforcement arrest|C0392351|[gora]
10. Presentation|C0449450|[idcn]
11. Ventricular Fibrillation by ECG Finding|C0344435|[lbtr]
12. Master of Science|C1513009|[inpr]
13. m/s|C0439493|[qnco]
14. CALD1 wt Allele|C3813548|[gngm]

The above are the concepts detected by MetaMap together with their respective CUI and semantic type. CUIs (Concept Unique Identifier) are identifiers used by UMLS to signal different meanings of the same word; for example, the two senses of culture are *'C0010453: Anthropological Culture'* and *'C0430400: Laboratory Culture'* in the UMLS release 2007AB. The semantic types are abbreviations to full words or phrases that explain the role of the concept; for example *"topp"* is linked to *"Therapeutic or Preventive Procedure"*, *"dsyn"* is linked to *"Disease or Syndrome"* and so on.

It is rather obvious that just submiting the summaries to MetaMap produces some completely irrelevant concepts (*Costa Rica*, *Law enforcement arrest*), a phenomenon that could be sourced at the linguistic structure of the summary. There are also some concepts (*Presentation*, *m/s*) that do not seem particularly helpful at answering the generic clinical question presented by the topic, which is *"What should be the treatment?"* In order to only keep the concepts of interest, their semantic types were used as flags. Only the concepts with the following semantic types were kept: [2]:

```
[fndg], [sosy], [patf], [menp], [mobd], [dsyn], [inpo], [tisu],
                    [topp], [diap], [orga]
```

Each of the kept concepts has a set of atoms (AUI) to its own, meaning that each concept has a set of words or phrases to describe it. For example, the *'Ventricular Fibrillation|C0042510|[dsyn]'* has the following atoms:

[2]For more information on semantic types, see:
`https://metamap.nlm.nih.gov/Docs/SemanticTypes_2013AA.txt`

```
"fibrillation;ventricular", "cardiac arrest and ventricular
    fibrillation", "fibrillation paroxysmal vent", "vf",
"fibrillations, ventricular", "vfib", "ventricular fibrillation",
  "fibrillation, ventricular", "fibrillation ventricular" "vf -
    ventricular fibrillation", "cardiac arrest - ventricular
      fibrillation", "paroxysmal ventricular fibrillation",
              "ventricular fibrillation paroxysm"
```

It can bee observed that a single MetaMap concept is linked to a variety of atoms (AUIs) even when the aforementioned flags are active. It will be very interesting to see how examples as the above affects the query performance.

III. **Experimental Results**

i. Evaluation Metrics

The metrics of interest at the CDS track are the following:

- *Inferred Average Precision (infAP)* [4]: This measure approximates the average precision even when relevance judgments are incomplete.
- *Inferred Normalized Discounted Cumulative Gain (infNDCG)* [4]: This measure approximates the usefulness, or gain, of a document based on its position in the result list using the graded relevance scale of the assessors.
- *R-Precision (R-prec)*: R-precision is the precision after R documents have been retrieved, where R is the number of relevant documents for the topic.
- *Precision at Rank 10 (P@10)*: The number of the articles within the top-10 results that are relevant to the topic.

ii. Official Runs

To explore the effect that this multitude of words will have at the final query, three main runs were submitted to TREC. The first one (*SA: Single Atom*) included the summary of each topic augmented with the first atom of each concept detected by MetaMap. The second run (*MA: Multiple Atoms*) included, besides the original summary, all the resulting atoms that were not already in it; in other words, synonym words or phrases were added to the query. The last run (*AA: All Atoms*) added to the summary the complete set of atoms detected with no filtering whatsoever.

Table 1: *Results for P@10, R-Precision, infAP, and infNDCG*

Run	P@10	R-Prec	infAP	infNDCG
SA	.2866	.1133	.0302	.2265
MA	.2866	.1031	.0280	.2006
AA	.2466	.0753	.0192	.1651
Average Median	.2633	.1219	.0195	.1858
Average Best	.6300	.2553	.0865	.4376

Table 1 shows the P@10, R-Precision, infAP, and infNDCG of the three runs submitted plus two rows labelled Average Median and Average Best. The latter two show the median and best performance per topic averaged over the total number of the topics according to TREC reported scores. It should be noted that Average Best indicates a hypothetical best system (or an upper bound) and Average Median indicates a hypothetical median system and are put here just for reference.

From Table 1, it can be observed that using atoms to augment the medical queries provided acceptable results. Adding a small number of words (SA/MA) is a safe enough approach that yields results mostly better than a hypothetical median system; however, adding the complete set of atoms for each concept detected by MetaMap (AA) in a medical query seems to disorientate the system, since the results are not very encouraging. The performance of these runs are also far away from the performance of a hypothetical best system.

IV. **Conclusions**

In this paper, the concept of atoms from the United Medical Language System (UMLS) was investigated as a method to augment medical queries. Many methods were explored, among them being one in which a single word was added to the original query or, another, in which a rather big list of words was used. It was observed that a larger number of words being added to a query leads usually to a decreased performance, while adding a few specific words works well.

Overall, two of the three runs submitted perform better than the hypothetical median system in three out of four evaluation measures. These results are modestly satisfying. Future work could be directed towards better semantic analysis of the medical summaries in order to extract even more precise concepts from MetaMap and, by extension, more accurate AUIs from UMLS Metathesaurus.

References

[1] George Drosatos, Stefanos Roumeliotis, Avi Arampatzis, and Eleni Kaldoudi. DUTH at TREC 2015 Clinical Decision Support Track. In: *Proceedings of The Twenty-Fourth Text REtrieval Conference (TREC 2015)*, Gaithersburg, Maryland, USA, November 17–20, 2015. National Institute of Standards and Technology (NIST) Special Publication 500-319.

[2] João Palotti and Allan Hanbury. TUW @ TREC Clinical Decision Support Track 2015. In: *Proceedings of The Twenty-Fourth Text REtrieval Conference (TREC 2015)*, Gaithersburg, Maryland, USA, November 17–20, 2015. National Institute of Standards and Technology (NIST) Special Publication 500-319.

[3] Aronson, A.R. Effective Mapping of Biomedical Text to the UMLS Metathesaurus: the MetaMap Program. In: *Proceedings of the American Medical Informatics Association Symposium.* pp. 17-21, 2001.

[4] E. Yilmaz, E. Kanoulas, and J. A. Aslam. A simple and efficient sampling method for estimating ap and ndcg. In: *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* , pp 603-610, ACM, 2008.