BIOINFORMATICS LETTER TO THE EDITOR Vol. 25 no. 20 2009, pages 2770–2771
doi:10.1093/bioinformatics/btp483

Data and text mining

Comment on 'MeSH-up: effective MeSH text classification for improved document retrieval'

Aurélie Névéol*, James G. Mork and Alan R. Aronson
US National Library of Medicine, Bethesda, MD 20894, USA

Received on June 26, 2009; revised on July 31, 2009; accepted on August 7, 2009
Advance Access publication August 11, 2009
Associate Editor: Jonathan Wren

Contact: neveola@ncbi.nlm.nih.gov

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 BACKGROUND

Information retrieval is an important task that requires specific attention in the biomedical domain where controlled vocabularies are available to characterize and organize textual content. A recent article published in *Bioinformatics* (Trieschnigg *et al.*, 2009) confirms that there is a continued interest in the community to address this problem and achieve 'improved document retrieval'.

As shown by the authors, the task of assigning controlled vocabulary descriptors or 'concepts' to documents is beneficial for information retrieval within a collection. While this task, called 'categorization' ('annotation' or 'classification') by the authors undoubtedly plays an important part in the retrieval process, it also plays a major role in summarizing the content of documents. We would like to stress this dual purpose of a task that aims at *characterizing* and *summarizing* the subject matter of texts. There is no inherent contradiction between *characterizing* and *summarizing*. However, for relatively large bodies of text such as documents, the focus is on providing a compact and selective description of the subject matter. On the other hand, for relatively shorter text such as information queries, the focus is on an exhaustive representation, that if possible, infers more than what is specifically spelled out.

These observations show that, in the context of information retrieval, the task of assigning controlled vocabulary descriptors to text actually branches out into two distinct tasks: *indexing*, where a limited number of descriptors denoting concepts that are substantively discussed are assigned to a document, and *query expansion*, where an exhaustive number of search terms are derived from an information query. As a result, it is not unreasonable to conjecture that different methods should be used for different tasks—or that a given method would require some degree of tailoring for each task.

2 COMMENT ON METHODS

2.1 Indexing

The assignment of Medical Subject Headings (MeSH®) descriptors can be viewed as a categorization problem in the sense that for a given document, the task consists of deciding whether a given heading should be assigned or not. Trieschnigg *et al.* (2009) rightfully state that 'the assignment of MeSH descriptors to text is a large multi-class and multi-label text classification problem' and that systems addressing the task should account for all MeSH descriptors versus a subset. We strongly agree with this assessment and recently proposed to increase the scope of the Medical Text Indexer (MTI) (Aronson *et al.*, 2004) from indexing based on ∼24 000 MeSH main headings to ∼580 000 MeSH indexing terms (main headings and main heading/subheading pairs) through subheading attachment (Névéol *et al.*, 2008). In our report of this work, we explicitly state that the intent behind the development of MTI is to provide assistance to MEDLINE® indexers and not to replace them. Consistent with this mission, MTI automatically provides a set of MeSH indexing recommendations that is limited in size and may include headings at different levels of specificity in a given MeSH hierarchy to make the task of an indexer easier by picking the adequate heading with a single click. In addition, in our 2008 study, we also report design decisions favoring recommendations that make sense to the indexers using the system over pure performance figures. The MeSH main heading recommendations provided by MTI are obtained through the combination of two methods followed by post-processing. The first MTI method uses MetaMap, a Natural Language Processing algorithm to extract Unified Medical Language System (UMLS®) concepts from biomedical text. Then, the non-trivial task of mapping the resulting UMLS concepts to MeSH main headings is performed using the 'Restrict-to-MeSH' algorithm (Fung and Bodenreider, 2005). Neither MetaMap nor 'Restrict-to-MeSH' is itself a MeSH 'classifier'. The former simply maps text to UMLS concepts, and the latter is a method that makes full use of UMLS semantics for mapping concepts between biomedical terminologies. The difference between limiting MetaMap processing to MeSH as used by Trieschnigg *et al.* (2009) and what is implemented in 'Restrict-to-MeSH' is illustrated by the example in the supplement. The second MTI method is adapted from the 'PubMed Related Citations' (PRC) algorithm (Lin and Wilbur, 2007), a modified k-NN algorithm.

2.2 Query expansion

In a previous retrieval experiment on three document sets including the OSHUMED collection, Kim *et al.* (2001) had shown that MeSH headings automatically obtained from MTI (referred to as the 'Indexing Initiative System' or IIS at the time) yielded an improvement over solely using title and abstract text. Kim *et al.* (2001) also noted in the discussion of their results that although the automatically assigned MeSH descriptors compared favorably to humanly assigned descriptors in the retrieval experiment, no conclusion could be drawn that a human searcher using MeSH terms for Boolean query would find the automatically assigned MeSH descriptors as useful as the humanly assigned descriptors.

In addition, a major benefit of using controlled vocabulary concepts in Information Retrieval is the possibility of exploiting the hierarchical relationships between concepts. For information retrieval in the MEDLINE

*To whom correspondence should be addressed.

Table 1. Attempt at reproducing indexing experiment of Trieschnigg *et al.*

Method	Experiment	MAP	P10	F1	Micro-F1
MTI	Trieschnigg *et al.*	0.2536	0.3200	0.4503	0.4415
	Névéol *et al.*	0.3243	0.3079	0.3677	0.3561
MetaMap	Trieschnigg *et al.*	0.1623	0.1910	0.3187	0.2968
	Névéol *et al.*	0.2333	0.2757	0.3329	0.3171
k-NN	Trieschnigg *et al.*	0.5052	0.4515	0.4074	0.4963
	Névéol *et al.*	0.2656	0.2893	0.2982	0.3000
PRC	Névéol *et al.*	0.7315	0.6011	0.6145	0.6010

database, the PubMed search algorithm uses relationships between MeSH headings by 'exploding' the search terms, so that documents indexed with a specific heading will be retrieved by a search on a more general, related heading. PubMed also employs the 'Automatic Term Mapping' (ATM) feature, which automatically maps a text query to MeSH for improved retrieval results. The benefits of this feature were formally assessed on the TREC collections recently (Lu *et al.*, 2009) and it is found that MeSH query expansion does not always improve retrieval.

3 COMMENT ON RESULTS

We find that the experiments performed by Trieschnigg *et al.* are difficult to reproduce.

Specifically, Table 1 shows the results we obtained (using TREC's 'treceval' package) when trying to replicate the indexing experiment using the following settings:

- MTI: default MTI setting with a weight of seven for the MetaMap method and two for the PubMed Related Citations.
- MetaMap: MTI setting with a weight of 1 for the MetaMap method and 0 for the PubMed Related Citations.
- k-NN: the k-NN method described by Trieschnigg *et al*. was implemented.
- PRC: Lin and Wilbur's algorithm was applied.

While some variation is to be expected, these results are quite different from those obtained by Trieschnigg *et al.* (Table 1). We observe that the results we obtain for MetaMap using the full Restrict-to-MeSH algorithm are superior to those obtained by Trieschnigg *et al*. This difference in performance can be explained by our different use of MetaMap, as illustrated by the example shown in the Supplementary Material.

However, we cannot explain the difference in MTI results. The drastic difference in k-NN results is also puzzling, even though two different implementations of the same algorithm are used. Furthermore, the results for sample text 'Reactive oxygen species and the regulation of cell death by the Bcl-2 gene family' given in the paper differ significantly from those obtained from the authors' online system.[1] In any case, the results obtained from the PubMed Related Citations algorithm are superior to that of any other method with the metrics used. We believe that this is largely due to the excellent recall performance of this method. In addition, we noticed that only 520 citations in the test corpus had an abstract. This means that 48% of the articles in the corpus had to be indexed using titles only. MTI and MetaMap are known to have much lower recall on 'title only' citations because they return very few terms based on an article title. As a result, the precision at 10 (P10) as computed by treceval would certainly be impacted negatively for citations where fewer than 10 terms were returned.

4 CONCLUSION

We have pointed out that the task of assigning MeSH descriptors to a body of text should be considered differently depending on the intended use of the resulting set of MeSH descriptors, and the type of text at hand. We provided some clarifications about NLM's tools and previous evaluations of those tools including through information retrieval experiments. We have found that experiments of Trieschnigg *et al.* were difficult to reproduce and that their results did not confirm previous work in the field. Finally, while a benchmark evaluation of indexing and query expansion methods was commendable, we feel that current challenges in MeSH indexing include an increase of the scope of the task (e.g. including subheadings) and an effort to meet the needs of the tool users (e.g. indexers) rather than solely focusing on abstract performance scores.

ACKNOWLEDGEMENTS

Funding: Intramural Research Program of the National Institutes of Health, National Library of Medicine.

Conflict of Interest: none declared.

REFERENCES

Aronson,A.R. *et al*. (2004) The NLM indexing initiative's medical text indexer. *Stud. Health Technol. Inform.*, **107**(Pt 1), 268–272.

Fung,K.W. and Bodenreider,O. (2005) Utilizing the UMLS for semantic mapping between terminologies. *AMIA Annu. Symp. Proc.*, 266–270.

Kim,W. *et al*. (2001) Automatic MeSH term assignment and quality assessment. *Proc. AMIA Symp.*, 319–323.

Lin,J. and Wilbur,W.J. (2007) Pubmed related articles: a probabilistic topic-based model for content similarity. *BMC Bioinformatics*, **8**, 423.

Lu,Z. *et al*. (2009) Evaluation of query expansion using MeSH in PubMed. *Information Retrieval*, **12**, 69–80.

Névéol,A. *et al*. (2008) A recent advance in the automatic indexing of the biomedical literature. *J. Biomed. Inform.* in press.

Trieschnigg,D. *et al*. (2009) MeSH Up: effective MeSH text classification for improved document retrieval. *Bioinformatics,* **25**:1412–1428.

[1] Accessed 21 June 2009 at http://www.ebi.ac.uk/Rebholz-srv/MeshUP.