A Rule-based Approach for Automatic Identification of Publication Types of Medical Papers

Abeed Sarker
Centre for Language Technology
Department of Computing
Macquarie University
NSW 2109 Australia
abeed.sarker@mq.edu.au

Diego Molla-Aliod
Centre for Language Technology
Department of Computing
Macquarie University
NSW 2109 Australia
diego.molla-aliod@mq.edu.au

Abstract *The medical domain has an abundance of textual resources of varying quality. The quality of medical articles depends largely on their publication types. However, identifying high-quality medical articles from search results is till date a manual and time-consuming process. We present a simple, rule-based, post-retrieval approach to automatically identify medical articles belonging to three high-quality publication types. Our approach simply uses title and abstract information of the articles to perform this. Our experiments show that such a rule-based approach has close to 100% precision and recall for the three publication types.*

Keywords Medical Document Classification, Post-retrieval Classification, Rule-based Classification, Evidence-based Medicine

1 Introduction

Medical practitioners seek high quality information when searching for evidence-based answers to clinical inquiries. The quality of a medical article depends on a number of factors including its publication type. Searching for and appraising high quality articles can be a cumbersome process and requires significant proportions of a practitioner's time when making clinical decisions [5, 7]. This problem is amplified by the large and growing number of available medical articles. The aim of our research is to reduce the time required for the appraisal process by automatic identification of the publication types of medical papers. We propose a simple rule-based approach that uses text from the article titles and abstracts to perform this classification. We show that our proposed approach is extremely efficient at correctly classifying three medical article types (Systematic Reviews, Meta-analyses and Randomized Controlled Trials), which are considered to be of high quality by the medical community, from a set of medical articles belonging to a range of publication types of varying quality levels.

Proceedings of the 15th Australasian Document Computing Symposium, Melbourne, Australia, 10 December 2010.

2 Background

2.1 Evidence-based Medicine

Evidence-based medicine (EBM) is the '*conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients*' [14]. Current clinical guidelines urge physicians to practice EBM when providing care for their patients. Good practice of EBM involves finding and appraising current medical evidence before making a decision. Therefore, it involves efficient use of information search and extraction strategies to identify good quality evidence [13].

EBM practitioners require comprehensive, specific bottom-line recommendations that directly answer clinical questions and hence, they often rely on sources of synthesized or pre-appraised evidence. However, databases with synthesized evidence (e.g. Cochrane Library[1]) only cover limited topics and in most cases practitioners have to rely on raw databases such as MEDLINE for information retrieval. MEDLINE, maintained by the US National Library of Medicine (NLM), comprises more than 18 million records and is available online (via the NLM PubMed[2] interface). A typical clinical query on this database returns thousands of results and in most cases assessing the quality of all the returned articles is not possible, particularly at point of care. This is the primary motivation behind implementing a system that performs post-retrieval classification to identify the quality of evidence of medical articles.

2.2 Strength of Evidence and Publication Types

The quality, strength or grade of a recommendation for a clinical query is based on a body of evidence typically found in more than one study. This usually takes into account (i) the level of evidence of the individual studies; (ii) the type of outcomes measured by these studies; (iii) the number, consistency and coherence of the evidence as a whole; and (iv) the relationship between benefits harms and costs [4].

[1] http://www.cochrane.org
[2] http://www.ncbi.nlm.nih.gov/PubMed

The level of evidence of an individual publication is tightly related to the type of publication. Medical publication types include (but are not limited to) Randomized Controlled Trials (RCTs), Systematic Reviews (SRs), Meta-Analyses (MAs), Practice Guidelines, Uncontrolled Clinical Trials, Single Case Studies, Cohort Studies, Tutorial Reviews and even personal opinions. Although all of them provide evidence of some form, the quality of their evidence varies significantly due to the different ways in which the studies are carried out. For example, a clinical trial consisting of a large number of randomly allocated subjects and carried out in a systematic and controlled manner (i.e. a RCT) has a higher level of evidence than a case study of a single patient. In other words, the outcomes presented in the former study are more reliable than the ones presented in the latter.

The connection between the type of publication and the strength of recommendation is generally acknowledged in the numerous grading scales. There are over 100 grading scales in use today [17]. The Strength of Recommendation Taxonomy (SORT) is one such grading scale and it is very popular in EBM practice [4]. SORT provides a uniform recommendation-rating system that can be applied throughout the medicine literature and its simplicity, straightforwardness and comprehensiveness increases its usefulness to practitioners. This taxonomy uses only three ratings A (strong), B (moderate) and C (weak) to specify the strength of recommendation (SOR) of a body of evidence. SORT provides an explicit link between the strength of recommendation and the publication type. Thus, an evidence of grade A may consist of high quality SRs, MAs, RCTs or even cohort studies with good follow-up. Figure 1 shows a pyramid of publication types arranged according to their usual levels of evidence. The pyramid does not explicitly show the SORT grades for the publication types shown. However, for a set of articles, the SORT grade can be derived from the pyramid. Usually, evidence obtained from articles belonging mostly to the top two levels in the pyramid are considered to be of grade A; those from mostly the middle of the pyramid are considered to be of grade B; and those from lower down the pyramid are considered to be of grade C.

2.3 Related Work

During our review of literature in this area, we did not find any previous work attempting to automatically classify medical documents with respect to publication types. However, there has been some research in the area of retrieval of clinically relevant articles. Hunt and McKibbon [9] present some key phrases that are useful for retrieving SRs while Montori et al. [11] use a set of terms including single words or phrases in abstracts or titles, subject headings, publication types etc. The slightly earlier approach proposed by Haynes et al. [8] is similar and relies quite heavily on the metadata associated with each article in MEDLINE instead of

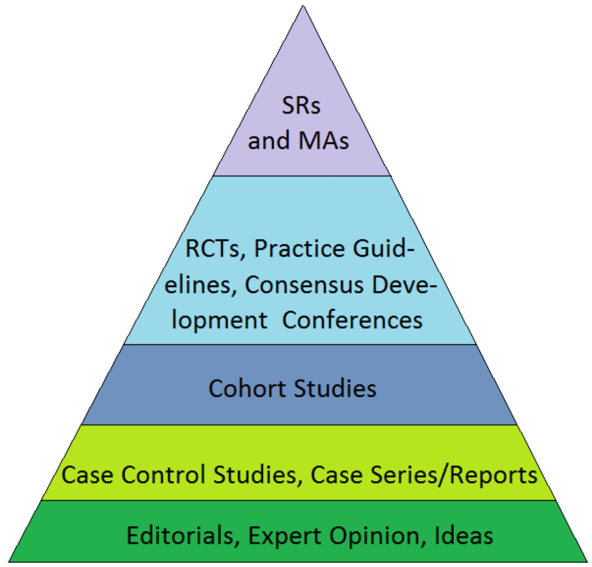

Figure 1: Level of evidence with respect to publication type.

the abstract and title texts only. Shonjania and Bero [15] also use metadata for retrieval and provide some PubMed search filters to identify SRs and show them to be quite effective. PubMed also points to some of the above mentioned sources to help practitioners formulate their search queries.

There has also been some research on automatic quality assessment of medical publications. Approaches based on word co-occurrences [6] and bibliometrics [12] have been proposed but these approaches do not integrate EBM recommendations for appraisal. Tang et al. [16], Aphinyanaphongs et al. [1] and Kilicoglu et al. [10] propose approaches for identifying high-quality medical articles that are more relevant for EBM. The post-retrieval re-ranking approach proposed by Tang et al. [16] is not directly comparable to ours because it does not directly take into account medical publication types and is applied to all articles returned by a search engine query (instead of formal, published papers only). Furthermore, their approach is only tested in a very specific sub-domain (i.e. Depression) within the much broader medical domain, which our approach attempts to work in. The other two approaches mentioned above are based on machine learning techniques and are also shown to be quite effective. However, these approaches also rely largely on the metadata accompanying each MEDLINE article. Metadata is only a moderate predictor of the clinical value of an article [3] and relying heavily on metadata associated with a MEDLINE article makes classification approaches suitable for this database only. Additionally, the semi-automatic approach used for indexing MEDLINE articles has evolved with time due to the increasing frequency of medical article publication. As a result, the metadata content may vary significantly between articles published at different times. Furthermore, MEDLINE does not have a 'PublicationType' tag for SRs (they are usually assigned the 'Review' tag together with non-systematic reviews), many articles do not have any 'PublicationType' tag assigned at all and many have multiple distinct tags for this category. An approach that relies solely on the article contents (i.e. titles and abstracts), such as the one we are proposing in this paper, would clearly overcome these problems. This technique can be applied after retrieval to cluster the articles based on their publication types, allowing the practitioner to easily identify the most suitable ones and extract evidence from them.

3 Methods

Abstracts and often titles of medical articles contain information about the types of studies and therefore provide evidence of their publication types. Our approach relies on regular expressions to identify relevant patterns (evidence) from titles and abstracts. At this point of research, we focus only on identifying SRs, MAs and RCTs since articles belonging to these publication types are most often associated with SOR level A, as explained in Section 2.2.

3.1 Rule Development

We developed the expressions used to classify articles by manually studying the titles and abstracts of articles belonging to each of the above mentioned publication types. We collected our development set from a mixture of sources. For articles which have associated 'PublicationType' tags in MEDLINE (e.g. RCTs and MAs) we retrieved about two hundred of each type. We studied each article individually, identified the evidence of publication type and developed patterns to pick up the evidence. During development of the rules, we used an incremental approach similar to the Ripple Down Rules [2] philosophy – after adding a new regular expression we tested its effect on our development set and added more expressions based on the articles that were not correctly identified. For example, in the case of RCTs we primarily developed expressions to detect evidence of randomization in the abstracts. Once evidence of randomization is found, we also developed expressions (from false positives) to detect evidence(s) of unacceptable randomization[3]. Some of the expressions used to identify RCTs are given below:

```
Evidence of Randomization:
'random.*allocate'
'randomi[sz]ed.*study'
'random.*clinical'
'design:.*random'
```

[3] Details about unacceptable randomization techniques for RCTs and other publication types can be found at http://www.nlm.nih.gov/mesh/pubtypes2004.html

```
Evidence of no or unacceptable
randomization:
'coin\W*flip'
'non\W*random'
'odd\W*even'
'uncontrolled\W*study'
```

For articles without an associated 'PublicationType' tag in MEDLINE (e.g. SRs), obtaining a large development set was considerably difficult. We therefore used a mixture of secondary sources of evidence such as the Journal of Family Practice[4] (JFP) and the Cochrane Library for obtaining about fifty of each and developed our expressions from that set. Furthermore, we studied search techniques suggested by PubMed[5] for efficient retrieval of SRs and developed expressions based on their suggestions. We also developed expressions based on search keywords and techniques suggested in the literature for obtaining articles of specific publication types [9, 11]. After studying the resulting development set, we observed that a relatively small number of carefully developed expressions is sufficient to achieve our goal and in our current approach we use a total of 25 patterns for SRs and MAs and 48 patterns for RCTs.

3.2 Application of Rules

We apply a decision list to identify the publication types of articles. Each article is initially assigned an empty tag and passed through a sequence of tests, each responsible for checking for patterns indicating a specific publication type. At any stage of the sequence, if sufficient evidence of a particular publication type is found (with no further evidence of negation), the article is tagged and removed. The sequence in which the operations are applied is very important as the number of false positives may increase significantly if the sequence is changed. For example, if SRs and MAs are not removed before searching for RCTs, many of the former are falsely tagged as the latter. This is because abstracts of SRs and MAs usually mention the number and types of studies that are being reviewed/analysed, which usually includes RCTs (along with other types of studies). The following list elaborates the actions performed at each stage of the sequence[6]:

1. Check title for evidence of SR or MA[7]

2. Check title for evidence of Practice Guideline or Consensus Development Conference

3. Check title for evidence of RCT

4. Check abstract text for evidence of SR or MA

5. Check abstract text for evidence of RCT

[4] http://www.jfponline.com/

[5] The techniques can be found at http://www.nlm.nih.gov /bsd/PubMed_subsets/sysreviews_strategy.html

[6] Steps 2 and 6 are not discussed in this paper

[7] The two types are grouped together since MAs are actually types of SRs.

6. Check for evidence of other low priority publication types (e.g. Evaluation, Cohort Studies, Multicentre Studies etc.)

While checking the abstract of an article for evidence, each sentence is searched separately. We have attempted other approaches such as searching the whole abstract and using a sliding window. However, we have found sentence-level searching to produce the best results primarily because evidence of publication or study type is usually stated or described in a single sentence of an article abstract. Once a pattern match occurs, the entire abstract is searched again to identify patterns that negate the evidence (such as unacceptable randomization techniques in the case of RCTs) and the article is only tagged if no evidence of negation is found. For the mentioned publication types, such a simplistic negation detection technique proves to be sufficient.

4 Results and Discussion

For reasons mentioned in Section 2, we did not depend on the MEDLINE metadata to annotate our test set. We required a set of test articles that were different from the development set and at the same time completely reliable. To achieve this, we used JFP to build our test data. From the Clinical Inquiries sections of the JFP issues, we identified medical articles that are explicitly mentioned (by the JFP authors) to be RCTs, MAs or SRs and were not present in the development set. Importantly, the chosen articles are not actually written by JFP authors, but are cited by them within JFP articles which provide evidence-based answers to clinical queries. Hence, the chosen articles come from a variety of sources and this enables us to test our approach on a diverse article collection. To obtain the article abstracts and titles, we searched for those medical articles in MEDLINE using PubMed and added them to our test set after manually annotating them based on the JFP classifications. Relying on JFP for the test data also allowed us to include articles from a wide range of medical topics, thus ensuring that our approach is not topic dependent. Also, to further prevent bias, all articles identified were added to the test set regardless of their structure/content and the abstracts of the articles were not reviewed during the annotation process. Such a labourious annotation process was necessary due to the lack of substantial reliable annotated data.

For our test set, we used a total of 294 articles including 111 SRs and MAs, 100 RCTs and 83 articles belonging to a mix of other publication types. Including a set of articles belonging to various other publication types was necessary to ensure that our approach does not only correctly tag SRs, MAs and RCTs but also leaves other types of articles untagged. The recall, precision and F-score values are shown in Table 1. For SRs and MAs, our approach produced perfect precision but failed to identify one SR. On the other hand, our approach tagged a total of 97 articles as RCTs, of which 96 were correctly identified.

Publication Type	**Recall**	**Precision**	**F-Score**
MA and SR	0.990	1.00	0.995
RCT	0.960	0.990	0.975

Table 1: Automatic classification results (sample size = 294).

In our post-test review, we discovered that in the case of RCTs, the falsely tagged article was a Review (non-systematic) which mentioned ‘one randomized, placebo-controlled study’ and was therefore picked up by our rules. As for the four RCTs that were not identified, none of their abstracts contained any evidence of randomization although for one of the RCTs, there was clear evidence of randomization in the full article text. In the case of SRs and MAs, the unpicked article was a SR in which the abstract did not contain any detail of the study type.

The results clearly indicate that a rule-based approach such as ours is very effective in classifying SRs, MAs and RCTs. The high f-scores can be attributed to the fact that articles belonging to these three publication types are very structured (since there are very specific guidelines that must be followed when writing these articles) and therefore their titles and abstracts almost invariably contain sufficient evidence of the type of publication, which can be automatically identified. Furthermore, since the approach does not take into account the metadata associated with each article, it can be applied to articles across various databases.

5 Conclusion and Future Work

In this paper, we have presented an automatic, rule-based approach for classifying medical articles with strong levels of evidence. The results presented here are a step towards a more ambitious goal of automatically identifying the SORs of sets of medical articles. Our results show that the approach is very promising and may be used for automatic classification of other types of medical articles as well. We did not experiment with any machine learning algorithm due to the little amount of annotated data but considering the good results obtained, machine learning is perhaps not necessary. Our future research will focus on testing this rule-based approach with more manually annotated documents. Also, the system will be extended to cover more publication types, which may be a harder problem to solve considering the lower quality of the structure in articles of lower priority.

Acknowledgements This research is jointly funded by Macquarie University and CSIRO.

References

[1] Y. Aphinyanaphongs, I. Tsamardinos, A. Statnikov, D. Hardin and C. F. Aliferis. Text categorization models for high-quality article retrieval in internal medicine. *Journal of the American Medical Informatics Association : JAMIA*, Volume 12, Number 2, pages 207–216, 2005.

[2] Paul Compton and Bob Jansen. Knowledge in Context: A Strategy for Expert System Maintenance. In *Proceedings of the 2nd Australian Joint Artificial Intelligence Conference*, pages 292–306, 1988.

[3] Dina Demner-fushman, Susan Hauser and George Thoma. The role of title, metadata and abstract in identifying clinically relevant journal articles. In *AMIA Annu Symp Proc*, 2005.

[4] Mark H Ebell, Jay Siwek, Barry D Weiss, Steven H Woolf, Jeffrey Susman, Bernard Ewigman and Marjorie Bowman. Strength of recommendation taxonomy (SORT): a patient-centered approach to grading evidence in the medical literature. *Am Fam Physician*, Volume 69, Number 3, pages 548–556, February 2004.

[5] John Ely, Jerome A Osheroff, Mark H Ebell, M. Lee Chambliss, DC Vinson, James J. Stevermer and Eric A. Pifer. Obstacles to answering doctors' questions about patient care with evidence: Qualitative study. *BMJ*, Volume 324, Number 7339, pages 710, 2002.

[6] Thomas Goetz and Claus-Wilhelm von der Lieth. PubFinder: a tool for improving retrieval rate of relevant PubMed abstracts. *Nucleic Acids Research*, Volume 33, pages W774–W778, 2005.

[7] Trisha Greenhalgh. *How to read a paper: The Basics of Evidence-based Medicine*. Blackwell Publishing, 3 edition, 2006.

[8] R Brian Haynes, Nancy Wilczynski, K Ann McKibbon, Cynthia J Walker and John C Sinclair. Developing Optimal Search Strategies for Detecting Clinically Sound Studies in MEDLINE. *J Am Med Inform Assoc*, Volume 1, Number 6, pages 447–458, 1994.

[9] Dereck L Hunt and K Ann McKibbon. Locating and appraising systematic reviews. *Ann Intern Med.*, Volume 126, Number 7, pages 532–538, 1997.

[10] Halil Kilicoglu, Dina Demner-Fushman, Thomas C. Rindflesch, Nancy L. Wilczynski and Brian R. Haynes. Towards automatic recognition of scientifically rigorous clinical research evidence. *J Am Med Inform Assoc*, Volume 16, Number 1, pages 25–31, January 2009.

[11] V. M. Montori, N. L. Wilczynski, D. Morgan, R. B. Haynes and . Optimal search strategies for retrieving systematic reviews from medline: analytical survey. *BMJ*, Volume 330, Number 7482, January 2005.

[12] Maksim Plikus, Zina Zhang and Cheng-Ming Chuong. PubFocus: semantic MEDLINE/PubMed citations analytics through integration of controlled biomedical dictionaries and ranking algorithm. *BMC Bioinformatics*, Volume 7, Number 1, pages 424–439, 2006.

[13] D. L. Sackett, R. B. Haynes, G. H. Guyatt and Tugwell P. *Clinical epidemiology: A basic science for clinical medicine*. Little Brown & Co. Inc., 2 edition, 1991.

[14] David L Sackett, William M C Rosenberg, J A Muir Gray, R Brian Haynes and W Scott Richardson. Evidence based medicine: what it is and what it isn't. *BMJ*, Volume 312, Number 7023, pages 71–72, 1996.

[15] Kaveh G Shonjania and Lisa A Bero. Taking Advantage of the Explosion of Systematic Reviews: An Efficient MEDLINE Search Strategy. *Effective Clinical Practice*, Volume 4, Number 4, pages 157–162, 2001.

[16] Thanh Tang, David Hawking, Ramesh Sankaranarayana, Kathleen Griffiths and Nick Craswell. Quality-Oriented Search for Depression Portals. In Mohand Boughanem, Catherine Berrut, Josiane Mothe and Chantal Soule-Dupuy (editors), *Advances in Information Retrieval*, Volume 5478 of *Lecture Notes in Computer Science*, Chapter 60, pages 637–644. Springer Berlin / Heidelberg, Berlin, Heidelberg, 2009.

[17] Suzanne West, Valerie King, Timothy S. Carey, Kathleen N. Lohr, Nikki McKoy, Sonya F. Sutton and Linda Lux. Systems to rate the strength of scientific evidence. http://www.arhq.gov/clinic/epcsums/strengthsum.htm, April 2002.