Reducing systematic review workload using text mining: opportunities and pitfalls

Claire Stansfield, Alison O'Mara-Eves and James Thomas
Evidence for Policy and Practice Information and Coordinating (EPPI-) Centre, Social Science Research Unit, UCL Institute of Education, University College London, London, United Kingdom

Abstract
This EAHIL workshop focussed on three applications of text mining to assist with screening citations for systematic reviews, and encouraged participants to discuss issues affecting their adoption. This paper outlines these applications and summarises the factors raised by participants in relation to their uptake. Key aspects to uptake include having an accepted advantage over existing approaches, coupled with training and user support.

Key words: systematic reviews; text mining; study selection; automation.

Introduction

This workshop focussed on using text mining to assist with citation screening, which is a necessary but time-consuming step in conducting a systematic review. The first half of the session described the processes and applications, which was followed by group discussions on challenges of adopting this technology for the different applications. We were pleased to have such an interested group and the 45 participants fully engaged in the discussions. The issues they raised will feed into guidance that the EPPI-Centre is developing on when and how to use this technology.

Overview

The citation screening process in a systematic review involves checking a citation against specific criteria to assess whether it is suitable for answering the review's research questions. Typically this is carried out on the titles and abstracts of citations that have been identified through systematic searches, before the full-text documents of relevant citations are retrieved. This can be a laborious and time-consuming process, with potentially tens of thousands of citations to be screened.

Text mining has the potential to automate at least some of this process with potential benefits including a) reducing the time spent screening; b) ranking the citations so the most relevant items are identified early on in the screening process; and c) providing a second-check to ensure relevant studies are not missed by human reviewers. Moreover, if screening time is reduced it also offers the possibility of conducting more sensitive searching as larger numbers of citations can be "screened". Putting all this together, this can change the approach to systematic reviewing. Such technology does not reduce the need for skilled information professionals in developing search strategies, as performance relies on good training data of a suitable sample of relevant and irrelevant studies. Some of the processes, however, currently have some limitations and need further evaluation.

Current research and opportunities

At the 2015 EAHIL Workshop, James Thomas presented an overview of the technologies and evaluations of their performance, and a live demonstration was performed on a participants' Cochrane register. We also had time for a brief snapshot on other applications of text mining: developing search strategies and mapping (obtaining an overview of the topics in a group of citations). An overview of the technology for use in systematic reviews that were discussed is shown in *Figure 1.*

In *Figure 1*, the technology for term recognition and

Address for correspondence: EPPI-Centre, Social Science Research Unit, 18 Woburn Square, UCL Institute of Education, London, UK WC1 0NR. Email: c.stansfield@ucl.ac.uk

automatic clustering are based on the corpus of text under analysis. In contrast, automatic classification analyses both text and decisions from a human screener so it can assign levels of relevance to the citations. Essentially it is trained by a human on a randomised sample of citations, though it needs a sufficient number of relevant and irrelevant citations for optimal performance. When a human continues screening, the automatic classifier is retrained. A register of studies that have previously been classified by humans may also provide this training data.

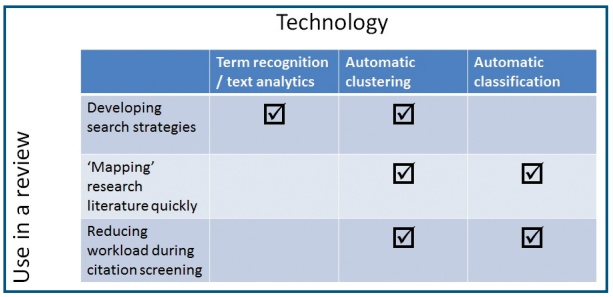

Figure 1. *Overview: how text mining can be used in systematic reviews*

Focusing specifically on the screening stage in systematic reviews, there are currently three main applications:

- *Screening prioritisation.* Using text mining to prioritise the order in which items are screened. After being "trained" by a human on a subset of known includes and excludes, the machine identifies words (and combinations of words) that are associated with includes and those with excludes, and lists all studies in order of likelihood of inclusion for the human/s to screen.
- *Double screening*. The use of text mining as a "second screener". At least one human screens the studies and their decisions are compared with the include/exclude recommendations of the machine. The researcher can specify how any conflicts are dealt with.
- (Semi-)automatic classification. The use of text mining to eliminate studies automatically or semi-automatically. After being "trained" by a human on a subset of known includes and excludes, the machine classifies all records as either includes or excludes.

Each of these three approaches has pros and cons, which were the focus of the group discussion (summarised later in this paper). Additionally, the research in this area is very new and fairly small, as highlighted in a systematic review of the methods used in text mining for screening (1). This review concluded that text mining to prioritise the order in which items are screened is suitable for use in reviews, and using text mining as a "second screener" may also be used cautiously. Using text mining to automatically classify studies should be considered promising, but its utility is not fully proven. In highly technical/ clinical areas, it may be used with a high degree of confidence; but more developmental and evaluative work is needed in other disciplines. One opportunity being investigated is part of the Cochrane Collaborations' Transform Project. Potentially, randomised controlled trials collected from crowdsourcing initiatives could be automatically directed to the most relevant Cochrane review group or systematic review (2).

Issues for adoption

Some of these applications raise questions on processes that are inherent for traditional systematic reviews. For example, is it acceptable to not screen all of the studies identified through searches? Does the technology perform sufficiently well to use? Does it actually save time?

This is also coupled with issues on adopting new innovations in general. Despite being available to systematic reviewers since 2006, text mining has not been widely adopted (3). Rogers (4) proposed five characteristics that affect the rate of adoption of innovations, which might be considered to explain the low uptake (Box 1). At the Edinburgh workshop, the groups discussed their relative importance in uptake of text mining for screening citations.

1. ***Perceived relative advantage*** (does it appear to have benefits to the user?)
2. ***Compatibility*** (is it consistent with past experiences and the needs/values of the user?)
3. ***Trialability*** (can the user try it out in their own work?)
4. ***Observability*** (are the results of the innovation visible to others?)
5. ***Complexity*** (is it perceived as easy to understand and use?)

Box 1. *Characteristics that affect the rate of adoption of innovations (adapted from Rogers 2003)*

Outcomes of group discussion on text mining for screening citations

The overall impression from the group discussions were that people were positive about the benefits of the technology and were open to exploring issues on its application. Participants expressed that perceived relative advantage was seen as most critical to uptake. Librarians and information scientists need to be able to demonstrate advantage to the reviewers. There were concerns that using new methods for a systematic review would also need general acceptance by publishers of reviews.

Trialability was considered important, which includes having an understanding of the technology to try it out without needing computer programmers. This was combined with specific concerns about access to software: is it open source or licensed; off-the-shelf or do they have to program themselves; and what support is needed to use it.

Training was seen as essential, and the development of guidance on its use was welcomed. There were also concerns about transparency: how would one know if a mistake had been made given that it is complex to understand how one has obtained the results. There is also perhaps a need to communicate differences between trained automatic classification and the relevance-ranking function that exists in commercial bibliographic databases.

Other issues included concern about literature in different languages, misspellings and symbols. Ease of importing datasets and the appropriateness for the topic area were also raised. One participant observed it might mean no need for removal of duplicates, and another participant was relieved that it would not be the end of manual screening all together.

Conclusion

We enjoyed discussing text mining with so many health librarians and information scientists. It was particularly useful to discover that people were generally open to the use of these technologies, with caveats related to the issues on adoption and use. Acceptance that the technologies had a relative advantage over existing approaches, coupled with thorough training and user support, were seen as critical to uptake. We aim to publish guidance on using these technologies in 2016.

Declaration

James Thomas is co-lead of the Cochrane "Transform" project, which is implementing some of the technologies discussed here. He also directs development and management of EPPI-Reviewer, the EPPI-Centre's software for systematic reviews.

The research project used for illustration was funded by the Medical Research Council, UK. The views and opinions expressed by authors are those of the authors and do not necessarily reflect those of the MRC.

REFERENCES

1. O'Mara-Eves A, Thomas J, McNaught J, Miwa M, Ananiadou S. Using text mining for study identification in systematic reviews: a systematic review of current approaches. Systematic Reviews. 2015;4:5. doi:10.1186/2046-4053-4-5
2. Project Transform [internet] 2015 [cited 2015 Jun 30] http://community.cochrane.org/transform
3. Thomas JM. Diffusion of innovation in systematic review methodology: Why is study selection not yet assisted by automation? OA Evidence-Based Medicine. 2013 Oct;1(2):12.
4. Rogers E. Diffusion of innovations. 5th ed. New York, NY: Free Press; 2003.