Text mining for screening efficiency? Testing within a Cochrane public health review

Alison Weightman1, James Thomas2, Philip Baker3, Yolanda Lovel-Too4, Daniel Francis3,4, Alison O’Mara-Eves2
1 Support Unit for Research Evidence (SURE), Cardiff University UK, Cochrane Information Retrieval Methods Group; 2 EPI Centre, Institute of Education UK, #CochraneTech; 3 Institute of Health and Biomedical Innovation, Queensland University of Technology, Australia, Cochrane Public Health Group; 6 Metro North Public Health Unit, Queensland Australia

Background

The requirement for dual screening of titles and abstracts to select papers to examine in full text can create a huge workload, not least when the topic is complex and a broad search strategy is required, resulting in a large number of results. An automated system to reduce this burden, while still assuring high accuracy, has the potential to provide huge efficiency savings within the review process.

Objectives

To undertake a direct comparison of manual screening with a semi-automated process using a machine classifier. The research was carried out as part of the current update of a Cochrane population level public health review on community wide interventions for increasing physical activity.

Methods

Authors hand selected studies for the review update, in duplicate, using the standard Cochrane methodology. Studies selected at title/abstract stage (T&A) for the original review (2011) and this 2014 update, as well as those included in the original review following full text assessment (FT), provided the ‘gold standard’ data.

Two text mining techniques were applied retrospectively:

A single ‘ranked list’: The gold standard data was used to train a machine classifier to ‘recognise’ the difference between a likely ‘include’ and ‘exclude’. The classifier then generated a list, ordered by the likelihood that they are relevant to the new review.

Active learning: Using the gold standard data, the process of active learning was simulated. Beginning with random samples of 5,10 or 20 ‘includes’ and ‘excludes’, the classifier was trained to generate a ranked list, boosting the relevant items to the top of the list. Manual screening was simulated with the most relevant items being prioritised for ‘manual’ screening and the ranked list being revised every 25 items ‘screened’.

Results

From a search retrieval set of 9555 papers, review authors manually excluded 9494 papers at title/abstract and 52 at full text, leaving 9 papers for inclusion in the review update.

Neither the Ranked List (Figure 1) nor Active Learning (Figure 2) were able to identify all of the titles and abstracts that had been identified manually without ‘looking’ at most of the irrelevant citations too; many relevant citations were found sooner than would otherwise be the case, but some ‘stragglers’ remained.

However, if the classifier doesn’t mind missing some T&As so long as no FT are missed, the burden would be reduced considerably - by up to 67% with the ranked list approach and 70% with active learning.

Discussion & Conclusions

This was a challenging review for text mining and performance does not reach the levels seen elsewhere in terms of the ability of the classifier to identify all potentially relevant titles and abstracts selected by manual screening. However, if the reviewer is willing to accept that some titles/abstracts may not be picked up but that all full text includes will be found, benefits may be huge and could reduce the screening burden by up to 70%. An enormous efficiency saving where a large number of results need to be screened.

Would reviewers be comfortable with the idea of not looking at everything?

If not, the use of text mining for the results from database searching can still: (i) identify the most relevant studies earlier in the process, enabling appraisal and data extraction to begin; (ii) be combined with supplementary search methods (eg reference list follow up, citation tracking and contact with experts) to minimise the risk of missing relevant papers.

Wanting to test this out?

Contact James Thomas (j.thomas@ioe.ac.uk). It may be possible to include your review in the current study, completion date July 2015

This work was partially funded by the Medical Research Council grant for Test Mining and Screening (Grant No. MR/J005837/1)