

Automatic Title and Abstract Screening in Healthcare Systematic Reviews

Lee A¹ and Flaxman A¹ and Salama J¹ and Afshin A¹

¹Institute for Health Metrics and Evaluation

Background. Systematic reviews are widely considered the preferred approach for synthesizing and summarizing the primary scientific evidence on a specific research question. The first step of a systemic review is evaluating the titles and abstracts of the articles identified through literature search. This step is generally the most time-consuming step of the process and constrains the scope of systematic reviews. As a part of the Global Burden of Disease study (GBD), we annually conduct systematic review of literature to identify representative data on prevalence of more than 400 diseases, injuries, and risk factors. To increase the speed and accuracy of the process, we sought to automate the process of title and abstract screening in healthcare literature using natural language processing (NLP).

Methods. The Global Burden of Disease study (GBD) is a systematic, scientific effort to quantify the comparative magnitude of health loss due to diseases, injuries, and risk factors by age, sex, and geography. As part of this project we have created a dataset of 58,326 labeled title and abstract dataset. These titles and abstracts are all labeled with a binary 1 or 0 representing “include” and “exclude” respectively. The title and abstracts were used as the training dataset for the NLP models. We ensembled two common NLP neural network architectures: a long short-term memory model (LSTM) and an attention model that uses Google’s word2vec model to map common words to vectors. While each one of these model architectures offer their own strengths and weaknesses, ensembling the two models together allows us to benefit from the strengths and mitigate the weaknesses. During model training, we focused on optimizing the weighted categorical crossentropy of the model.

Results. Our results showed a 91.0% accuracy in correctly classifying whether or not a title and abstract should be included for analysis or not. Given that the model’s output prediction is a probability that it is associated with the “include” and “exclude” classes, we assessed the performance of the model when looking at guesses the model made when it was >90%, >95% and greater than >99% certain. The accuracies were 98.4%, 99.1%, and 99.7%. Our models were able to process 100 title and abstracts in less than one second, greatly increasing the speed of the screening process.

Conclusion. Here we have proposed a new method for screening scientific articles that uses a ensembled LSTM and attention model that vastly improves both the speed and quality of this process. Though this model has been trained on epidemiological articles, the model can be easily trained and tuned to work with clinical and medical articles as well. This can help medical professionals and researchers alike greatly improve the speed and accuracy of their systematic literature reviews.