Short Review

DOI: http://dx.doi.org/10.53043/2347-3894.acam90025

Short Review on srBERT: Automatic Article Classification Model for Systematic Review Using BERT

Seon Choe[1], Sungmin Aum[2,3], Ju Han Kim[1]*

[1]Seoul National University Biomedical Informatics (SNUBI), Division of Biomedical Informatics, Seoul National University College of Medicine, Seoul, 03080, South Korea

[2]Korea Institute of Science and Technology (KIST), Robotics and Media Institute 5, Hwarang-ro 14-gil, Seongbuk-gu, Seoul

[3]University of Science and Technology (UST), Division of Nano and Information Technology, 34113, Gajeong-ro, Yuseong-gu, Daejeon, Republic of Korea

Asian Journal of Complementary and Alternative Medicine. Volume 10 Issue 01

Published on: 25/01/2022

***Author for Correspondence:** Ju Han Kim, Seoul National University Biomedical Informatics (SNUBI), Division of Biomedical Informatics, Seoul National University College of Medicine, Seoul, 03080, South Korea, E-mail: juhan@snu.ac.kr

Cite this article as: Choe S, Aum S, Kim JH. *Short Review on srBERT: Automatic Article Classification Model for Systematic Review Using BERT*. Asian Journal of Complementary and Alternative Medicine, Vol 10(1), 16-19:2022.

ABSTRACT

Systematic reviews (SRs) have been recognized as the most rigorous and reliable approach to enable evidence-based medicine. However, the considerable workload required to create SRs prevents reflecting the latest knowledge. This study automated the classification of included articles by adopting the Bidirectional Encoder Representations from Transformers (BERT) algorithm. By pretraining with abstracts of articles and a generated vocabulary fine-tuned with titles of articles, the proposed srBERTmy overcomes the training data insufficiency while improving performance in both classification and relation-extraction tasks. Despite the limitation of model vulnerabilities based on training dataset quality, the results demonstrated the feasibility of automatic article classification using machine-learning (ML) approaches to support SR tasks.

Keywords: systematic review, process automation, deep learning, text mining

BACKGROUND

A systematic review (SR) is a comprehensive review of all relevant evidence to answer a research question [1–4]. It is recognized throughout the various types of clinical research for its high reliability [5], enabling evidence-based medicine in clinical practice [6].

However, due to the time-consuming and labor-intensive process for SRs, requiring an average of 67.3 weeks from protocol to publication [7], most SRs are insufficient in reflecting the latest research findings [8]. Despite suggestions to support parts of the SR process using advanced machine-learning (ML) techniques [9–14], such models are still unable to achieve high precision scores, focusing more on comparatively simple and intuitive tasks [15–18].

This study was devised based on the difficulties we experienced during the actual SR process. We attempted to automate the screening stage, retrieving all relevant literature based on a predefined research question. These tasks were error-prone and time-consuming, constituting a significant portion of the entire SR process. Moreover, ML has limited application because of the shortage of training data with domain-specific multilingual corpora.

We addressed these limitations minimizing the additional effort for gathering data while maintaining the existing workflow by adopting the Bidirectional Encoder Representations from Transformers (BERT) [19] algorithm. We then used most of the data generated during SR creation.

METHOD

The data generated while working on a previous SR were used to train and validate the model. Two different tasks were tested for the two types of datasets [20–25]: datasetA with 3,268 articles [25, 26] and datasetB with 409 case studies. The first task was to classify the included articles using datasetA and its adjusted version, whose class composition was adjusted from a ratio of 9.08:1 to 2.45:1 by generating dummy data. The second task using datasetB was to extract the relations of elements (RE) from the title of the articles.

We pretrained the $srBERT_{my}$ model using abstracts of included articles with a vocabulary obtained by tokenizing the articles. We then developed other submodels ($srBERT_{my}$, $srBERT_{mix}$, and original BERT) by differentiating the pretrained data and vocabulary. Fine-tuning was applied to all the pretrained models, using the titles of included articles.

RESULTS

Our proposed srBERTmy model achieved superior results for almost every performance index in both classification and relation-extraction tasks. The $srBERT_{my}$ model, pretrained using 250,000 (355,000) steps, exhibited the highest performance in the first task, with an accuracy of 94.35% (89.38%), an area under the receiver operating characteristic curve of 0.77 (0.9), and an F1 score of 66.12 (78.64) on the original and (adjusted) datasets. In the second task, the model trained on 100,000 steps achieved an accuracy of 93.5% with a loss of 27%, outperforming the other evaluated models, including the original BERT model.

Compared with the gold standard generated by our manual work, the poorly-filtered cases were accepted as having sufficient error to require reconfirmation for title ambiguity. From these results, we verified the applicability of our model to an actual SR study and expectable synergy between the mutual studies.

DISCUSSION

Our research demonstrated the possibility of automatically classifying articles to support SR tasks and reusing BERT-based models.

However, there were limitations to consider in developing our model. We designed a multilingual model to analyze as many varied articles as possible without language restriction. Although our multilingual model was efficient, exhibiting adequate performance on datasets composed of both Chinese and English, it was challenging to assess the optimization of each language and reflection of their characteristics on the model.

Furthermore, we observed model vulnerabilities where the learning performance is highly biased by the observed data and influenced by the sufficiency of the training datasets with unique terminology. We verified that the misclassified cases were similar in that they were insufficiently secured in the training data. Data sufficiency is an inevitable challenge of natural languaging processing (NLP) models in specialized domains. Along with the increasing demand for NLP in various domains, model optimization could be improved by the cooperation of experts to build their own corpus for their field, such as bioBERT [27] and clinical BERT [28], where BERT models have been trained with a focus on the medical field.

We expect srBERT can also contribute to the literature, given that a BERT model pretrained appropriately to a field of interest could be reused by additionally training only detailed topics.

Aside from its state-of-the-art performance compared with other models, the srBERT model had the potential to be used to screen newly updated data or create new SRs if a similar corpus is used. Because fine-tuning is inexpensive in terms of computational cost compared with pretraining, this form of transfer learning enables researchers to exploit powerful deep neural network models without having access to a high-end computing environment.

LIST OF ABBREVIATIONS

SR: Systematic Review

NLP: Natural Language Processing

BERT: Bidirectional Encoder Representations from Transformers

TPU: Tensor Processing Unit

AUC: Area under The Curve

RE: Relation Extraction

SVC: Support Vector Classification

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable

CONSENT FOR PUBLICATION

Not applicable

AVAILABILITY OF DATA AND MATERIALS

The pre-trained data format and weights of srBERT are available at https://github.com/SEONCHOE/

COMPETING INTERESTS

The authors declare that they have no competing interests

AUTHOR CONTRIBUTIONS

SC conducted learning data extraction, organized the results and wrote the manuscript. SMA applied the algorithm, conducted parameter tuning. JHK reviewed and confirmed the manuscript.

ACKNOWLEDGEMENTS

Not applicable

FUNDING ACKNOWLEDGMENTS.

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by The Healthcare Bigdata Showcase Project by Korea Disease Control and Prevention Agency in the Republic of Korea (no.4800-4848-501) and Education and Research Encouragement Fund of Seoul National University Hospital.

REFERENCES

1. Siddaway AP, Wood AM, Hedges LV. How to Do a Systematic Review: A Best Practice Guide for Conducting and Reporting Narrative Reviews, Meta-Analyses, and Meta-Syntheses. Annu Rev Psychol 2019; 70: 747-70.
2. Pati D, Lorusso LN. How to Write a Systematic Review of the Literature. Herd 2018; 11: 15-30.
3. Uman LS. Systematic reviews and meta-analyses. J Can Acad Child Adolesc Psychiatry 2011; 20: 57-9.
4. Clarke M, Hopewell S, Chalmers I. Reports of clinical trials should begin and end with up-to-date systematic reviews of other relevant evidence: a status report. J R Soc Med 2007; 100: 187-90.
5. Gupta S, Rajiah P, Middlebrooks EH, Baruah D, Carter BW, Burton KR, et al. Systematic Review of the Literature: Best Practices. Academic Radiology 2018; 25: 1481-90.
6. Cohen A, Adams C, Yu C, Yu P, Meng W, Duggan L, et al. Evidence-based medicine, the essential role of systematic reviews, and the need for automated text mining tools; 2010. p376-80.
7. Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. BMJ Open 2017; 7: e012545.
8. Jaidee W, Moher D, Laopaiboon M. Time to update and quantitative changes in the results of cochrane pregnancy and childbirth reviews. PloS one 2010; 5: e11553-e.
9. Bragge P, Clavisi O, Turner T, Tavender E, Collie A, Gruen R. The Global Evidence Mapping Initiative: Scoping research in broad topic areas. BMC medical research methodology 2011; 11: 92.
10. Snilstveit B, Vojtkova M, Bhavsar A, Stevenson J, Gaarder M. Evidence & Gap Maps: A tool for promoting evidence informed policy and strategic research agendas. J Clin Epidemiol 2016; 79: 120-9.
11. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005; 8: 19-32.
12. Qi X-S, Bai M, Yang Z-P, Ren W-R. Duplicates in systematic reviews: A critical, but often neglected issue. World Journal of Meta-Analysis 2013; 1: 97-101.
13. Qi X, Yang M, Ren W, Jia J, Wang J, Han G, et al. Find duplicates among the PubMed, EMBASE, and Cochrane Library Databases in systematic review. PloS one 2013; 8.
14. Jiang Y, Lin C, Meng W, Yu C, Cohen AM, Smalheiser NR. Rule-based deduplication of article records from bibliographic databases. Database : the journal of biological databases and curation 2014; 2014: bat086-bat.
15. Kiritchenko S, de Bruijn B, Carini S, Martin J, Sim I. ExaCT: automatic extraction of clinical trial characteristics from journal publications. BMC Med Inform Decis Mak 2010; 10: 56.
16. Thomas J, McNaught J, Ananiadou S. Applications of text mining within systematic reviews. Research Synthesis Methods 2011; 2: 1-14.
17. Ananiadou S, Rea B, Okazaki N, Procter R, Thomas J. Supporting Systematic Reviews Using Text Mining. Social Science Computer Review 2009; 27: 509-23.
18. Wallace BC, Small K, Brodley CE, Lau J, Trikalinos TA. Deploying an interactive machine learning system in an evidence-based practice center: abstrackr. In: Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium. Miami, Florida, USA: Association for Computing Machinery; 2012. p819–24.
19. Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In; 2018. parXiv:1810.04805.
20. Wang P, Yang J, Liu G, Chen H, Yang F. [Effects of moxibustion at head-points on levels of somatostatin and arginine vasopressin from cerebrospinal fluid in patients with vascular dementia: a randomized controlled trial]. Zhong Xi Yi Jie He Xue Bao 2010; 8: 636-40.
21. Chen H, Wang P, Yang J, Liu G. [Impacts of moxibustion on vascular dementia and neuropeptide substance content in cerebral spinal fluid]. Zhongguo Zhen Jiu. 2011;31:19-22. Chinese.
22. Ganghui LYJ. Effects of combination of acupuncture and moxibustion with chinese drugs on lipid peroxide and antioxidase in patients of vascular dementia. World Journal of Acupuncture-Moxibustion 1998.
23. Liang Y. Effect of acupuncture-moxibustion plus chinese medicinal herbs on plasma TXB2, 6-Keto-PGF1α in patients with vascular dementia. World Journal of Acupuncture-Moxibustion 1998.
24. Wang P, Yang J, Liu G, Chen H, Yang F. [Effects of moxibustion at head-points on levels of somatostatin and arginine vasopressin from cerebrospinal fluid in patients with vascular dementia: a randomized controlled trial]. Zhong Xi Yi Jie He Xue Bao. 2010;8:636-40. Chinese. doi:10.3736/jcim20100706.
25. Choe S, Cai M, Jerng UM, Lee J-H. The Efficacy and Underlying Mechanism of Moxibustion in Preventing Cognitive Impairment: A Systematic Review of Animal Studies. Experimental neurobiology 2018; 27: 1-15.
26. Aum S, Choe S, Cai M, Jerng UM, Lee J-H. Moxibustion for cognitive impairment: a systematic review and meta-analysis of animal studies. Integrative Medicine Research 2020: 100680.

27. Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. In; 2019. parXiv:1901.08746.

28. Alsentzer E, Murphy JR, Boag W, Weng W-H, Jin D, Naumann T, et al. Publicly Available Clinical BERT Embeddings. 2019.