Analyzing health risk on technology of mRNA vaccine by integrating publications and Twitter data

Munan LI [a,b];   Liang WANG [a]
a.School of Business Administration, South China University of Technology, Guangzhou China
b.Guangdong key laboratory on innovation methods and decision management system, Guangzhou China

Abstract: The vaccine race brought about by the COVID-19 pandemic has greatly accelerated the research and development (R&D) of mRNA vaccine and the corresponding technologies. In 2021, The MIT Technology Review (MTR) ever reported that mRNA vaccine has become one of ten breakthrough technologies in the world. However, either some views from the research communities and social media comments delivered a certain worry about the health risk caused by the emerging technology of mRNA vaccine. According to some sound from the publications and social media, although the injection of mRNA vaccine has certain resistance to the specified virus, it also has some side effects, which may bring certain health risks for the objects of injection. In order to further explore the health risks of mRNA vaccine, this paper proposed a novel model for named entity recognition (NER) to capture risk signals based on two data sources: Twitter and Web of Science (WOS). Firstly, the text analysis capability of PubMedBERT in biomedical field is utilized to extract text features and generate dynamic word vectors. Then, conditional random field (CRF) is used to predict the results of tag sequences. In order to solve the problems of small size of data and difficulty in entity label mark about risk entity recognition (RER), the method of transfer learning is conducted to pre-train the model based on i2b2-2010 dataset that is popular in the medical informatics, and then fine-tune the obtained pre-training model in the text data, which effectively improves the recognition effect. Through empirical analysis, the text model has a certain effect on the health risk identification about vaccine technology of mRNA, which also outputs a certain implication for both vaccine production and vaccine injection population, and the corresponding policy-making decision.