Journal of the American Medical Informatics Association : JAMIA, 2021, vol.28, N10, pages 2184-2192

DeepADEMiner: a deep learning pharmacovigilance pipeline for extraction and normalization of adverse drug event mentions on Twitter

Magge A., Tutubalina E., Miftahutdinov Z., Alimova I., Dirkson A., Verberne S., Weissenbacher D., Gonzalez-Hernandez G.
Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

OBJECTIVE: Research on pharmacovigilance from social media data has focused on mining adverse drug events (ADEs) using annotated datasets, with publications generally focusing on 1 of 3 tasks: ADE classification, named entity recognition for identifying the span of ADE mentions, and ADE mention normalization to standardized terminologies. While the common goal of such systems is to detect ADE signals that can be used to inform public policy, it has been impeded largely by limited end-to-end solutions for large-scale analysis of social media reports for different drugs. MATERIALS AND METHODS: We present a dataset for training and evaluation of ADE pipelines where the ADE distribution is closer to the average 'natural balance' with ADEs present in about 7% of the tweets. The deep learning architecture involves an ADE extraction pipeline with individual components for all 3 tasks. RESULTS: The system presented achieved state-of-the-art performance on comparable datasets and scored a classification performance of F1 = 0.63, span extraction performance of F1 = 0.44 and an end-to-end entity resolution performance of F1 = 0.34 on the presented dataset. DISCUSSION: The performance of the models continues to highlight multiple challenges when deploying pharmacovigilance systems that use social media data. We discuss the implications of such models in the downstream tasks of signal detection and suggest future enhancements. CONCLUSION: Mining ADEs from Twitter posts using a pipeline architecture requires the different components to be trained and tuned based on input data imbalance in order to ensure optimal performance on the end-to-end resolution task.

http://dx.doi.org/10.1093/jamia/ocab114

Keywords

drug safety, information extraction, natural language processing, pharmacovigilance, social media mining