Drug Safety (2019) 42:95–97
https://doi.org/10.1007/s40264-018-0766-8

EDITORIAL

Towards Drug Safety Surveillance and Pharmacovigilance: Current Progress in Detecting Medication and Adverse Drug Events from Electronic Health Records

Feifan Liu[1] · Abhyuday Jagannatha[2] · Hong Yu[2,3,4,5]

Published online: 16 January 2019

Large-scale drug safety surveillance and pharmacovigilance are key components of effective drug regulation systems, clinical practice, and public health programs [1]. Although the efficacy and safety of a drug must be demonstrated in a series of clinical trials prior to approval [2], many adverse drug events (ADEs) are detected only after a drug has been marketed when it is used by a larger and more diverse population than during clinical trials. Adverse drug events discovered after a drug is in broad use can be a significant cause of morbidity and mortality. Thus, effective and accurate post-market drug surveillance is in urgent demand for the protection of public health and the reduction of healthcare expenditures due to ADE-related hospital complications [3–5].

Spontaneous reporting systems [6–9] have been traditionally used for pharmacovigilance. However, this type of data is inherently passive because except for drug companies' spontaneous reporting systems, reporting is voluntary, and studies have shown that as many as 90% of serious ADEs go unreported [10]. Electronic health records (EHRs) contain real-time real-world clinical data gathered during routine clinical care, offering a potentially more proactive approach to pharmacovigilance [2]. Therefore, EHRs for post-market surveillance play an important role in the new paradigm of drug regulation [11]. More importantly, compared with structured data or coded data in EHRs, unstructured clinical narratives provide more information on ADE documentation. A study shows that only 9020 (28.6%) out of 31,531 patients with documented statin side effects had the relevant ADE recorded in a structured format [12]. Therefore, developing advanced natural language processing (NLP) techniques to unpin ADE information from EHR narratives will greatly facilitate proactive, accurate, and efficient post-market drug safety monitoring on a large scale.

In 2010, i2b2 partnered with the VA Salt Lake City Health Care System and organized an NLP open challenge [13], which supported community efforts in applying NLP to extract medication and treat targets and caused adverse events from EHR narratives. However, the annotation schema defined for that challenge only covers a limited set of entities relevant for pharmacovigilance. To further/better assess the current methodological progress in this research area, we organized the "NLP Challenge for Detecting Medication and Adverse Drug Events from Electronic Health Records (MADE1.0)" in 2018, which offers larger scale, expert-annotated clinical notes labeled with more fine-grained clinically named entities and relations related to drug safety surveillance. There are 15 teams from seven countries registered in this challenge and in total 41 runs from 11 teams were submitted.

Part of this theme issue of *Drug Safety* is to present recent advances in mining unstructured information from clinical narratives in the context of drug safety surveillance and pharmacovigilance. There are five articles from the MADE1.0 challenge, including an overview paper and four

Part of a theme issue on "NLP Challenge for Detecting Medication and Adverse Drug Events from Electronic Health Records (MADE 1.0)" guest edited by Feifan Liu, Abhyuday Jagannatha and Hong Yu.

✉ Hong Yu
hong_yu@uml.edu

1 Department of Quantitative Health Sciences and Radiology, University of Massachusetts Medical School, Worcester, MA, USA

2 College of Information and Computer Sciences, University of Massachusetts, Amherst, MA, USA

3 Department of Computer Science, University of Massachusetts, 220 Pawtucket St, Lowell, MA 01854-2874, USA

4 Department of Medicine, University of Massachusetts Medical School, Worcester, MA, USA

5 Bedford Veterans Affairs Medical Center, Bedford, MA, USA

research papers invited from top performance teams participating in the challenge.

The first paper by Jagannatha et al. [14] provides an overview of the MADE1.0 challenge. First, the article describes the MADE1.0 corpus, including the details about the annotation process and a comprehensive annotation schema. The authors report the Fleiss's Kappa score of 0.628 and 0.424 for the inter-annotator agreement on named entity annotation and relation annotation, respectively. Second, the authors introduce the three subtasks defined in the challenge: Named Entity Recognition (NER), Relation Identification (RI), and Joint Relation Extraction (NER-RI), followed by a comprehensive report of system submissions for the challenge. Finally, an ensemble-based system aggregation has shown improved performance, suggesting that the top performing systems develop in a different but complementary manner.

Wunnava et al. [15] present a three-layer deep learning architecture for the NER subtask, consisting of a bi-directional long short-term memory (BiLSTM) layer for character-level encoding, a BiLSTM layer for word-level encoding, and a conditional random fields (CRF) layer for structured prediction. To better handle the noisy format of clinical notes, they built a rule-based sentence and word tokenizer leading to a better performance compared with using an off-the-shelf Natural Language Toolkit [16]. Their system achieved the best micro F1 score of 0.829 for NER, and they found character-level encoding and CRF sequence inference contribute to performance improvement.

Yang et al. [17] applied a similar BiLSTM-CRF structure for NER, which is combined with a support vector, machine-based relation extraction system to address all the three tasks. They trained BiLSTM-CRF in two stages: optimize parameters based on validation data and then train a final model using both validation and training data, leading to better results than using the validation-optimized model directly. Their experiments demonstrate that developing separate classifiers to handle intra- and inter-sentence relations respectively obtained better performance (F1 score of 0.8466) than one single classifier for both (F1 score of 0.8304).

Dandala et al. [18] employed two deep learning architectures in their challenge: a BiLSTM-CRF model for NER and a BiLSTIM model with an attention mechanism for RI. In addition to character/word level embeddings, part-of-speech embeddings were also utilized for input encoding in both models. Based on the observation that "adverse drug events" and "indications" entities have a semantic overlap with "other sign and symptoms", they experimented with a joint modeling method where those three types of entities are first merged into one category for the NER model and their relations with medications determined by the RI model were in turn used to distinguish those three types of entities. Experimental results show the joint modeling approach outperformed the standard sequential model for the integrated NER-RI subtask (micro F1 of 0.653 vs. 0.624).

Peterson et al. [19] explored traditional machine learning models, CRF for NER and random forest for RI, which were shown more computationally efficient and thus easily deployed in real-world applications without depending on a special high-performing infrastructure. As part of the feature engineering effort, they included word embeddings as clustering features trained with Mini-batch K-means, in which multiple cluster sizes and compound cluster features were also examined. Compared with the counterpart deep learning models, their system achieved competitive overall results through effective feature engineering, yielding the best micro F1 of 0.8684 for the RI subtask.

While the performance reported in this challenge is promising, there is much room for further improvement, especially for the complex joint NER-RI task. The design of better learning algorithms and the availability of more labeled data are two important aspects contributing to improved system performance. Another future direction is to validate and increase the existing systems' generalizability on larger scale datasets from diverse clinical subspecialties. That may require more effort in building annotated data as well as exploring effective domain adaption techniques for data-scarce subspecialties. Finally, it would be essential to investigate how to effectively integrate a large volume of diverse, dynamic, distributed structured or unstructured data from different sources such as spontaneous reporting system reports, EHRs, insurance claims, medical literature, and social media for collective ADE signal detection.

Data mining EHRs for drug safety surveillance, especially mining unstructured narratives through NLP, will remain an active research topic. The innovative approaches reported in this issue, which were motivated by the MADE1.0 challenge, will lay a solid foundation for further advancing methodological development and system deployment towards more intelligent drug safety surveillance.

Acknowledgements We thank the editor and the manuscript authors for their contributions to this issue. We also thank all the reviewers for their comments and thoughtful suggestions for improving the submitted drafts.

Compliance with Ethical Standards

Funding It was supported by the National Heart, Lung, and Blood Institute (NHLBI) of the National Institutes of Health (R01HL125089). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest Feifan Liu, Abhyuday Jagannatha, and Hong Yu have no conflicts of interest that are directly relevant to the content of this editorial.

References

1. Jeetu G, Anusha G. Pharmacovigilance: a worldwide master key for drug safety monitoring. J Young Pharm. 2010;2:315–20.
2. Coloma PM, Trifirò G, Patadia V, Sturkenboom M. Postmarketing safety surveillance: where does signal detection using electronic healthcare records fit into the big picture? Drug Saf. 2013;36:183–97.
3. Bates DW, Spell N, Cullen DJ, Burdick E, Laird N, Petersen LA, et al. The costs of adverse drug events in hospitalized patients. JAMA. 1997;277:307.
4. Nebeker JR, Hoffman JM, Weir CR, Bennett CL, Hurdle JF. High rates of adverse drug events in a highly computerized hospital. Arch Intern Med. 2005;165:1111–6.
5. Hakkarainen KM, Hedna K, Petzold M, Hägg S. Percentage of patients with preventable adverse drug reactions and preventability of adverse drug reactions: a meta-analysis. PLoS One. 2012;7:e33236.
6. Polepalli Ramesh B, Belknap SM, Li Z, Frid N, West DP, Yu H. Automatically recognizing medication and adverse event information from Food and Drug Administration's Adverse Event Reporting System narratives. JMIR Med Inform. 2014;2:e10.
7. Botsis T, Nguyen MD, Woo EJ, Markatou M, Ball R. Text mining for the Vaccine Adverse Event Reporting System: medical text classification using informative feature selection. J Am Med Inform Assoc. 2011;18(5):631–8.
8. Lindquist M. VigiBase, the WHO global ICSR database system: basic facts. Drug Inf J. 2008;42:409–19.
9. Alvarez Y, Hidalgo A, Maignen F, Slattery J. Validation of statistical signal detection procedures in Eudravigilance post-authorization data: a retrospective evaluation of the potential for earlier signalling. Drug Saf. 2010;33:475–87.
10. Hazell L, Shakir SAW. Under-reporting of adverse drug reactions: a systematic review. Drug Saf. 2006;29:385–96.
11. Moore TJ, Furberg CD. Electronic health data for postmarket surveillance: a vision not realized. Drug Saf. 2015;38:601–10.
12. Skentzos S, Shubina M, Plutzky J, Turchin A. Structured vs. unstructured: factors affecting adverse drug reaction documentation in an EMR repository. AMIA Annu Symp Proc. 2011;2011:1270–9.
13. Uzuner Ö, South BR, Shen S, DuVall SL. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. J Am Med Inform Assoc. 2011;18:552–6.
14. Jagannatha A, Liu F, Liu W, Yu H. Overview of the first natural language processing challenge for extracting medication, indication, and adverse drug events from electronic health record notes (MADE 10). Drug Saf. 2019. https://doi.org/10.1007/s40264-018-0762-z.
15. Wunnava S, Qin X, Kakar T, Sen C, Rundensteiner EA, Kong X. Adverse drug event detection from electronic health records using hierarchical recurrent neural networks with dual-level embedding. Drug Saf. 2019. https://doi.org/10.1007/s40264-018-0765-9.
16. Bird S, Klein E, Loper E. Natural language processing with Python: analyzing text with the natural language toolkit. Newton: O'Reilly Media, Inc; 2009. http://www.datascienceassn.org/sites/default/files/Natural%20Language%20Processing%20with%20Python.pdf.
17. Yang X, Bian J, Gong Y, Hogan WR, Wu Y. MADEx: a system for detecting medications, adverse drug events, and their relations from clinical notes. Drug Saf. 2019. https://doi.org/10.1007/s40264-018-0761-0.
18. Dandala B, Joopudi V, Devarakonda M. Adverse drug events detection in clinical notes by jointly modeling entities and relations using neural networks. Drug Saf. 2019.
19. Chapman AB, Peterson KS, Alba PR, DuVall SL, Patterson OV. Detecting adverse drug events with rapidly trained classification models. Drug Saf. 2019. https://doi.org/10.1007/s40264-018-0763-y.