De-identification of free-text descriptions of suspected adverse drug reactions using deep learning

Eva-Lisa Meldau, Sara Hedfors Vidlin, Lucie Gattepaille, Henric Taavola, Lovisa Sandberg, Yasunori Aoki, G. Niklas Norén
Uppsala Monitoring Centre, Uppsala, Sweden

1 Introduction

Automatic de-identification of free-text descriptions of suspected adverse drug reactions may enable analyses that help improve patient safety without compromising confidentiality. New medicines are extensively tested before they are approved for real-world use, but additional adverse reactions are often identified once broader groups of patients begin to use the medicines in less carefully controlled settings. This is the focus of pharmacovigilance, which is the science and activities relating to the detection, assessment, understanding, and prevention of adverse effects or any other drug-related problem. VigiBase is the World Health Organization's global database of individual case safety reports. As of May 2018, VigiBase contains 17 million reports from 131 countries. These reports have been submitted by health professionals and patients, and include both structured data and free-text narrative descriptions of the adverse events experienced by the patient. Both the structured data and the free-text descriptions can be crucial to evaluate the link between the medicine and the event [Karimi *et al*., 2014], as well as its impact, but concerns about patient privacy and confidentiality may limit their availability to scientists and professionals working to ensure safe and effective use of medicines. If the free-text descriptions could be de-identified by automatically removing personal identifiers, they could be shared more reliably and allow more thorough evaluation of suspected adverse drug reactions.

De-identification of free-text descriptions is more challenging than the de-identification of structured data, since elements that serve as personal identifiers must first be identified in the free text using natural language processing. In recent years, advances in deep learning have enabled improved natural language processing for various tasks [Collobert *et al*., 2011; Mikolov *et al.*, 2010] and some researchers have explored the use of deep neural networks specifically for de-identification [Dernoncourt *et al.*, 2017; Li *et al.*, 2016]. We are not aware of research projects looking at the de-identification of free-text descriptions of suspected adverse drug reactions, but the related task of de-identifying free-text descriptions in medical records was the focus of the i2b2 challenge in 2014 [Stubbs and Uzuner, 2015], which also provides an annotated reference set for training and validation.

2 Aim

To develop a method for de-identification of free-text narrative descriptions of suspected adverse drug reactions; presented here, are intermediate results of ongoing research.

3 Method

Our de-identification method is an ensemble which combines the predictions of several models through logistic regression. The ensemble assigns the binary label 'personal' or 'non-personal' to each token of the narrative, and for this study, we considered names, dates, and locations as personal identifiers, and all other tokens as not personal identifiers ('non-personal'). Our ensemble method includes deep neural networks trained to assign different labels to the tokens in the narrative, such as 'Patient name', 'Doctor name', 'Date' and 'City'. The first layer of all the deep neural networks used here is an embedding layer that is pre-trained using GloVe word embeddings [Pennington *et al.*, 2014]. The next layer is a Long Short-Term Memory in some networks, and three stacked Convolutional Neural Network layers in the others. A probability output for each of the labels is produced by a dense layer with a softmax function. If the 'non-personal' label is predicted with a probability greater than or equal to 90% then the final output of that neural network is 'non-personal'. Otherwise, its final output is the 'personal' label with highest probability. The ensemble also includes a logistic regression model, based on a window of words around the token of interest, which selects the output label in the same way as the neural networks. The ensemble further includes a standard part-of-speech tagger, as well as a rule-based component that uses regular expressions and dictionary lookups to label 'personal' and 'non-personal' parts of the text.

Large amounts of annotated data are required for training deep neural networks but, at this point of our research, there is a shortage in annotated VigiBase narratives. We therefore used the i2b2 de-identification challenge data set as a starting point for our model training. Fine-tuning the neural networks on VigiBase data will be explored during future work. We

trained the individual algorithms on the 521 medical records in the training set of the 2014 i2b2 de-identification challenge data, and we trained the ensemble on 3/4 of the 269 records in the validation set from the same challenge. The method was evaluated on the remaining 1/4 of the i2b2 2014 validation set which was held out for that purpose, as well as on 300 VigiBase narratives that had been annotated as part of an earlier research project [Sahlström, 2015].

4 Results

In our evaluation against the held-out 1/4 of the i2b2 validation data, 95% of the personal identifiers was correctly recalled with a precision of 90%. On VigiBase narratives, the method trained on i2b2 data reached a recall of 90% and a precision of 45% for personal identifiers. For both the i2b2 data set and for VigiBase, partial misses of personal identifiers such as not removing 'of' in '4th of May' or not removing 'Home' in 'Perth Nursing Home' count as false negatives. This decreases nominal recall even though these tokens on their own do not convey enough information to compromise patient confidentiality.

5 Conclusions

The proposed method can remove personal identifiers in free-text narrative descriptions of suspected adverse drug reactions with high recall and intermediate precision even though it was trained only on medical records. Fine-tuning, and possibly re-training parts of, the method directly on reports of suspected adverse drug reactions can be expected to improve performance even further.

Acknowledgments

The authors are indebted to the national centres who make up the WHO Programme for International Drug Monitoring and contribute reports to VigiBase. However, the opinions and conclusions of this study are not necessarily those of the various centres nor of the WHO.

References

[Collobert *et al.*, 2011] Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 12(Aug):2493-2537, 2011.

[Dernoncourt *et al.*, 2017] Franck Dernoncourt, Ji Young Lee, Ozlem Uzuner, and Peter Szolovits. De-identification of patient notes with recurrent neural networks. *Journal of the American Medical Informatics Association*, 24(3):596-606, 2017.

[Karimi *et al.*, 2014] Ghazaleh Karimi, Kristina Star, Marie Lindquist, and I Ralph Edwards. Clinical stories are necessary for drug safety. *Clinical Medicine*, 14(3):326-327, 2014.

[Li *et al.*, 2016] Kun Li, Yumei Chai, Hongling Zhao, Xiaofei Nan, and Yueshu Zhao. Learning to Recognize Protected Health Information in Electronic Health Records with Recurrent Neural Network.. In *Natural Language Understanding and Intelligent Applications*, Springer, 2016.

[Mikolov *et al.*, 2010] Tomáš Mikolov, Martin Karafiát, Lukáš Burget, Jan Černocký, and Sanjeev Khudanpur. Recurrent neural network based language model. In *Eleventh Annual Conference of the International Speech Communication Association*, 2010.

[Pennington *et al.*, 2014] Jeffrey Pennington, Richard Socher, and Christopher Manning. GloVe: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014.

[Sahlström, 2015] Jakob Sahlström. Automatic de-identification of case narratives from spontaneous reports in VigiBase (M.S. thesis). Uppsala University, 2015.

[Stubbs and Uzuner, 2015] Amber Stubbs A, Özlem Uzuner. Annotating longitudinal clinical narratives for de-identification: The 2014 i2b2/UTHealth corpus. *Journal of Biomedical Informatics*, 58(Suppl): S20-29, 2015.