Language-agnostic pharmacovigilant text mining to elicit side effects from clinical notes and hospital medication records

Benjamin Kaas-Hansen[1], Davide Placido[2], Cristina Rodríguez[2], Hans-Christian Thorsen-Meyer[3], Simona Gentile[4], Anna Nielsen[2], Søren Brunak[2], Gesche Jürgens[1], and Stig Andersen[1]

[1]Zealand University Hospital Roskilde
[2]University of Copenhagen
[3]Copenhagen University Hospital
[4]Region Zealand

January 29, 2022

Abstract

Aim To create a drug safety signalling pipeline associating latent information in clinical free text with exposure profiles to highlight potential adverse drug reactions to single drugs and drug pairs. Methods All inpatient visits of a 500,000-patient sample from two Danish regions, between 18 May 2008 and 30 June 2016. Tokens from clinical notes recorded within 48 hours of admission were operationalised with a fastText embedding. For each of the 10,720 single-drug and drug-pair exposures from doorstep medication profiles, we trained a feed-forward neural network predicting the risk of exposure using embedding vectors as inputs. Results 2,905,251 inpatient visits comprised 13,740,564 doorstep drug prescriptions; the median number of prescriptions was 5 (IQR: 3-9) and in 1,184,340 (41%) admissions patients used [?]5 drugs concurrently. 10,788,259 clinical notes were included, with 179,441,739 tokens retained after pruning. Of 345 single-drug signals reviewed, 28 (8.1%) represented possibly undescribed relationships; 186 (54%) signals were clinically meaningful. 16 (14%) of the 115 drug-pair signals were possible interactions and 2 (1.7%) were known. Conclusion We built a language-agnostic pipeline for mining associations between free-text information and medication exposure without manual curation, by predicting not the likely outcome of a range of exposures, but the likely exposures for outcomes of interest. Our approach may help overcome limitations of text mining methods relying on curated data in English and makes our method appealing in settings that must make sense of non-English free text for pharmacovigilance.

Hosted file

`2022-01-19 - Manuscript.docx` available at https://authorea.com/users/458002/articles/554614-language-agnostic-pharmacovigilant-text-mining-to-elicit-side-effects-from-clinical-notes-and-hospital-medication-records