

Automatic adverse drug reaction extraction from electronic health records

The IXA group at the University of the Basque Country (UPV/EHU) is using natural language processing to extract adverse effects from health records written in Spanish

RESEARCH 09/03/2021



Sara Santiso. Photo: Mitxi. UPV/EHU

Researchers from the IXA group at the UPV/EHU are collaborating with Osakidetza (the Basque Regional Health Service) to create a system for automatically extracting adverse drug reactions from electronic health records written in Spanish. The researchers have conducted different tests using both machine learning and deep learning, with the aim of building a robust model for extracting relations between drug-disease pairs based on clinical text mining.

Patients' electronic health records convey crucial information. The application of natural language processing techniques to these records may be an effective means of extracting information that may improve clinical decision making, clinical documentation and billing, disease prediction and the detection of adverse drug reactions. Adverse drug reactions are a major health problem, resulting in hospital re-admissions and even the death of thousands of patients. An automatic detection system can highlight said reactions in a document, summarise them and automatically report them.

In this context, the Basurto University Hospital and the Galdakao Hospital 'were interested in creating a system that would use natural language processing techniques to analyse patient health records in order to automatically identify any adverse effects' explains the engineer Sara Santiso, who also holds a PhD in Computer Science. After the hospitals contacted the IXA group at the UPV/EHU, several researchers started working to build a robust model with which to extract adverse drug reactions from electronic health records written in Spanish, based on clinical text mining.

To this end 'not only have we used techniques based on traditional machine learning algorithms, we have also explored deep learning techniques, reaching the conclusion that these are better able to detect adverse reactions' explains Santiso, one of the authors of the study. Machine learning and deep learning imitate the way the human brain learns, although they use

different types of algorithms to do so.

Difficulties finding a corpus in Spanish

Santiso underscores the difficulties the team encountered when trying to find a large enough corpus with which to work: 'At first, we started with only a few health records, because they are difficult to obtain due to privacy issues; you have to sign confidentiality agreements in order to work with them' she explains. The research team has found that 'having a larger corpus helps the system learn the examples contained in it more effectively, thereby giving rise to better results'.

Through this study, which was carried out with health records written in Spanish, 'we are contributing to closing the gap between clinical text mining in English and that carried out in other languages, which accounts for less than 5% of all papers published in the field. Indeed, the extraction of clinical information is not yet fully developed due (among other things) to the potential for extracting information from other hospitals and in other languages' claims the researcher.

Although natural language processing has been of inestimable help in the computer-aided detection of adverse drug reactions, there is still room for improvement: 'To date, systems have tended to focus on detecting drug-disease pairs located in the same sentence. However, health records contain implicit information that might reveal underlying relations (for example, information about antecedents might be relevant for determining the causes of an adverse event). In other words, future research should strive to detect both explicitly and implicitly-stated inter-sentence relationships'. Moreover, another issue that should be the subject of future research is the lack of electronic health records written in Spanish.

Complementary information

This study was carried out by Sara Santiso, under the supervision of Alicia Pérez and Arantza Casillas, from the IXA research group at the University of the Basque Country (UPV/EHU).

bibliographic reference

Sara Santiso * , Alicia Pérez, Arantza Casillas

Adverse Drug Reaction extraction: Tolerance to entity recognition errors and sub-domain variants

Computer Methods and Programs in Biomedicine

DOI (Digital Object Identifier): 10.1016/j.cmpb.2020.105891

sharing

