

Visualising Patterns Associated with Adverse Drug Reactions in French Forums

Nour ALLAM^{a,b}, Bissan AUDEH^a, Marie-Christine JAULENT^a and Cedric BOUSQUET^{a,1}

^a*Sorbonne Université, INSERM, Université Paris 13, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé, Paris, France*

^b*ESIEE Paris graduate school of engineering, Noisy-le-Grand, France*

Abstract. As social media are an interesting source of information for pharmacovigilance, we implemented a novel visualisation method for pharmacovigilance specialists applied to French discussion forums. A word embedding model was trained on posts to facilitate the identification of patterns associated with adverse drug reactions.

Keywords. Natural Language Processing, Pharmacovigilance, Word Embedding, Machine Learning, Drug Surveillance

1. Introduction

Nowadays, it has become frequent for patients to describe adverse drug reactions in discussion forums. As a result, an immense amount of posts might contain precious information for pharmacovigilance. Reviewing such voluminous data by pharmacovigilance specialists is almost impossible without the use of natural language processing and machine learning techniques. Before applying these techniques, detecting and visualizing patterns in data is a major step in exploratory data analysis. This step allows better understanding of data which leads to better strategy for further knowledge extraction and interpretation. We propose an unsupervised learning approach for visualizing patterns associated with adverse drug reactions. Our visualisation is based on the output of word2vec [1], a word embedding language model that regroups tokens that co-occur in similar contexts. We used a sample from a dataset consisting of co-occurrences between drugs and medical conditions previously annotated using a supervised approach [2]. Then we performed 2D representations of word embeddings of these annotations to figure if it is possible to highlight relevant patterns associated with adverse drug reactions.

2. Methods and Results

A total of 665.662 messages were extracted from the medical discussion forum Doctissimo² from January 2020 to June 2020. After preprocessing and tokenizing the

¹ Corresponding Author, Cédric Bousquet; E-mail: cedric.bousquet@chu-st-etienne.fr.

² French medical discussion forum <www.doctissimo.fr>.

