

NLP in Support of Pharmacovigilance

Fabio Rinaldi and **Lorenzo Ruinelli** and **Roberta Noseda**
Oscar William Lithgow Serrano and **Sandra Mitrovic**
fabio@idsia.ch

Abstract

Pharmacovigilance (PV) is the science and activities related to the detection, assessment, understanding, and prevention of adverse effects of drugs, or of any other drug related problem. It is a critical component of drug safety, and it is essential for the safe and effective use of medicines.

In Switzerland Healthcare professional who suspect an adverse drug reaction in their patients are legally required to report it to one of the five PV center. In those centers, PV experts manually assess the validity of received alerts to further process only retained “active” spontaneous safety reports, thus making causality assessment and sending the information to Swissmedic.

In this project, we aim to investigate the potential of state of the art NLP methods applied to the automated processing of clinical records in order to provide reliable signals for the potential presence of adverse drug reactions (ADRs) in the context of pharmacovigilance. Concretely, if the project is successful, the partner hospital will be able to significantly improve the detection of ADRs, and once a similar approach is adopted at other Swiss hospitals, it could have a major downstream impact on public health.

The project is structured as a collaboration between the the regional PV Center of Southern Switzerland, based at the Ticino Cantonal Hospital (EOC), and the NLP group of the Dalle Molle Institute of Artificial Intelligence in Lugano. The PV center will be responsible for collecting and annotating the datasets necessary for the project.

We are starting from an initial (imperfect) version of the annotations obtained through the method of distant supervision. Specifically, the data present in the existing reports will be automatically mapped (as far as possible) onto the original discharge letters from which they were manually extracted.

The problem is framed as two distinct NLP tasks. The first task consists in distinguishing discharge letters which contain ADRs (which should be further processed) from those which do not, it is thus a simple classification problem. The second task consists in the identification of the elements that need to reported to swissmedic, such as drugs administered to the patient, their dosages, comorbidities, age and sex, etc., it is thus a named entity recognition problem.

The project aims at establishing a solid baseline for both tasks, using traditional, efficient, and low cost technologies, before exploring the cost-benefit trade-off of employing LLMs.