

running the self-management planning module (4) for the first time. The plan for self-management will be updated and revised as more patient data is added. The activity logging module (2) runs continuously since it provides data from the wristband regarding the patient's physical activity and sleep patterns. The main goal of this module is to verify whether the patient follows the suggestions in the recommended plan (e.g., for physical activity). As for the initial screening data, the activity log data is also sent to the server, and becomes input to the periodic run of the self-management planning module (4). The user interaction module (3) is basically a question/answering module, typically initiated by the patient when using the SELFBACK app and browsing through the given information.

Overall, the aim of the SELFBACK system is to solve the problems of reinforcement and adherence to a self-management plan by offering an evidence-based system that allows personalized follow-up and advice to the patient.

Link:

SELFBACK project:
<http://research.idi.ntnu.no/selfback/>

References:

- [1] P. Wändell et al.: "Most common diseases diagnosed in primary care in Stockholm", Sweden, in 2011. *Fam Pract* 30:506-513, 2013.
- [2] M. van Tulder et al.: "Chapter 3. European guidelines for the management of acute nonspecific low back pain in primary care. *Eur Spine J* 15:s169-s191, 2006.

Please contact:

Kerstin Bach, Paul Jarle Mork, Agnar Aamodt
Norwegian University of Science and Technology
E-Mail: kerstin.bach@idi.ntnu.no,
paul.mork@ntnu.no,
agnar.aamodt@idi.ntnu.no

Twitter can Help to Find Adverse Drug Reactions

by Mark Cieliebak, Dominic Egger and Fatih Uzdilli

Drugs are great! We all need and use drugs every now and then. But they can have unwanted side-effects, referred to as "adverse drug reactions" (ADRs). Although drug manufacturers run extensive clinical trials to identify these ADRs, there are still over two million serious ADRs in the U.S. every year – and more than 100,000 patients in the U.S. die due to drug reactions, according to the U.S. Food and Drug Administration (FDA) [1]. For this reason, we are searching for innovative and effective ways to find ADRs.

Identifying ADRs is an important task for drug manufacturers, government agencies, and public health. One way to identify them before a drug goes to market is through clinical trials. Governments worldwide also have diverse surveillance programs in order to identify ADRs once the drugs are in use by consumers. For example, official websites such as MedWatch allow both patients and drug providers to submit ADRs manually. However, only a very small fraction of all ADRs is submitted to these systems – experts estimate that over 90% of all reactions go unreported.

Twitter can help!

On the other hand, there are millions of messages on Twitter that discuss medications and their side-effects. These messages contain data on drug usage in much larger test sets than any clinical trial will ever have. Inspired by this, research teams worldwide, including our team at Zurich University of Applied Sciences, are beginning to utilize these messages for ADR detection. The goal is

to automatically find relevant messages, to "understand" their content, and to extract structured data about the drugs and (unwanted) reactions.

A typical approach for ADR detection uses Natural Language Processing (NLP) to analyze tweets automatically. Input for the system is the entire stream of Twitter messages. Each individual tweet is analyzed, using a classification system as shown in Figure 1: The tweet is preprocessed and a set of relevant properties ("features") is extracted. Then, a classifier decides whether the tweet mentions an ADR. This classifier is based on machine learning and was trained beforehand on thousands of sample tweets that were tagged by humans. Finally, a system for named entity extraction is used to output a drug name and associated ADRs.

This approach is similar to technologies for sentiment analysis, which decide whether a tweet is positive or negative. Sentiment analysis is already success-

fully applied, for instance, in market monitoring, customer support and social media analysis.

State-of-the-art

Our ADR system, which implements the technologies shown above, achieves a success rate of 32% (measured in F1-score). This is comparable to other academic ADR systems: in an open international competition this year, even the best systems achieved only a success rate of approximately 40% [2].

For a preliminary evaluation on real-world data, we applied our ADR system to the full Twitter stream. The low precision of the system resulted in 20% of all tweets being classified as ADR. This is way too high; there are not that many ADR tweets in the Twitter stream. For this reason, we pre-filtered the stream with a list of 1678 drug names. Out of about 50 million tweets, this resulted in 13,000 tweets referencing drugs. Using the ADR system on this reduced set yielded 2800 tweets. We expect to find

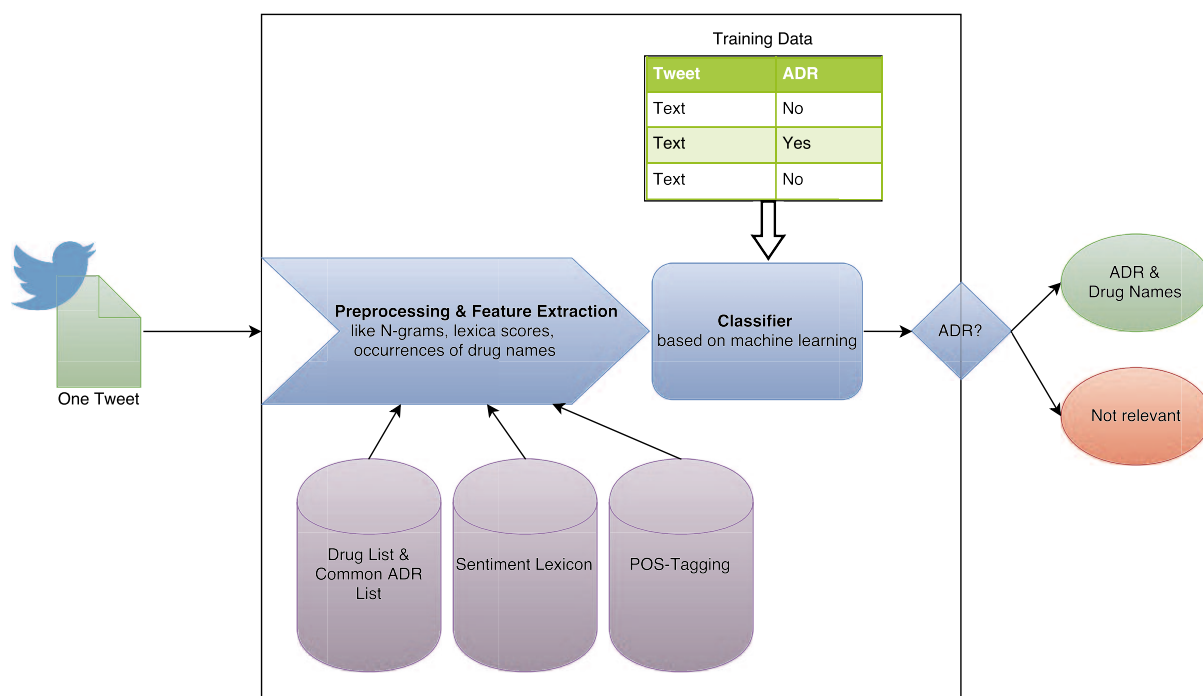


Figure 1: Typical system for ADR detection using machine learning.

more than 60% of these to be true ADR tweets.

Future improvements

Automatic detection of ADRs on Twitter (or other social media channels) is still a very young discipline, which only started some five years ago. There are only a few teams working on the topic at the moment, and a first large-scale benchmark dataset was only published in 2014 [3]. However, we expect a significant improvement in detection rates in the near future, owing in part to the existence of several new technologies in machine learning, such as word embedding and deep learning. These have already been successfully applied to other text analysis tasks and have

improved existing benchmark scores there. Applying these technologies to ADR detection will probably help to increase the detection rate significantly. In addition, our team is working on a system that not only analyzes the text of a tweet, but also its context: the timeline of the user, other messages in the same geographic or temporal context etc. This will allow us to “step back” from an isolated event (a single tweet) and see the “whole picture” of the discourse on Twitter.

References:

- [1] J. Lazarou, B.H. Pomeranz, and P.N. Corey: “Incidence of Adverse Drug Reactions in Hospitalized Patients: A Meta-analysis of

Prospective Studies”, JAMA 279(15):1200-1205, 1998.

- [2] Pacific Symposium on Biocomputing: <http://psb.stanford.edu/workshop/wkshp-smm/>

- [3] R. Ginn et al.: “Mining Twitter for Adverse Drug Reaction Mentions: A Corpus and Classification Benchmark”, BioTxtM, 2014.

Please contact:

Mark Cieliebak
School of Engineering
Zurich University of Applied Sciences (ZHAW)
Tel: +41 58 934 72 39
E-mail: ciel@zhaw.ch

Trust for the “Doctor in the Loop”

by Peter Kieseberg, Edgar Weippl and Andreas Holzinger

The “doctor in the loop” is a new paradigm in information driven medicine, picturing the doctor as authority inside a loop supplying an expert system with data and information. Before this paradigm is implemented in real environments, the trustworthiness of the system must be assured.

The “doctor in the loop” is a new paradigm in information driven medicine, picturing the doctor as authority inside a loop with an expert system in order to support the (automated) decision making with expert knowledge. This information not only includes support in pattern finding and supplying external

knowledge, but the inclusion of data on actual patients, as well as treatment results and possible additional (side-) effects that relate to previous decisions of this semi-automated system.

The concept of the “doctor in the loop” is basically an extension of the increas-

ingly frequent use of knowledge discovery for the enhancement of medical treatments together with the “human in the loop” concept (see [1], for instance): The expert knowledge of the doctor is incorporated into “intelligent” systems (e.g., using interactive machine learning) and enriched with additional