Number 104 January 2016

ERCIM NEWS

www.ercim.eu

Special theme:

Tackling Big Data in the Life Sciences

Also in this issue:

Research and Society:

"Women in ICT Research and Education"

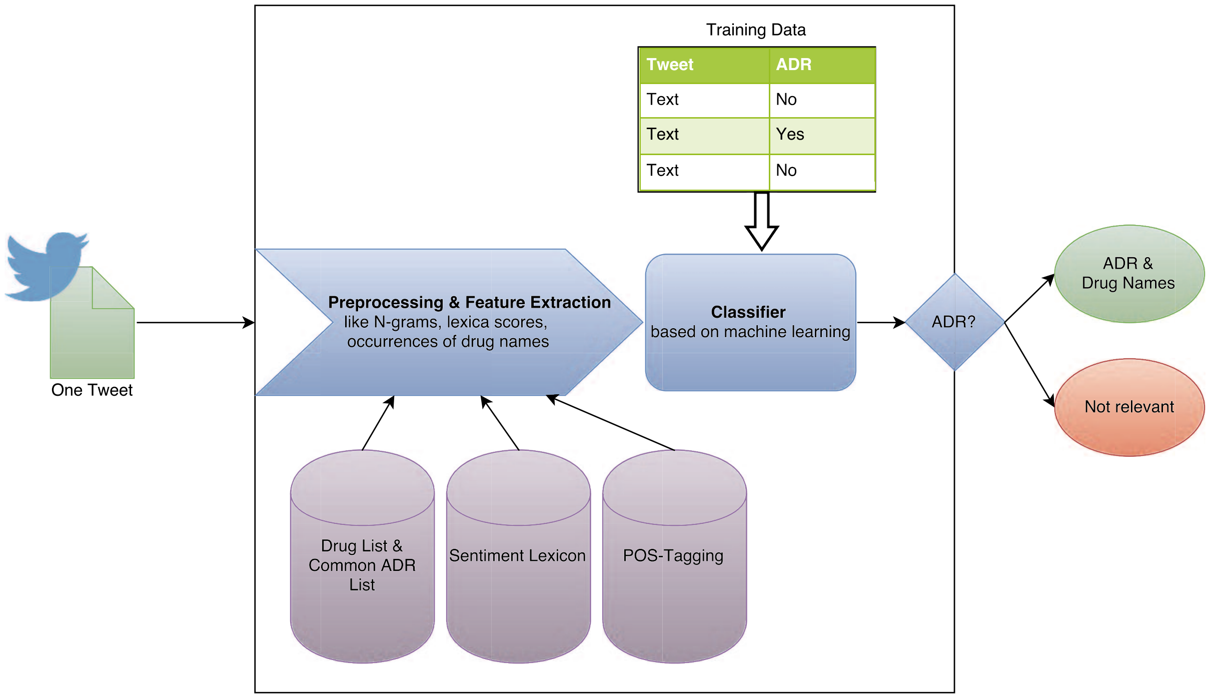
Training Data
Tweet
ADR
Text
No
Text
Yes
Text
No
One Tweet
Preprocessing & Feature Extraction
like N-grams, lexica scores, occurrences of drug names
Classifier
based on machine learning
ADR?
ADR & Drug Names
Not relevant
Drug List & Common ADR List
Sentiment Lexicon
POS-Tagging

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 increasingly 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 information and expert know-how. Using machine learning algorithms, medical knowledge and optimal treatments are identified. This knowledge is then fed back to the doctor to assist him/her (see Figure 1).

Manipulation Security and Trust

The implementation of the doctor in the loop concept has met several challenges – both of a technical nature and in other areas. One challenge is gaining the acceptance of such systems by doctors themselves, who are often not researchers, but medical practitioners.

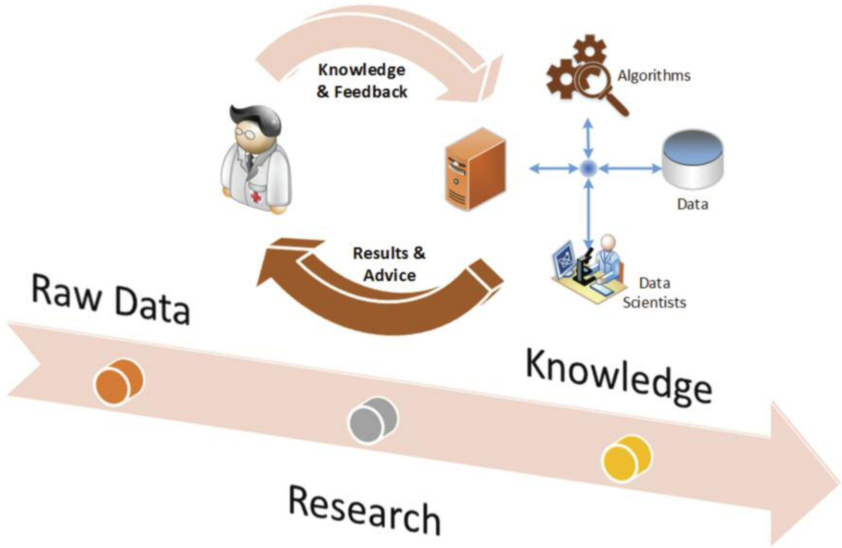

Figure 1: The Doctor in the Loop.

While privacy and security in biomedical data applications have been discussed extensively in recent years (see [2] for an overview), the topic of trust has been neglected. Nevertheless, it is very important that the trustworthiness of the systems can be guaranteed in order to make the abstract concept of a doctor in the loop practicable.

In this medical context, it is absolutely crucial to ensure that the information sent by the doctor to the machine learning algorithms cannot be manipulated following its submission. In order to guarantee this, a new approach for protecting the underlying data based on a hash chain has been proposed explicitly for doctor in the loop systems (see [3]). This approach takes advantage of the fact that large amounts of data are typically stored in databases. While these are often identified with the abstract data model, they are in reality complex software systems providing a multitude of internal mechanisms that have various functions, such as enhancing performance. This approach utilizes the transaction mechanism for providing transaction safety (i.e., if a crash occurs, the database is brought back into a well-defined state) in order to protect the information sent to the system by the doctor against manipulation, even if the manipulation is carried out by the administrator of the database.

Legal Issues

Legal issues are often overlooked by purely technical solutions, but are vital for any doctor that is actually participating in such an approach as expert. Important questions that need to be addressed include: What information can be shared between the doctor and the expert system, and what levels of data and privacy protection need to be applied? And who is responsible for the correctness of the results derived from the combination of human knowledge and machine learning algorithms? Although this is an important field, no guidelines are currently available. The issue is further complicated by the differences between national legislations even between member states of the European Union. Defining workflows that clinical doctors can reliably apply without the fear of prosecution lies thus in the focus of the planned RDA (Research Data Alliance) Workgroup “Security and Trust” that held its “Birds of Feather” session on the 6th RDA Plenary on September 24th in Paris [L1]. One of the major goals of this workgroup is to draw together a set of best practices and guidelines in the area of data driven medical research.

Special focus in CBmed

CBmed [L2] is an interdisciplinary research centre, founded in 2014, with the aim of providing biomarker research, mainly in the areas cancer, metabolism and inflammation. One of the core features of this centre is the tight incorporation of ICT as a horizontal research area that provides data and techniques to all other projects. This does not only apply to applications associated with the doctor in the loop, but also other areas, such as data and privacy protection, the development of new data mining techniques and tools for the efficient analysis of large amount of “–omics” data.

For the CBmed consortium, the model of the doctor in the loop offers abundant possibilities, especially in the area of data driven research. Considering the recent surge in big data related research and developed tools, this approach is expected to be one of the major drivers in medical research in the years to come.

Links:

[L1]: https://rd-alliance.org/rda-working-group-data-security-and-trust-wgdst-p6-bof-session.html

[L2]: http://www.cbmed.org/en/

References:

[1] W. S. Levine, ed.: “The control handbook”, CRC press, 1996.

[2] P. Kieseberg, H. Hobel, S. Schrittwieser, E. Weippl, A. Holzinger, “Protecting anonymity in data-driven biomedical science”, in “Interactive Knowledge Discovery and Data Mining in Biomedical Informatics” (pp. 301-316), Springer Berlin Heidelberg, 2014.

[3] P. Kieseberg, J. Schantl, P.Fruehwirt and E. R. Weippl, A. Holzinger: “Witnesses for the Doctor in the Loop," in 2015 International Conference on Brain Informatics & Health (BIH), 2015.

Please contact:

Peter Kieseberg
SBA Research, Vienna, Austria
E-mail: pkieseberg@sba-research.org