Digital Personalized Health and Medicine
L.B. Pape-Haugaard et al. (Eds.)

doi:10.3233/SHTI200375

Classification of the Severity of Adverse Drugs Reactions

Raphaël CHAUVET[1,2], Cédric BOUSQUET[1], Agnès LILLO-LELOUET[3], Ilan ZANA[1], Ilan BEN KIMOUN[1], Marie-Christine JAULENT[1]

[1] *Sorbonne Université, INSERM, Université Paris 13, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé, Paris, France.*
2 EISTI, International Engineering School, 95000 Cergy-Pontoise, France
[3] *Centre Régional de Pharmacovigilance, Hôpital Européen Georges-Pompidou, Assistance Publique – Hôpitaux de Paris, Paris, France.*

Abstract. This poster presents a non-exhaustive study of machine learning classification algorithms on pharmacovigilance data. In this study, we have taken into account the patient's clinical data such as medical history, medications taken and their indications for prescriptions, and the observed side effects. From these elements we determine whether the patient case is considered serious or not. We show the performances of the different algorithms by their precision, recall and accuracy as well as their learning curves.

Keywords: Classification, Machine Learning, Pharmacovigilance, Decision support tool.

1. Introduction

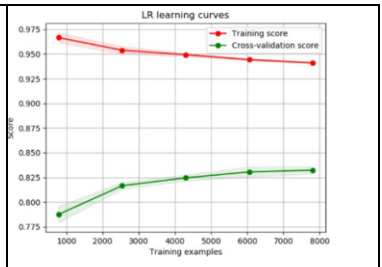
Since 2017 in France, citizens can report directly to health authorities any adverse event related to health products. The objective of the presented work is to study to which extend machine learning algorithms could predict the seriousness of the situation from the description of the situation by patients. We used structured data recorded by health professionals from the verbatim of the patient to train machine learning algorithms. The main perspective is to apply the resulting prediction system to social media for pharmacovigilance monitoring. Although, signal detection in pharmacovigilance is historically based on statistical methods, more and more machine learning can make predictions for the detection of side effects [1]. Andrew M Wilson in [2] concluded that machine learning will not replace traditional pharmacovigilance technics but could collaborate to detect uncommon drug-related effects or reducing adverse drugs effects identification time.

2. Methods and Results

To carry out the study we considered a database that includes all the cases reported by patients from 2010 to 2019 resulting in a base of 13026 patient cases described by 50 variables from HEGP (European Hospital Georges Pompidou), Paris, France. Cases reported in this database are standardized by pharmacovigilant using MedDRA terms. 8699 cases are annotated as serious by health professionals. The aim of the study is to search the most efficient supervised learning algorithm that will enable the pharmacovigilant to prioritize serious cases. In a first step, all values of a variable are considered as a bag of word and translated in terms of presence/absence ("one-hot-encoding"). This operation allows to keep the inter-variable interaction for poly-medical or poly-pathological patients.

Before applying an algorithm, the dataset is divided in two, 75% of the database will serve as a training base and 25% will be used to test the learning performance. This process is repeated 5 times in order to have a general tendency of the performances of the different methods. Then we apply 9 different state-of-art clustering models to our dataset thanks to available libraries: *Decision Tree, Bagging Tree, Boosting Tree (AdaBoost), Gradient Boosting Machine, K-Nearest Neighbours (K-NN), Logistic Regression, Neural Network, Random Forest, Support Vector Machine (SVM).* The next figure left shows the results in terms of precision (proportion of good prediction if the case is severe) and recall (percentage of serious cases correctly identified). The model with the best performance in terms of accuracy and recall is the logistic regression (next figure right). It does not over-fit the data and is very accurate (cf. LR learning curves next figure right).

Model	Recall	Precision
Decision tree	0.73 +/- 0.15	0.73 +/- 0.11
Bagging	0.75 +/- 0.17	0.77 +/- 0.11
AdaBoost	0.68 +/- 0.22	0.77 +/- 0.15
Gradient Boosting	0.67 +/- 0.25	0.79 +/- 0.16
K-NN	0.74 +/- 0.13	0.72 +/- 0.10
Logistic Regression	0.77 +/- 0.19	0.81 +/- 0.11
Neural Network	0.75 +/- 0.13	0.75 +/- 0.09
Random Forest	0.76 +/- 0.18	0.79 +/- 0.11
SVM	0.50 +/- 0.00	0.34 +/- 0.00



3. Discussions and conclusion

It is important to note that the performance of machine learning methods depends on the input data that is used for learning. The same method on a different data set will give different results. Similarly, as new several observations are saved to the database, the performance of the same method will change over time. During this first study, it was decided to work initially with the default values of the algorithms because we did not have a priori knowledge to force the value of certain parameters. One possible improvement will be to choose optimal settings for all models and re-compare their performances later. For that we could use a search grid where all the parameters would be stored and build the models that would test all the possible combinations to determine the best of all.

References

[1] Voss EA, Boyce RD, Ryan PB, van der Lei J, Rijnbeek PR, Schuemie MJ. Accuracy of an automated knowledge base for identifying drug adverse reactions. J Biomed Inform. 2017 Feb;66:72-81. doi: 10.1016/j.jbi.2016.12.005. Epub 2016 Dec 16.

[2] Andrew M Wilson, Lehana Thabane and Anne Holbrook. Application of data mining techniques in pharmacovigilance. Br J Clin Pharmacol. 2004 Feb; 57(2): 127–134.