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

doi:10.3233/SHTI200120

Automatic Classification of Discharge Letters to Detect Adverse Drug Reactions

Vasiliki FOUFI[a,1], Kuntheavy ING LORENZINI[b], Jean-Philippe GOLDMAN[a], Christophe GAUDET-BLAVIGNAC[a], Christian LOVIS[a] and Caroline SAMER[b]
[a]*Division of Medical Information Sciences, Geneva University Hospitals & University of Geneva, Switzerland*
[b]*Clinical Pharmacology and Toxicology, Geneva University Hospitals, Switzerland*

Abstract. Adverse drug reactions (ADRs) are frequent and associated to significant morbidity, mortality and costs. Therefore, their early detection in the hospital context is vital. Automatic tools could be developed taking into account structured and textual data. In this paper, we present the methodology followed for the manual annotation and automatic classification of discharge letters from a tertiary hospital. The results show that ADRs and causal drugs are explicitly mentioned in the discharge letters and that machine learning algorithms are efficient for the automatic detection of documents containing mentions of ADRs.

Keywords. Adverse drug reaction, pharmacovigilance, text mining, document classification, supervised machine learning

1. Introduction

Adverse drug reactions (ADRs) affect 7 to 17% of hospitalized patients [1,2] and can result in serious morbidity, mortality and high costs. They are largely underreported, making active pharmacovigilance useful. The detection of ADRs can be performed through the review of electronic medical records (EMR) [3] or regular ward visits by a trained health professional [4], which can be time-consuming. In this context, data mining techniques, focusing on the automated identification of ADRs from the patient EMR can be helpful [5,6]. These techniques include structured data analysis as well as text mining, including natural language processing (NLP) [7,8]. In this context, techniques for the automatic classification of clinical documents have been proven effective [9–11].

The aim of this study is to assess the feasibility of using NLP techniques to detect the presence or absence of ADRs in discharge letters written in French and extracted from patients hospitalized in a tertiary hospital via a hybrid –machine learning and rule-based– method. In this paper, we will present the supervised learning method. Particularly, three machine learning algorithms for document classification have been applied and evaluated. For the creation of the training and test datasets, manual processing of 300 discharge letters was performed. The results show that ADRs are reported in the documents and that NLP tools are efficient for their automatic detection.

[1] Corresponding Author, Division of Medical Information Sciences, Geneva University Hospitals & University of Geneva, Switzerland; E-mail: vasiliki.foufi@unige.ch.

2. Method

2.1. Data collection

Our study was approved by the local ethics committee (study number: 2016-02107). Hospitalized adults for whom a specialized consultation from clinical pharmacologists (in 2015 and 2016) had identified the occurrence of serious ADRs, during or leading to hospitalisation, were included in the study. Out-patients and cases of non-serious ADRs were not included. Based on these criteria, a dataset of 100 positive discharge letters (presence of ADR) and 200 negative letters (absence of ADR) was constituted.

2.2. Data processing

2.2.1. Manual annotation

For the creation of the training and test datasets, an expert manually annotated 100 discharge letters (positive dataset) and validated 200 letters (negative dataset). Based on specific guidelines, sequences of the following categories were annotated:

1. Drugs

The drugs category is divided in 3 sub-categories: a) commercial names, b) international nonproprietary names (INN), c) therapeutic class.

2. ADRs

Occurrences of ADRs and their consequences, symptoms, laboratory values are annotated. ADRs are divided in 3 sub-categories: a) names (hépatite/*hepatitis*), b) periphrases (perturbation des tests hépatiques/*liver test abnormalities*), c) characteristics (hémoglobine à 75 g/l/*haemoglobin at 75 g/l*).

3. Trigger words

Words like imputabilité/*causality*, stoppé/*stopped*, suspect/*suspect* that imply the presence of an ADR are annotated.

4. Drug indications

Indications for drugs entailed in ADRs are annotated.

2.2.2. Automatic classification of discharge letters

For the automatic classification of the discharge letters into positive or negative, a supervised learning approach was followed. The dataset is composed of the positive letters containing at least one annotation of ADR and the negative letters validated by the expert. For this task, three machine learning algorithms widely used for text classification tasks were applied: Support Vector Machine (SVM), Naïve Bayes Classifier, and Linear Classifier. From the whole dataset, 80% was used for training and 20% for testing.

3. Results

3.1. Manual annotation

Out of 100 letters of the positive dataset, 87 letters contained at least one annotated sequence. The mean length of a discharge letter is 785 words. In total, 1471 sequences were annotated. These results are summarized in Table 1:

Table 1. Number of occurrences per annotation category.

Annotation category	Number of occurrences	Unique occurrences
Commercial name	170	76
International nonproprietary name	210	87
Therapeutic class	126	59
Trigger word	293	156
ADR	441	217
Characteristic	130	103
Drug indication	37	28
Periphrasis	64	35

From the 200 discharge letters considered as negative (absence of ADR), 47 reported an ADR and 2 were empty. Therefore, the final negative dataset consists of 151 letters.

3.2. Automatic document classification

Three well-known classification methods implemented in the Scikit-learn python library [12] were applied and compared: Support Vector Machine, Naive-Bayes, and Linear Classifier. For the SVM model, the radial basis function was selected with a 'scale' gamma. The Multinomial Naive-Bayes was trained with default parameters. Eventually, the Linear Classifier is based on a stochastic gradient descent with a lower stopping criterion than default (tol=1e-6). Figure 1 represents the classification accuracy of each model after 50 iterations. For each iteration, the test set represents 20% of the whole corpus, i.e. 17 positive documents and 30 negative documents (k-fold = 0.8). The training time is ~100ms for each SVM iteration whereas it takes 2mn for both the Naïve Bayes and the Linear Classifier.

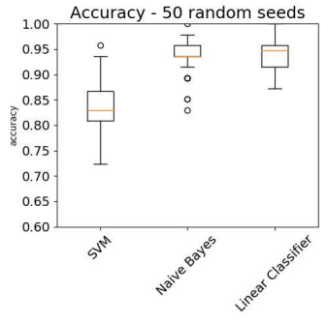

Figure 1. Mean accuracy over 50 iterations.

The SVM classifier achieved 0.83 accuracy, Naïve Bayes achieved 0.94 accuracy and the Linear Classifier 0.94. Complete results are shown in Table 2:

Table 2. Evaluation results of the automatic classification.

Classification method	Class	Precision	Recall	F1 score	Test set
SVM	Positive	0.80	0.97	0.88	17
	Negative	0.92	0.60	0.73	30
Naïves Bayes	Positive	0.93	0.97	0.95	17
	Negative	0.95	0.88	0.91	30
Linear Classifier	Positive	0.97	0.93	0.95	17
	Negative	0.90	0.95	0.92	30

The confusion matrices in Figures 2-4 display the performance of each classifier at the classification task:

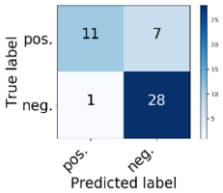

Figure 2. SVM.

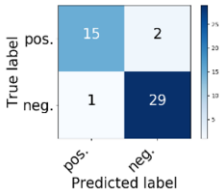

Figure 3. Naïve Bayes.

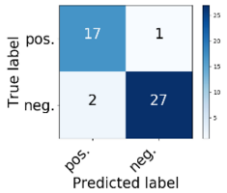

Figure 4. Linear Classifier.

4. Discussion

The manual annotation of positive discharge letters by an expert showed that ADRs were explicitly mentioned in most cases (>80%). A significant bias is that the test population were patients who had already received a specialized pharmacology consultation that had identified the ADR, thereafter mentioned in the discharge letter. Drugs were almost equally mentioned as commercial names, INN and therapeutic classes, and this has to be taken into account for their automated detection, especially given the diversity of commercial names in different countries. Trigger words were frequently present (293 occurrences); therefore, they constitute useful tools for the automatic detection of ADRs. ADRs were most frequently mentioned as plain terms, such as MedDRA (Medical Dictionary for Regulatory Activities) derived terms (i.e. *hepatitis*), but were also described as periphrases or as laboratory characteristics in many cases (approximately 200 occurrences) which makes their automatic detection challenging given the fact that the distinction between the 3 sub-categories was not always straightforward even for the human annotator.

For the automatic classification task into positive and negative discharge letters, three machine learning algorithms were applied and evaluated on the dataset. Naïve Bayes and Linear Classifier achieved the same mean accuracy over 50 iterations (0.94) and high precision and recall (Table 2).

A major limitation of this study is that the dataset was manually processed by only one annotator. Also, the classifiers should be applied and evaluated on a larger dataset.

5. Conclusion

In this study, we presented the methodology used for the manual and automatic processing of discharge letters generated in a tertiary hospital in the aim to automatically detect the presence or absence of ADRs. A dataset of 300 discharge letters written in French was manually processed and the output was used to train and test three machine learning algorithms for document classification. After comparison, we concluded that Naïve Bayes and Linear Classifier performed better than SVM at this task.

The manual annotation of the dataset from another expert will serve to create a gold standard corpus. In a next step, the trigger words and sequences describing the presence of ADRs identified during the manual annotation will be included in the rules that are being developed for the automatic identification and extraction of ADRs and their relations with causal drugs. Then, the hybrid method –machine learning and rule-based– will be applied and evaluated on the gold standard dataset.

Acknowledgements

This research is funded by the HUG's Private Foundation and the HUG's Medical Director PRD (Projets Recherche & Développement) project funding.

References

[1] Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. - PubMed - NCBI, (n.d.). https://www.ncbi.nlm.nih.gov/pubmed/9555760 (accessed October 15, 2019).
[2] Frequency of adverse drug reactions in hospitalized patients: a systematic review and meta-analysis. - PubMed - NCBI, (n.d.). https://www.ncbi.nlm.nih.gov/pubmed/22761169 (accessed October 15, 2019).
[3] Identification of Adverse Drug Events from Free Text Electronic Patient Records and Information in a Large Mental Health Case Register, (n.d.). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4537312/ (accessed October 15, 2019).
[4] Methods and systems to detect adverse drug reactions in hospitals. - PubMed - NCBI, (n.d.). https://www.ncbi.nlm.nih.gov/pubmed/11735652 (accessed October 15, 2019).
[5] Data mining on electronic health record databases for signal detection in pharmacovigilance: which events to monitor? - PubMed - NCBI, (n.d.). https://www.ncbi.nlm.nih.gov/pubmed/19757412 (accessed October 15, 2019).
[6] Data-mining-based detection of adverse drug events. - PubMed - NCBI, (n.d.). https://www.ncbi.nlm.nih.gov/pubmed/19745372 (accessed October 15, 2019).
[7] Using text-mining techniques in electronic patient records to identify ADRs from medicine use. - PubMed - NCBI, (n.d.). https://www.ncbi.nlm.nih.gov/pubmed/22122057 (accessed October 15, 2019).
[8] Automated detection of adverse events using natural language processing of discharge summaries. - PubMed - NCBI, (n.d.). https://www.ncbi.nlm.nih.gov/pubmed/15802475 (accessed October 15, 2019).
[9] Clinical Document Classification Using Labeled and Unlabeled Data Across Hospitals. - PubMed - NCBI, (n.d.). https://www.ncbi.nlm.nih.gov/pubmed/30815095 (accessed October 15, 2019).
[10] Machine learning in automated text categorization, (n.d.). https://dl.acm.org/citation.cfm?id=505283 (accessed October 15, 2019).
[11] K.J. Dreyer, M.K. Kalra, M.M. Maher, A.M. Hurier, B.A. Asfaw, T. Schultz, E.F. Halpern, and J.H. Thrall, Application of recently developed computer algorithm for automatic classification of unstructured radiology reports: validation study, *Radiology*. **234** (2005) 323–329. doi:10.1148/radiol.2341040049.
[12] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research*. **12** (2011) 2825–2830.