

Initial Experiments for Pharmacovigilance Analysis in Social Media Using Summaries of Product Characteristics

Leonardo Campillos-Llanos^a, Cyril Grouin^a, Agnès Lillo-Le Louët^b, Pierre Zweigenbaum^a

^a LIMS, CNRS, Université Paris-Saclay, F-91405 Orsay, France

^b CRPV Paris-HEGP, Hôpital Européen George Pompidou, 20 rue Leblanc, 75015 Paris, France

Abstract

We report initial experiments for analyzing social media through an NLP annotation tool on web posts about medications of current interests (baclofen, levothyroxine and vaccines) and summaries of product characteristics (SPCs). We conducted supervised experiments on a subset of messages annotated by experts according to positive or negative misuse; results ranged from 0.62 to 0.91 of F-score. We also annotated both SPCs and another set of posts to compare MedDRA annotations in each source. A pharmacovigilance expert checked the output and confirmed that entities not found in SPCs might express drug misuse or unknown ADRs.

Keywords:

Natural Language Processing; Pharmacovigilance; Social media

Introduction

According to the World Health Organization (WHO), pharmacovigilance is "the science and activities relating to the detection, assessment, understanding and preventing of adverse effects or any other drug-related problem".¹ The drug development process has several steps, from discovery in a laboratory, to preclinical research and clinical development involving patients. Nonetheless, after approval, complete and definitive information about drug safety is not available. Moreover, the drug use may change, the benefit/risk may evolve and health authorities need any information available.

Self-reports of adverse drug reactions (ADRs) are scarce: the French National Agency for Drug Safety (ANSM) evaluated that only 5% are reported by patients. Because ADRs are known late, except in case of highly publicized events (e.g., H1N1 flu), social media is used to improve pharmacovigilance efficiency [1,2]. However, web-based data generally contain colloquial jargon that is hard to process with common Natural Language Processing (NLP) tools, calling for dedicated approaches [3]. Pharmacovigilance also addresses abuse and drugs misuse, which involves "situations where a medicinal product is intentionally and inappropriately used not in accordance with the terms of the marketing authorization" [4].

We present an ongoing work on pharmacovigilance analyses in health fora written in French, with a special focus on drug misuse. We present the experiments we made based on a comparison of the Medical Dictionary for Regulatory Activities (MedDRA) codes annotated in web messages and the codes identified in Summaries of Product Characteristics (SPCs). As far as we know, this source has not been commonly used for NLP in health social media; this is our main contribution. We annotated pathological conditions in social media and SPCs using an NLP pipeline designed to identify such entities [5]. This tool was improved for normalizing annotations based on codes from MedDRA [6] and the Anatomical Therapeutic Classification (ATC) [7]. These methods identify potential drug misuse and unknown ADRs.

Background

Social media are useful to identify both adverse effects and drug misuse [1,2,8]. Different types of drug misuse may occur; Bigeard and colleagues report a comprehensive typology [9]. Misuse may be related to whether the drug is prescribed by a practitioner, or taken as self-medication. In case of a prescription drug, misuse may involve interaction with other drugs, an incorrect frequency, duration or dose of medication intake. Other situations involve not respecting the drug intake (e.g., if levothyroxine is not taken on an empty stomach), or problems of conservation (e.g., to keep an eye drop solution open for more than 2 weeks). Indication misuse involves the intake of a medication for an unrelated pathology (e.g., using baclofen for alcohol addiction). Another situation of interest is drug abuse, which occurs when users take drugs without any pathology but searching for a specific effect (e.g., a psychotropic effect, as is the case with the *purple drank* cocktail based on codeine cough syrup with soda).

Detection of drug misuse in social media is hard to identify automatically, since knowledge on a given medical drug indications or posology are needed to infer *unexpected* user's intake behaviors. The few works undertaken to detect drug misuse have applied supervised methods and annotated data by pharmacovigilance experts [10].

In the absence of enough available annotated data, we resorted to methods for generating training data over unlabeled samples. Distant supervision approaches [11] are close to that paradigm and have been applied in the medical domain [12]. Our method, however, relies on the annotation of knowledge

¹ <https://bit.ly/2FtlYX5> [Accessed March 2019]

sources (SPCs in our context of pharmacovigilance) and then compares the output to annotations in web fora.

Methods

We conducted two types of experiments. First, we conducted a *supervised experiment* with a small set of messages annotated by pharmacovigilance experts regarding positive or negative misuse. Second, we tested to which extend comparing MedDRA codes found in messages from web fora and in SPCs provide cues of drug misuse and unknown side effects.

Corpus for Supervised Experiments

Corpus 1 contains posts written in French by consumers, extracted from several public fora from the health domain: AlloDocteurs,² Atoute,³ Baclofène,⁴ Doctissimo Médicaments,⁵ eSanté,⁶ Journal des Femmes,⁷ and Vulgaris.⁸ In these posts, different drugs and pharmacological substances are discussed:

Agomelatine (Valdoxan®): an antidepressant, which may be misused for insomnia, panic and anxiety attacks.

Baclofen: a muscle relaxant, used to treat addictions (especially alcoholism) out of any marketing authorization.

Duloxetine (Cymbalta®): an antidepressant drug.

Exenatide (Bietta®): an antidiabetic injection; a dose misuse (i.e., users took more injections than what is prescribed) has been detected by pharmacovigilance experts.

Myolastan (Tetrazepam®): a muscle relaxant; both misuse (indication and posology) and drug abuse has been reported.

Messages were collected with a framework for extracting data from web fora [13] and then processed using an NLP pipeline to annotate pathological and medical drug entities. Pharmacovigilance experts—pharmacists, or physicians qualified in Pharmacology—revised a selection of 1178 posts and manually annotated them as expressing misuse or not.

Corpus for Comparing Internet Posts and SPCs

Corpus 2 gathers posts from similar kind of consumers' web. We chose three specific and actual drug topics (regarding the media coverage and patients' interest), each in different fora:

1. Baclofen: data come from the Atoute website and were used in semi-automatic methods to detect drug misuse.
2. Levothyroxine: this drug replaces or provides more **thyroid hormone**, and a new formulation was marketed in August 2017. Messages come from the Vivre sans thyroïde forum.⁹
3. Vaccines: in France, only 3 vaccines (diphtheria, tetanus, poliomyelitis) were mandatory until January 1st, 2018. Since that date, 11 vaccines are

compulsory: tetanus, diphtheria, poliomyelitis, and 8 additional ones: pertussis, polio, measles, mumps, rubella, hepatitis B, haemophilus influenza bacteria, pneumococcus, and meningococcus C. We used posts from the Doctissimo website.¹⁰

For each topic, we extracted 100 messages according to two inclusion criteria: presence of drug name and pathology names, and a limit of words (we avoided long generic discussions). In case of lack of messages with drug names and pathology entities, we extracted new messages to gather up to 100 messages per topic. We also used SPCs available from the French authorities¹¹ for baclofen, levothyroxine, and the twenty marketed products to perform the newborns immunization for the eleven vaccines: ACT-HIB®, Boostrix Tetra®, Engerix B®, Fendrix®, HBVaxPro®, Hexyon®, Imovax Polio®, Infanrix Tetra®, Infanrix Quinta®, Infanrix Hexa®, M-M-RVaxPro®, Menjugate®, Neisvac®, Pentavac®, Pneumovax®, Prevenar®, Priorix®, Repevax®, Revaxis®, and Tetravac®. We focused on sections describing indications, counter-indications or adverse drug reactions.

Annotation of Messages with an NLP Pipeline

We applied an NLP pipeline on both corpus: 1178 posts annotated by pharmacovigilance experts, and 300 concerning levothyroxine, baclofen and vaccines. As previously explained [5], the pipeline has modules for normalization, tokenization, Part-of-Speech tagging and concept annotation based on machine learning, namely Conditional Random Fields (CRF) [14]. Because the tool was improved since the first evaluation made on the first set of web fora [5], we evaluated the annotations of the 300 posts. We manually checked those annotations to build a gold standard, using BRAT (Figure 1) [15]. We evaluated the annotations through pre-precision, recall and F1-score (F1) metrics using BRATEval [16].

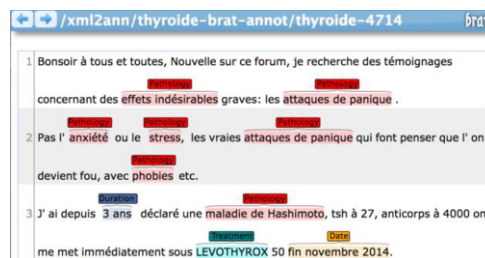


Figure 1 - Sample of forum message annotated with BRAT

We also applied the annotation tool on the summary of product characteristics (SPCs) detailed before. Thus, pathological entities in indications covered by those selected drugs are annotated and labeled with MedDRA codes. We also revised the annotations of sections Indications and Side effects in SPCs, using the same methodology to revise annotated posts.

Experiments in a Supervised Context

We followed the procedures applied by Bigeard and colleagues [10]. We tested Naïve Bayes (NB) and

² www.allodocteurs.fr [Accessed March 2019]

³ www.atoute.org [Accessed March 2019]

⁴ www.baclofene.com [Accessed March 2019]

⁵ <http://medicament.doctissimo.fr> [Accessed March 2019]

⁶ <https://www.e-sante.fr> [Accessed March 2019]

⁷ www.journaldesfemmes.fr [Accessed March 2019]

⁸ www.vulgaris-medical.com [Accessed March 2019]

⁹ www.forum-thyroide.net [Accessed March 2019]

¹⁰ www.doctissimo.com [Accessed March 2019]

¹¹ <https://bit.ly/2UNLuN6> [Accessed March 2019]

Multinomial Naïve Bayes (MBN) algorithms on the 1178 posts annotated as expressing misuse or not by pharmacovigilance experts. We tested different features:

- Word tokens in message
- Word roots in message
- Anatomical Therapeutic Classification (ATC) codes of medical drugs in each message
- MedDRA codes of pathologic entities in message
- 3-grams
- 3-character-grams

We used ScikitLearn Software for this set of experiments [17]. In all contexts, we applied 10-fold cross-validation and used an 80% and 20% ratio of training and test sets, respectively.

We also tested different subsets of messages with regard to the number of messages annotated as misuse. In real-life contexts, most posts will not contain any misuse behavior or ADR. The datasets will suffer from *class imbalance* (i.e., most samples will not express misuse nor ADRs), a common problem with supervised machine-learning algorithms [18]. In a context where most posts bear a negative class, a random classifier or a classifier labeling all samples as negative will certainly have good accuracy, even though it does not make use of any linguistic or knowledge-based feature. We thus tested different ratios of messages annotated as positive or negative misuse:

- The full corpus of messages (1178 posts),
111 messages annotated as positive misuse (~10:1 ratio)
- A subset of 336 messages: 111 classified as positive misuse and 225 as negative misuse (~2:1 ratio)
- A subset of 246 messages: 111 classified as positive misuse and 135 as negative misuse (~1:1 ratio)

Comparing Messages in Social Media and SPCs

Once we applied the NLP tool to annotate the pathological entities in the SPCs, we assumed these coded pathologies set up the list of correct uses to be found in messages over the Internet. Conversely, all pathologies related to one of those drugs found in a message (missing in that list of *expected* pathologies) may be a drug misuse or unexpected ADR. Because we used the same annotation schema and tool, we could compare MedDRA codes in both sets and extracted a list of candidate terms. Finally, a pharmacovigilance expert and coauthor of this work (ALL)—a physician, qualified in Pharmacology, with 20 years of expertise—checked the selected entities to confirm misuse behavior or unknown ADRs.

Normalization

We used MedDRA for coding terms of pathological entities. Following Bousquet *et al.* [19], we coded Lower-level terms (LLT) for expressions of pathologies (*verbatim terms*), but also mapped these LLTs to preferred terms (PT) for pharmacovigilance analyses. We used the UMLS® [20] Concept Unique Identifiers (CUIs) to map term variants referring to the same concept. Normalization rules were applied considering inflection (singular/plural, diacritics, syntactic variants of multiwords); and Levenshtein distances were used to get the candidate term with closer string distance from a term variant.

Results

Evaluation of the NLP Pipeline

We annotated a total of 2249 pathologies and medications in forums, and 6772 in the SPCs (Table 1; we report the count of both annotations and types, i.e., different annotated items). Table 2 shows the evaluation results of each subset of posts.

Table 1 - Number of annotated entities (total items and types)

	# words	# annotations (types)	
		Pathologies	Medical drugs
Levothyroxine	18274	511 (230)	314 (67)
Baclofen	12941	259 (133)	386 (52)
Vaccines	10336	390 (185)	389 (83)
Total posts	41551	1160 (480)	1089 (192)
[avg per post]	[138.5]	[3.87 (1.6)]	[3.63 (0.6)]
Total in SPCs	88919	3687 (815)	3085 (320)
[avg per SPC]	[4041.8]	[167.6 (37.1)]	[140.2 (14.5)]

Table 2 - Evaluation of the NLP annotation pipeline (P: Precision; R: Recall; F1: F-score; Avg: average)

	Pathologies			Medical drugs		
	P	R	F1	P	R	F1
Levothyroxine	0.80	0.68	0.74	0.98	0.85	0.91
Baclofen	0.84	0.70	0.76	0.89	0.80	0.84
Vaccines	0.84	0.72	0.77	0.82	0.78	0.80
Avg (posts)	0.83	0.70	0.76	0.93	0.83	0.88
SPCs	0.92	0.90	0.91	0.76	0.90	0.82

In web posts, higher F1 scores were obtained when annotating medications rather than pathological entities. This is mainly due to the higher number of different pathological entities (480) and also to the difficulty in annotating expressions of pathological conditions in *patient language* (e.g., *crevé*, ‘worn out’ stands for *fatigue*). Annotation results of messages concerning the baclofen and levothyroxine show higher F1 scores; this might be due to the fact that messages discussing newborns vaccination contain more different drug names (83). Results of pathological entities in posts regarding the levothyroxine might also be due to a higher number of different pathological entities in this forum (230). Because SPCs feature a lower degree of *patient language*, annotation of pathologies achieved a higher F1 score than in web fora.

Results of Supervised Experiments

As expected, the best results were obtained on the full corpus, either using Naïve Bayes or Multinomial Naïve Bayes (Table 3; we only report the results of the best features on the test set). The experiments on the other corpora configurations helped us better understand the features that *really* helped the classifier to learn and distinguish positive and negative

misuse. The best results were mainly obtained with these features: ATC codes, MedDRA codes and word roots. We had similar results as those reported by Bigeard *et al.* [10].

Table 3 - Results of classifiers in supervised context: Naïve Bayes (NB, above) and Multinomial Naïve Bayes (MNB, below). The label ratio is the proportion of posts annotated as positive or negative misuse. 2:1 stands for 2 negative posts per 1 positive; P: Precision; R: Recall; F1: F-score

	Label ratio	Features	P	R	F1
NB	~ 10:1 (all data)	ATC codes	0.94	0.93	0.91
		+ MedDRA codes			
		+ word roots			
	~ 2:1	+ 3-grams	0.63	0.67	0.62
		ATC codes + MedDRA codes + word roots			
MNB	~ 1:1	Tokens + 3-grams + ATC codes; or Tokens + 3-grams	0.88	0.88	0.88
	~ 10:1 (all data)	ATC codes	0.93	0.92	0.89
		+ MedDRA codes			
		+ word roots			
	~ 2:1	ATC codes + MedDRA codes + word roots	0.71	0.70	0.62
		Tokens + 3-grams			
	~ 1:1	+ ATC codes ; or Tokens + 3-grams	0.82	0.82	0.82

Results of comparing messages in social media and SPCs

We extracted from posts 301 pathological entities that were not documented in SPCs. Most were related to Levothyroxine (166) and Baclofen (103). Pathologic conditions related to vaccines (32) concerned Boostrix®, Engerix®, Infanrix®, Neisvac®, Pentavac®, Prevenar®, Priorix®, Repevax® and Revaxis®. We observed that some vaccines did not appear in the selected posts (ACT-HIB®, Fendrix®, HBVaxPro®, Hexyon®, Imovax Polio®, Infanrix Tetra®, M-M-RVaxPro®, Pneumovax® and Tetravac®). The pharmacovigilance expert considered that only 3 cases might be misuse related to Baclofen (1.3% of candidate items). However, one case is ambiguous: we cannot state if, when the user mentioned the unexpected pathology, he/she meant to link it to an indication related to the intake of Baclofen. No other misuse cue was confirmed with regard to other drugs. The expert identified 68 undocumented ADRs (22.6% of selected items; 28 need more context to be confirmed). Unknown ADRs mostly involved levothyroxine (52 cases), baclofen (6), Engerix® (2), Infanrix® (2), Pentavac® (3), Priorix® (1) and Repevax® (2).

Qualitative Evaluation

We analyzed messages to detect linguistic cues expressing drug indication or misuse (Table 4). Misuse due to incorrect

dose might be expressed with specific verbs. However, in the Corpus 1, we observed that experts did not always annotate as misuse some contexts with those cues. Exact validation of drug doses reported by web users might indeed be within the range of correct doses that could be a *user-perceived misuse*. We estimate a linguistic analysis needs to be complemented by knowledge-driven approaches; e.g., analyses of doses in posts may be compared with ranges of doses approved by authorities. Likewise, we noticed that possible adverse drug reactions or misuse events were not detected due to the lack of MedDRA terms, especially when users write narrative descriptions of events or use non-technical expressions.

Table 4 – Samples of linguistic cues of indication, misuse or unknown adverse drug reactions (ADRs)

Misuse	<i>surdosé</i> ('overdosed')
	<i>mauvais dosage</i> ('bad dose')
	<i>tu ingurgites le triple de ce qui est recommandé</i> , 'you take 3 times more than what is recommended'
	<i>prendre des doses de cheval de X</i> ('take a raging / strong dose of X')
Indication	<i>je prends X pour ...</i> ('I take X for')
	<i>X utilisé comme / pour</i> ('X used as / for')
ADRs	<i>X m'empêchait de dormir</i> ('X keeps me awake')

Results of the Normalization Step

Applying the normalization rules on Corpus 2, we mapped to CUIs 263 out of the 344 different types of pathologies (76.4% of types), and 199 to MedDRA codes (57.8% of types). Errors were due to spelling, syntactic variation (*treatment failure* vs. *failure of treatment*), inflection (*panic attack* vs. *panic attacks*), derivation (*depressive* vs. *depression*), abbreviations (*rgo* vs. *reflux gastroesophagique*) or errors in CUI mappings.

Discussion

Communication in Internet fora is asynchronous and asymmetric, without specific interlocutors. This impacts the way medical information is expressed: incomplete, informal and creative expressions for health conditions abound, which make it difficult concept normalization and automatic analyses through NLP. Comparing pathological entities documented in SPCs medical drugs and unexpected pathologies in social media needs quality term detection and normalization. Our work is thus preliminary and suffers from the limitation that terms in patient language remained still unannotated, or entities were not normalized to accurate terms or CUIs. Moreover, we did not check the quality and correctness of the normalization step. We would like to explore normalization techniques based on word-embeddings and deep-learning. The comparison method was weak for detecting misuse, but helped in finding new ADRs; this opens the door to future work.

Regarding our supervised experiment, we lack enough data for training our model and generalizing our predictions on drug misuse to new datasets. We want to annotate SPCs of more medical drugs to gather a database of annotations to be

used in future work, especially for distant supervision approaches.

Using social media brings up other limitations related to: 1) the fact that users may not necessarily post their misuse behavior or ADRs; and 2) privacy concerns: despite users post contents to be publicly available, careful anonymization protocols are required, as we applied in the project [13].

Conclusions

We presented a method and initial experiments on pharmacovigilance analyses on social media based on NLP annotations of web fora. Through a supervised experiment with a minimal set of data, we showed that classification models might perform adequately. However, lacking of enough data to address current medications of interest, we resorted to Summary of Product Characteristics (SPCs) to overcome the data bottleneck. This approach, as far as we know, has not commonly being used and might be a source of knowledge for contrasting data reported in SPCs and unexpected users' health or conditions. A pharmacovigilance expert confirmed possible cases of misuse and some unknown ADRs. We make available the annotated data of SPCs.¹²

We highlight that our methods do not aim at replacing human decisions concerning users' behavior that could be drug misuse. These need pharmacovigilance experts to be validated; luckily, automated methods supporting these tasks make this validation faster and easier to be conducted.

Acknowledgements

The Agence Nationale de Sécurité du Médicament funded this work (grant ANSM-2016-S-076, PHARES project). We greatly thank the LIMICS team (U. Paris 13) for the data, and the pharmacovigilance team for annotating those posts.

References

- [1] A. Sarker, R. Ginn, A. Nikfarjam, *et al.*, Utilizing social media data form pharmacovigilance: A review, *J Biomed Inform* **54** (2015), 202-212.
- [2] A. Nikfarjam, A. Sarker, K. O' Connor, *et al.*, Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labelling with word embedding cluster features, *J Am Med Inform Assoc* **22(3)** (2015), 671-681
- [3] D. Cameron, G.A. Smith, R. Daniulaityte, *et al.*, PREDOSE: A semantic web platform for drug abuse epidemiology using social media, *J Biomed Inform* **46** (2013), 985-997.
- [4] European Medicines Agency (EMA). Guideline on good pharmacovigilance practices (GVP) Module VI. Available at: <https://goo.gl/NuGcdj> [Accessed March 2019]
- [5] F. Morlane-Hondère, C. Grouin, and P. Zweigenbaum, Identification of Drug-Related Medical Conditions in Social Media, *Proc of LREC*, Portoroz, Slovenia (2016).
- [6] E.G. Brown, L. Wood, and S. Wood, The Medical Dictionary for Regulatory Activities (MedDRA), *Drug Safety* **20** (1999), 109-117.
- [7] World Health Organization (WHO), Anatomical therapeutic chemical (ATC) classification index (2006), WHO Collab. Centre for Drug Statistics Methodology
- [8] J. Kalyanam, T. Katsuki, G.R. Lanckriet, *et al.*, Exploring trends of nonmedical use of prescription drugs and polydrug abuse in the Twittersphere using unsupervised machine learning, *Addictive Behavior* **65** (2017), 289-295.
- [9] E. Bigeard, N. Grabar, and F. Thiessard, Typology of Drug Misuse Created from Information Available in Health Fora, *Stud Health Technol Inform*, **247** (2018), 351-355.
- [10] E. Bigeard, N. Grabar, and F. Thiessard, Detection and analysis of drug misuses. A study based on social media messages. *Frontiers in Pharmacology* **9** (2018), p. 791.
- [11] M. Mintz, S. Bills, R. Snow and D. Jurafsky, Distant supervision for relation extraction without labeled data. *Proc. of ACL* (2009)
- [12] H. Poon, K. Toutanova, C. Quirk. Distant supervision for cancer pathway extraction from text. *Pacific Symposium on Biocomputing* (2014), 120-131.
- [13] B. Audeh, M. Beigbeder, A. Zimmermann, *et al.* Vigi4Med scraper: a framework for web forum structured data extraction and semantic representation. *PloS one*, **12(1)**, (2017), e0169658.
- [14] T. Laverigne, O. Cappé, and F. Yvon, Practical very large scale CRFs, *Proc. of ACL* (2010), 503-514.
- [15] P. Stenetorp, S. Pyysalo, G. Topić, *et al.*, BRAT: a web-based tool for NLP-assisted text annotation, *Proc. of the Demonstrations at EACL* (2012), 102-107.
- [16] K. Verspoor, A. Jimeno Yepes, L. Cavedon, *et al.*, Annotating the biomedical literature for the human variome. *Database*, (2013).
- [17] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.* Scikit-learn: Machine learning in Python, *J Machine Learning Res* **12** (2011), 2825-2830.
- [18] S. Santiso, A. Casillas, and A. Pérez, The class imbalance problem detecting adverse drug reactions in electronic health records, *Health Inform J* (2018).
- [19] C. Bousquet, É. Sadou, J. Souvigniet, *et al.*, Formalizing MedDRA to support semantic reasoning on adverse drug reaction terms, *J Biomed Inform* **49** (2014), 282-291.
- [20] O. Bodenreider, The Unified Medical Language System (UMLS): integrating biomedical terminology, *Nucleic acids research* **32**(suppl_1) (2004), D267-D270.

Addresses for correspondence

Leonardo Campillos-Llanos, leonardo.campillos@limsi.fr

Cyril Grouin, cyril.grouin@limsi.fr

¹² <https://bit.ly/2OlddCa> [Accessed March 2019]