

Study of online health discussion fora for the detection of medication misuses

Elise Bigeard^{1 2} Frantz Thiessard² Natalia Grabar¹

¹Univ. Lille, CNRS, UMR 8163 - STL - Savoirs Textes Langage, F-59000 Lille, France

²Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, team ERIAS, UMR 1219, F-33000 Bordeaux, France

6th June 2018
ICHI 2018 New York City USA

Outline

Introduction

Methods

Results

Goal

Automatic detection of drug misuse in internet forums

Goal

Misuse definition for our purpose (different from the WHO definition) :

Use of a prescription drug for a purpose or in a fashion not consistent with medical guidelines

Goal



Examples of misuses

- ▶ "I was taken to hospital because I messed with sleeping pills to get smashed"
- ▶ "Hello I was supposed to take decapeptyl 1,5 mg this evening but I didn't have the prescription on me and I ended taking 3 mg, did this ever happen to someone?"
- ▶ "When I was not feeling well last week, I was already thinking about stopping xanax against my doctor's instructions."

translated from French

Related work

Drug use in social media

[O'Connor et al., 2014] Adverse drug reactions on *Twitter*

[Yang et al., 2015] Unsupervised mining of ADR from *Medhelp*

[Sarker et al., 2015] Review of ADR monitoring with social media

Drug misuse in social media

[Kalyanam et al., 2017] Non medical use of opioids on *Twitter*

[Cameron et al., 2013] Semantic platform for drug abuse epidemiology in social media

Corpus

- ▶ From French health forum *Doctissimo*
- ▶ Must contain a drug
- ▶ Must be less than 2,500 characters long
- ▶ 1,900 messages
 - ▶ 1,400 messages randomly selected
 - ▶ 500 messages with various drugs

Manual annotation

Categories :

- ▶ **No use** "Can someone share their experience with xanax?"
- ▶ **Normal use** "I took a xanax I feel better already"
- ▶ **Misuse** "I just took an entire box of xanax do you think it will be enough?"

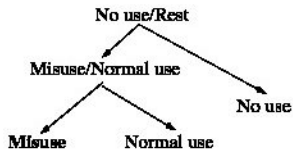
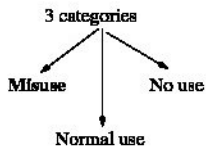
2 annotators including 1 pharmacologist

Results of annotation

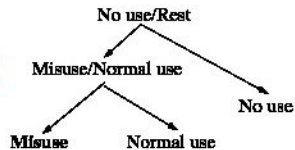
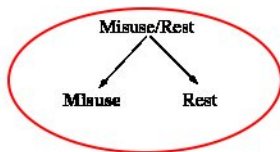
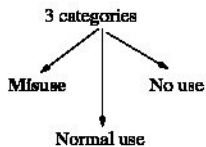
Kappa : 0.46 moderate agreement

No use	53%	999 messages
Normal use	30%	746 messages
Misuse	8%	155 messages

Classification task



Classification task



Indexing

Each message is indexed with ATC and ICD-10 generic codes

"I'd like to know if it's possible to take **xanax** while I breastfeed because I have **social anxiety**"

- ▶ xanax : N05
- ▶ social anxiety : F40

Indexing

Drugs lexicon :

- ▶ Theriaque
- ▶ *base publique du médicament*

Disorders lexicon :

- ▶ layman disorder names (anxiety)
 - ▶ Wikidata
- ▶ morphological family (anxious)
 - ▶ Lexique.org
 - ▶ Jeux de Mots [Lafourcade, 2007]

Features

- ▶ Lemmatized and vectorized text of the messages
- ▶ ATC and/or ICD-10 indexing

Algorithms

- ▶ NaiveBayes Family
- ▶ NaiveBayes Multinomial
- ▶ Random Forest
- ▶ J48 (trees)
- ▶ Simple Logistic
- ▶ SMO (SVM)

Results

- ▶ up to 0,773 F-measure
 - ▶ NaiveBayes, with drugs indexing
- ▶ ATC codes give small increase

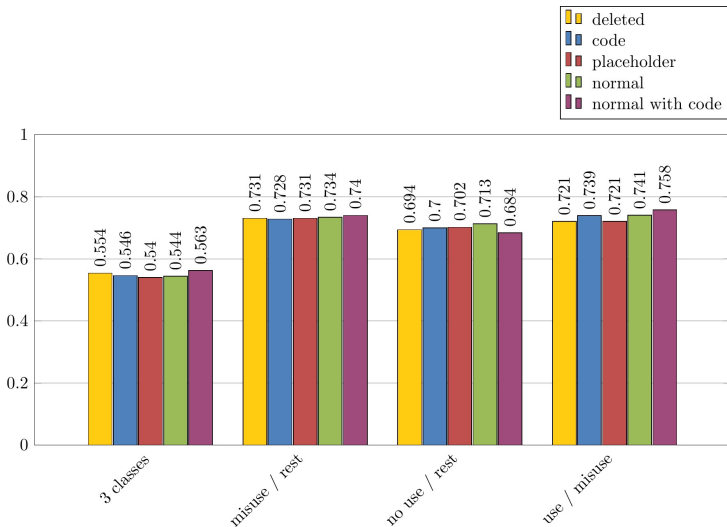
Role of drugs and disorders

new experiment :
remove/change drugs and disorders in messages to see if it impairs
the classification

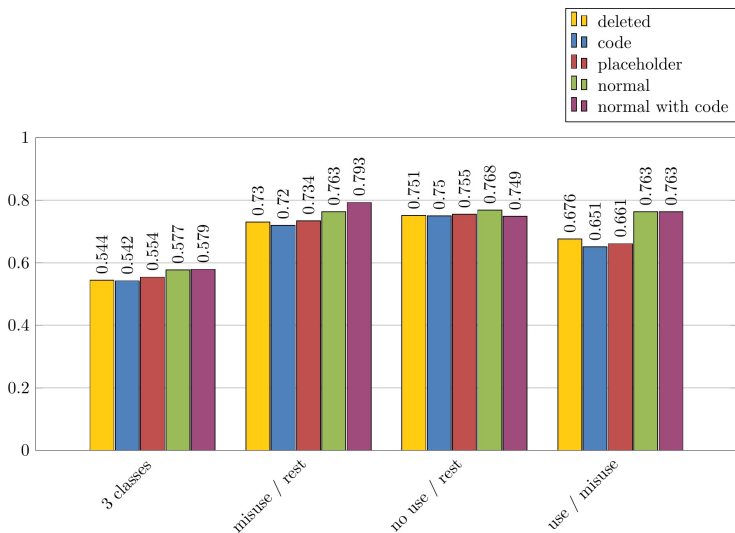
Role of drugs and disorders

- ▶ deleted drug/disorder
"I'm taking"
- ▶ name of the drug/disorder replaced by the generic code
"I'm taking N06"
- ▶ name of the drug/disorder replaced by a unique placeholder
"I'm taking \$drug\$"
- ▶ unmodified text
"I'm taking seroplex"
- ▶ unmodified text with the code added to it
"N06 | I'm taking seroplex"

Results for drugs



Results for disorders



Conclusion

Conclusion


- ▶ It's possible to detect misuse with simple supervised classification


Future work


- ▶ Larger corpus
- ▶ More varied forums and disorders

Thank you

Questions?

- 

Cameron, D., Smith, G. A., Daniulaityte, R., Sheth, A. P., Dave, D., Chen, L., Anand, G., Carlson, R., Watkins, K. Z., and Falck, R. (2013).
 PREDOSE: a semantic web platform for drug abuse epidemiology using social media.
46(6):985–997.
- 

Kalyanam, J., Katsuki, T., Lanckriet, G. R. G., and Mackey, T. K. (2017).
 Exploring trends of nonmedical use of prescription drugs and polydrug abuse in the twittersphere using unsupervised machine learning.
65:289–295.
- 

Lafourcade, M. (2007).
 Making people play for lexical acquisition.
7th Symposium on Natural Language Processing.
jeuxdemots.

-  O'Connor, K., Pimpalkhute, P., Nikfarjam, A., Ginn, R., Smith, K. L., and Gonzalez, G. (2014).
Pharmacovigilance on twitter? mining tweets for adverse drug reactions.
pages 924–933.
-  Sarker, A., Ginn, R., Nikfarjam, A., O'Connor, K., Smith, K., Jayaraman, S., Upadhaya, T., and Gonzalez, G. (2015).
Utilizing social media data for pharmacovigilance: a review.
Journal of Biomedical Informatics.
-  Yang, M., Kiang, M., and Shang, W. (2015).
Filtering big data from social media—building an early warning system for adverse drug reactions.
Journal of Biomedical Informatics.