Detection and analysis of medical misbehavior in online forums

Elise Bigeard[1] Natalia Grabar[1]

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

23th October 2019
SNAMS 2019 Grenada Spain

Medication non-compliance definition

Medication non-compliance : Any situation when the patient is not following the instructions of the physician / drug leafleat regarding medication

Medication non-compliance examples

- Errors, missed intake
- Underuse to avoid a side effect
- Overuse to get a stronger effect
- Misuse of neuroleptics for psychotropic effects
- Etc

Drug non-compliance examples

- "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."

Goal

Goal : Detect these situations in messages written by patients on Internet health forums

Method : Supervised classification

Corpus

- in French
- From French health forum *Doctissimo*
- We select the messages that contain at least one drug
- and are at most 2,500 characters long

Supervised method

Supervised classification between two categories :

- **Compliance** "I took a xanax I feel better already"
- **Non-compliance** "I just took an entire box of xanax do you think it will be enough?"

Results of annotation

1850 messages manually annotated

Compliance	93%	1717 messages
Non-compliance	7%	133 messages

Second round of annotation

- We train a NaiveBayes classifier with these messages
- And manually annotate the messages assigned non-compliant

Second round of annotation

	first round	second round	total
Compliance	1717	816	2533
Non-compliance	133	218	351

Preprocessing

- Text is lemmatised
- Grammatical words are removed
- Numbers are neutralised
- Diacritics are removed

Classifier

- Features : Vectorized text + ATC code of medications
- Algorithm :
  - Naive Bayes
  - Naive Bayes Multinomial
  - J48
  - Simple Logistic
- Evaluation
  - 10-folds cross validation
  - Results shown for the non-compliance category

Results

	Precision	Recall	F-mesure
Naive Bayes	0.281	**0.610**	0.385
Naive Bayes Multinomial	0.339	0.581	**0.428**
J48	0.313	0.199	0.243
Simple Logistic	**0.394**	0.011	0.022

(Reached F-mesure of 0,513 after publication)

Results

- Enough to detect non-compliance
- But not enough for the results to be used unfiltered

Analysis

We manually analyse and categorise the content of the messages

Medication mentioned

Code	Class	Occurrences	
N05	Psycholeptics	168	47.9 %
N06	Psychoanaleptics	134	38.18%
G03	Sex hormones	32	9.12%
N02	Analgesics	31	8.8%
A03	Gastrointestinal disorders	14	4.0%
N07	Nervous system, other	10	2.8%
R06	Antihistamines	8	2.3 %
H03	Thyroid therapy	7	2.0%
N03	Antiepileptics	6	1.7%
Other	Other	71	20.2 %

Types of non-adherence

Category	Messages	Percentage
under-use	101	28.7%
over-use	96	27.3%
contraindication	28	7.9%
error	24	6.8%
misuse	22	6.2%
automedication	20	5.7%
method	11	3.1%

Tags associated with under-use

Category	Messages	Percentage
discontinuation	55	54.4%
under-dosage	18	18%
addiction and habituation	10	9.9%
over-use	9	8.9%
error	3	3.0%
over dosage	3	3.0%
method	1	1%
no prescription	1	1%

Under-use : Why ?

- Experiences side effects (19 messages)
- Afraid of addiction and/or withdrawal effects (9 messages)
- Medication is not effective (5 messages) or useful (2 messages)
- Afraid of possible side effects (4 messages)
- Afraid of the medication for no specified reason (4 messages)

Tags associated with over-use

Category	Messages	Percentage
over dosage	43	44.8%
addiction and habituation	15	15.6%
no prescription	10	10.4%
under-use	9	9.4%
misuse	8	8%
contraindication	8	8.3%
alcohol	7	7.3%
discontinuation	4	4.2%
psychotropic	4	4.2%
under dosage	3	3.1%
error	2	2.1%
intake refusal	1	1.0%
method	1	1.0 %

Tags associated with over-use

Messages talking about withdrawal process

Category	Messages	Percentage
over dosage	43	44.8%
addiction and habituation	15	15.6%
no prescription	10	10.4%
under-use	9	9.4%
misuse	8	8%
contraindication	8	8.3%
alcohol	7	7.3%
discontinuation	4	4.2%
...		

Tags associated with contraindication

Category	Messages	Percentage
Alcohol	11	39.2%
Pregnancy	10	35.7%
Over-use	8	28.6%
Addiction and habituation	7	25.0%
Over dosage	5	17.9%
Psychotropic	2	7.1%
Discontinuation	1	3.6%
Misuse	1	3.6%
Automedication	1	3.6%

Tags associated with misuse

Category	Messages	Percentage
Over-use	8	36.4%
Weight loss	7	31.8%
Psychotropic use	5	22.7%
Addiction and habituation	3	13.6%
Over dosage	2	9.1%
Involuntary	1	4.5%
Alcohol intake	1	4.5%
Contraindication	1	4.5%
No prescription	1	4.5%

Conclusion

- Supervised classification with machine learning reaches 0.428 of F-mesure
- Over and under-use are the most common situations (30% each)
- Under-use happens mostly to avoid side-effects
- Over-use happens mostly because of addiction