Detecting drug non-compliance in Internet fora using information retrieval and machine learning approaches

Elise Bigeard[1]   Frantz Thiessard[2]   Natalia Grabar[1]

[1]Univ. Lille, CNRS, UMR 8163 - STL - Savoirs Textes Langage, F-59000 Lille, France
[2]U Bordeaux, Inserm, Bordeaux Population Health Research Center, team ERIAS, UMR 1219, F-33000 Bordeaux, France

30th August 2019
Medinfo 2019 Lyon France

Drug non-compliance definition

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

Drug non-compliance definition

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

- ► 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 : detect these situations in messages written by patients on Internet health forums

Method :

- ▶ Supervised (Information retrieval and machine learning)
- ▶ Unsupervised (Information retrieval)

Supervised method

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?"

Corpus

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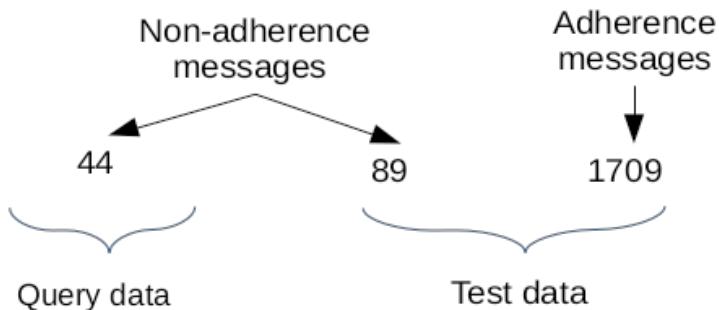- and are at most 2,500 characters long

Results of annotation

1850 messages manually annotated

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

Separation of train/test

Training

keywords = all words from training data with weight according to
frequency

Results

	Precision	Recall	F-mesure
Top 10	0.400	0.045	0.081
Top 20	0.300	0.067	0.109
Top 100	**0.500**	0.561	0.529

Results

	Precision	Recall	F-mesure
Top 10	0.400	0.045	0.081
Top 20	0.300	0.067	0.109
Top 100	**0.500**	0.561	0.529
NaiveBayes	0.761	0.897	0.824

Unsupervised method

Unsupervised

We search topics related to non-adherence among 20,000 messages

Unsupervised

Topics

- ▶ Weight changes
- ▶ Suicide
- ▶ Over-use
- ▶ Recreational medication use

Weight changes

- Medication taken to lose weight
- Medication not taken because of a side effect of gaining weight

Weight changes

Keywords :

- poids, kilo, grossir, maigrir
- weight, kilo, to take weight, to lose weight

Weight changes

Among 20 top messages

- ► 17 about the side effect
- ► 1 about medication to lose weight

Weight changes

"I need informations about seroplex (...) From what I gathered 90% of messages speak about weight gain. For me, I don't care about adverse effects like headache, stomach ache, nausea, tiredness, etc... But there's no way this medication is going to give me even one single kilo !"

Suicide

Among 20 top messages

- ▶ 10 about increased suicide risk of antidepressants
- ▶ 1 about a suicide attempt during medication withdrawal

Suicide

"I was afraid of taking antidepressants, I had to see 3 shrinks
before deciding to take them because I was afraid of being worse,
of side effects like suicide..."

Overuse

Keyword : boxes

Overuse

"I used to take the equivalent of 6 **boxes** of codoliprane under various forms for more than 10 years."

Overuse

Among 20 top messages

- ▶ 6 about overuse
- ▶ 3 about high dosages that may be overuse
- ▶ 2 about suicide attempts
- ▶ 2 where people propose to share unused boxes of medication

Recreational drug use

- drug, to drug (different than medical drug in French)
    - 15 messages refering to medication as drug
    - "I was drugged with xanax for 3 years"

Recreational drug use

- ▶ drug, to drug (different than medical drug in French)
  - ▶ 15 messages refering to medication as drug
  - ▶ "I was drugged with xanax for 3 years"
- ▶ to be high
  - ▶ 9 messages are misuse where the effect is seeked
  - ▶ 10 messages are under-use where the effect is avoided
  - ▶ "I have to take xanax I only take 0,25 mg instead of 50 mg or I would be high"

Conclusion

- Supervised classification with machine learning reaches 0.815 of F-mesure
- Information retrieval can detect specific cases of non-adherence
  - 0.45 average precision
  - Unsupervised