

Social Media Mining for Toxicovigilance

Automatic Monitoring of Prescription Medication Abuse from Twitter

Abeed Sarker (@sarkerabeed)



Health Language Processing Lab

Research Associate

Department of Biostatistics,
Epidemiology and Informatics

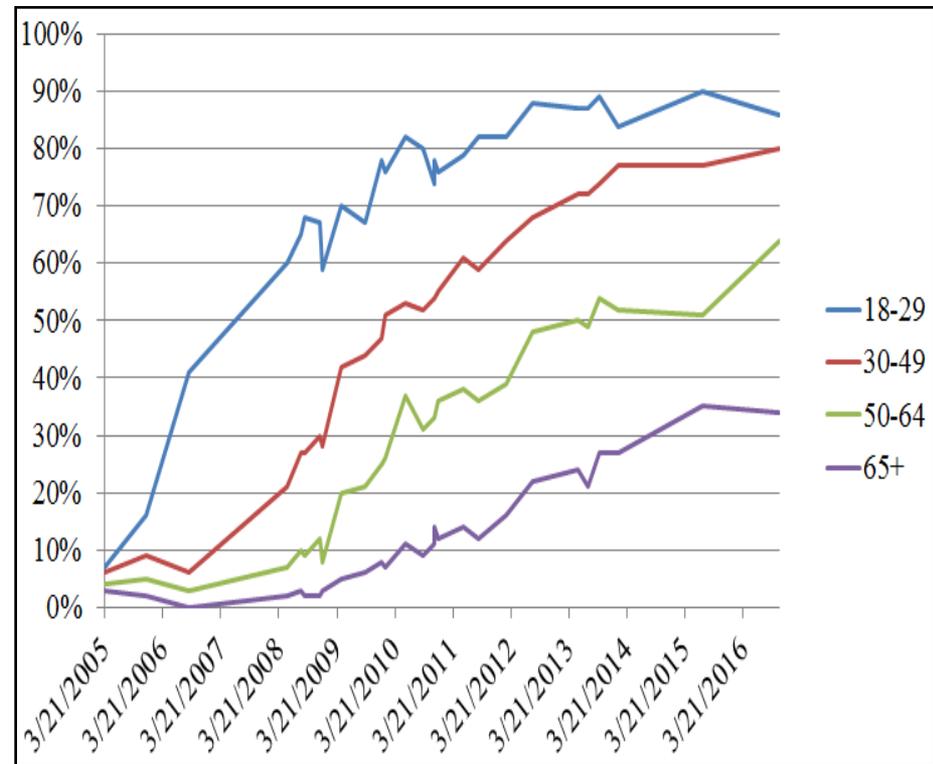
09/20/2017



Why social media?

- ◆ **26% of internet users actively discuss health information. Of that group¹**
 - 30% *changed behavior* as a result
 - 42% discussed *current* medical conditions
- ◆ **Abundance of health-related data**
- ◆ **Abundance of data on PM abuse**

Trends— U.S.A



Trends in social media usage among different age groups since 2005.²

Challenges of mining social media data

- ◆ **Most data is in text form**
- ◆ **Noisy**
 - Most of the data is irrelevant
 - Bots
 - Lack of context (particularly in Twitter)
- ◆ **Use of colloquial language**
 - Challenging to decipher
 - Drug abuse information may use street language
- ◆ **Machine learning algorithms show lower performance on social media data**
- ◆ **Addressing these challenges may open up new opportunities**

Past work on social media mining for health

- ◆ **Pharmacovigilance**
 - Adverse drug reaction detection, extraction, concept normalization
- ◆ **Nutritional supplements' safety assessment**
- ◆ **User sentiment assessment for medications**
 - Also general sentiment analysis
- ◆ **Cohort identification and monitoring (for pregnancy)**
 - Cohort detection and *timeline* collection
 - Cohort categorization and information extraction
 - Medication exposure and pregnancy outcomes (ongoing)
- ◆ **Toxicovigilance/addictovigilance**
 - Detection of PM abuse + temporal patterns of PM abuse

PM abuse information on social media

Abuse	Non-abuse
Did I just snort an oxy ? Is it like 10:30? What can I say I know a Friday when I smell one.....	I love how sometimes you're a personification of an oxymoron and sometimes minus the " oxy "
I poured 5 shots of whiskey into my lemonade and I took 4 OxyContin and I feel so great because I'm not feeling at all	Seroquel is prescribed. i use valerian root sometimes too. mostly i don't sleep
about to be cracked on adderall to survive today	speaking of oxycodone .. i need to take mine. This pain is ridiculous
i'm just gonna shower and overdose on Seroquel so I'll sleep until morning	Y'all got to lay off the Percs y'all looking crazy out here
popped Adderall tonight hahahah let's finish this 100 page paper	Where the Percs at

Automated detection of PM abuse

◆ Goals:

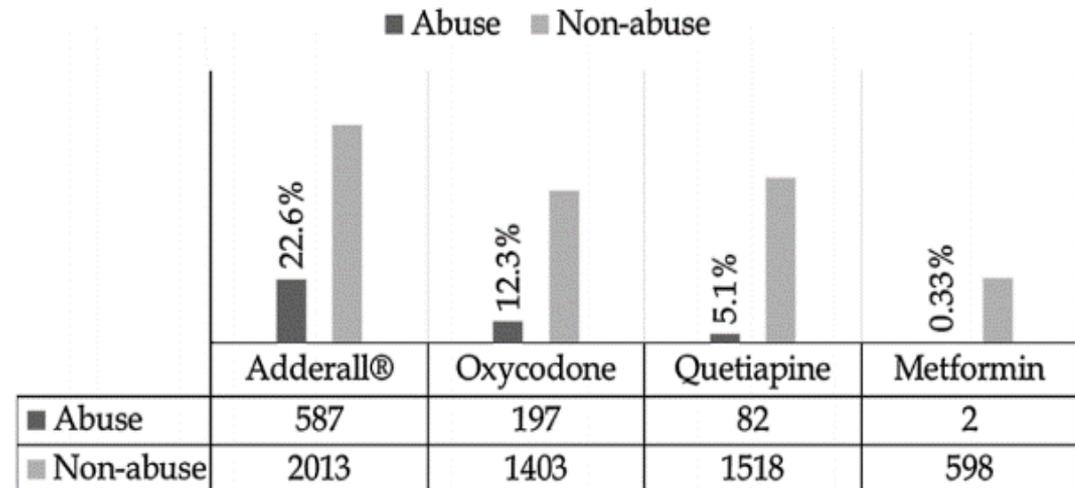
1. To verify that abuse-prone PMs have significantly higher abuse information compared to non-abuse-prone PMs
2. To assess if abuse information can be automatically detected via natural language processing and machine learning
3. To assess if automatically detected temporal patterns of abuse match manually verified ones

◆ Twitter as the source of data

- Chosen medications: oxycodone, quetiapine and Adderall® (plus metformin)

◆ Annotation

- 6400 tweets
- 2 annotators



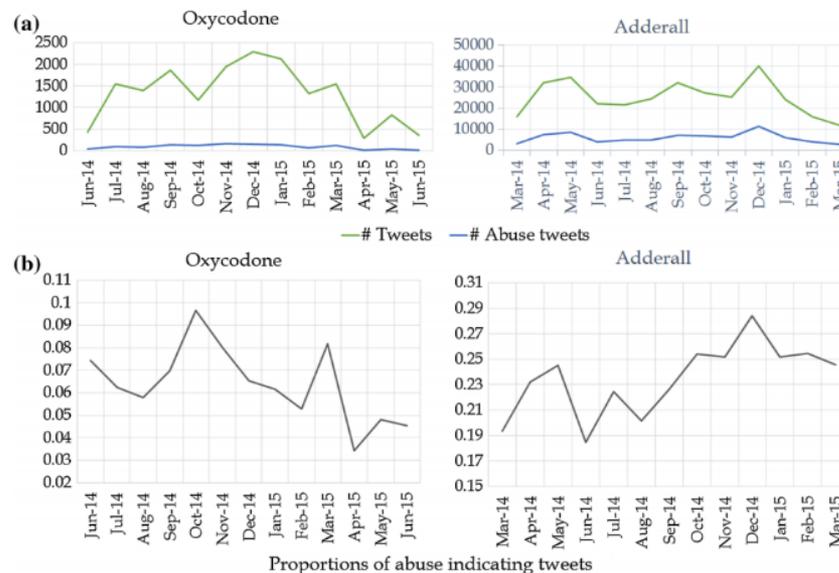
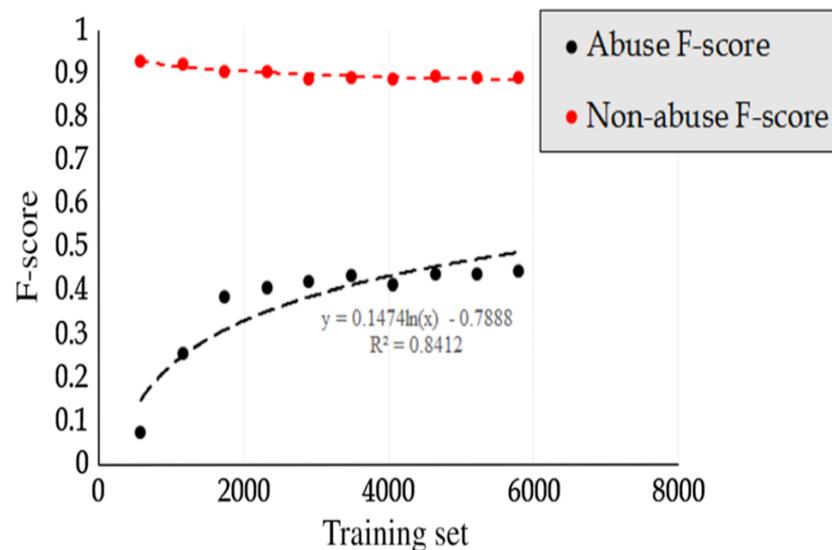
Automatic classification

- ◆ **Supervised classification using annotated data for training and some feature engineering**
- ◆ **Features**
 - N-grams, abuse-indicating lexicons, synonym detection and word clusters
- ◆ **Evaluation on annotated data via 10-fold cross validation**
- ◆ **Trained classifier run on ~130k tweets over ~1 year period to analyze temporal patterns**

Results

Classifier	Abuse F-score	Non-abuse F-score	Accuracy
NB	0.39	0.84	75%
wSVM	0.45	0.89	81%
Stacking	0.46	0.89	82%

- Stacking multiple classifiers produced best performance
- Trendline suggests that ~50k training tweets will lead to abuse class F-scores ~0.80
- Automatically detected patterns of abuse over time (for Adderall®) resembled Hanson et al. (2013) [2]



Summary

- ◆ **Social media encapsulates an abundance of data about PM abuse and the abusers**
- ◆ **Converting from data to information and from information to knowledge requires the development of NLP and ML techniques**
- ◆ **Artificial intelligence can aid relevant medical practitioners leverage social media to address the PM abuse crisis**
 - Collaborative, interdisciplinary research is essential
- ◆ **The question is not *if* social media can play a role, but *what* role it can play and *how* it can be best utilized**

Contact

Abeed Sarker

Email: abeed@penmedicine.upenn.edu

Twitter: @sarkerabeed

Phone: +1-215-746-1700 (office)

Phone: +1-602-474-6203 (personal)

Website: abeedsarker.com (personal)

