Social media mining for pharmacovigilance—challenges and progress

Presented By: Abeed Sarker
Email: abeed.sarker@asu.edu
DIEGO Lab
Department of Biomedical Informatics
Arizona State University

Project funded by
NLM/NIH grant number 1R01LM011176-01
Social media mining pipeline

- Data collection
- Annotation
- Resource creation (e.g., lexicons)
- Classification
- Information extraction
- Normalization
- Case studies
Adverse drug reaction (ADR) monitoring from social media

Over 770,000 people are injured or die each year in hospitals from ADRs

- Improved reporting mechanisms recommended

30% of adults are likely to share health related information on social media

Twitter has over 645 million users

- 9,100 tweets per second


A taste of Twitter ADR lingo

- @#### HA! Not if you're on #Seroquil. EXTREMELY vivid dreams that stay in conscious memory. Very #Freaky! Any idea why?
- Off to see the gi consultant this week. Hope there's something other than humira to try as not working also hair falling out.
- The 100mg tabs of trazodone my gp prescribed are too much, now that I don't take them every night. Still zombieish after an hour awake
- Gone from 50mg to 150mg of Serequel last night. Could barely wake up this morning and I feel like my body is made of lead
Data collection and annotation

Phonetic spelling variants for capturing misspelled medication names\(^1\) (http://diego.asu.edu/Publications/ADRSpell/ADRSpell.html)

- Seroquel -> siroquil, seroquil etc.

Binary and full ADR annotations\(^2,3\)

Multiple trained annotators + pharmacology expert to resolve annotation disagreements

2 O’Connor et al. Pharmacovigilance on Twitter. AMIA Annu Symp Proc. 2014.
Works to calm **mania** or **depression** but **zonks** me and scares me about **diabetes** issues reported.

- **Indication:** mania (C0338831)
- **Indication:** depression (C001157)
- **ADR:** drowsiness (C0013144)
- **Other:** diabetes
- **Indication:** crying (C0010399)
- **Adverse reaction:** emotional indifference (C0001726)

stops me from **crying** most of the time, **blocks most of my feelings**
Annotated resources

15,000+ binary annotations and 2000+ full annotations have been made publicly available: http://diego.asu.edu/downloads/

Other annotations:
- Prescription medication abuse
- Drug related chatter
Text classification

Generate a large set of features, representing semantic properties (e.g., sentiment, polarity, and topic), from short text nuggets

- Combine training data from different corpora in attempts to boost classification accuracies
- Effort in resource creation pays off

Other text classification tasks:
- Drug abuse classification
- Drug safety classification

Resources at: http://diego.asu.edu/Publications/ADRCClassify.html
3 Patki et al. Mining adverse drug . going beyond extraction. BioLinkSig. 2014.
ADR extraction

Goal: to extract exact mentions of ADRs and other information

Traditional, lexicon-based approaches perform poorly on social media text

a) #Schizophrenia\textsuperscript{indication} #Seroquel did not suit me at all. Had severe \textsuperscript{tremors} \textsubscript{ADR} and weight \textsuperscript{gain} \textsubscript{ADR}.

b) I felt awful, it made my \textsuperscript{stomach hurt} \textsubscript{ADR} with bad \textsuperscript{heartburn} \textsubscript{ADR} too, \textsuperscript{horrid taste in my} \textsuperscript{mouth} \textsubscript{ADR} tho it does tend to clear up the \textsuperscript{infection} \textsubscript{indication}.

Traditional, lexicon-based approaches perform poorly on social media text
Word clusters and conditional random fields

Our approach using conditional random fields outperforms lexicon based approaches

Particularly ambiguous ADRs captured by word cluster feature

I had the side effect of a \textit{bloody nose}^{ADR} and hated it.

Made me feel \textit{numb}^{ADR} and \textit{apathetic}^{ADR} to pretty much everything ... made me \textit{gain about 40 lbs}^{ADR}.

Working well no side effects from this besides \textit{cotton mouth}^{ADR}.

1 Nikfarjam et al. Pharmacovigilance from social media.. sequence labeling with word embedding cluster features. \textit{JAMIA}. 2015.
Publication resources: \url{http://diego.asu.edu/Publications/ADRMine.html}
## Examples of extracted ADRs

<table>
<thead>
<tr>
<th>Type</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADR</td>
<td>zap noises</td>
</tr>
<tr>
<td>ADR</td>
<td>pancreatitis</td>
</tr>
<tr>
<td>ADR</td>
<td>digestive problems</td>
</tr>
<tr>
<td>ADR</td>
<td>sleep no problem</td>
</tr>
<tr>
<td>Indication</td>
<td>anxiety</td>
</tr>
<tr>
<td>ADR</td>
<td>feel like a zombie</td>
</tr>
<tr>
<td>ADR</td>
<td>hair falling out</td>
</tr>
<tr>
<td>ADR</td>
<td>sleep</td>
</tr>
</tbody>
</table>
Concept normalization

A set of rule-based techniques followed by semantic similarity based techniques\(^1\)

Best F-score: 0.603

\(^1\) Emadzadeh et al. TO BE PUBLISHED. 2016.
## Normalization examples

<table>
<thead>
<tr>
<th>Extracted ADR</th>
<th>Expected</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>feel down</td>
<td>c0011570-Depression</td>
<td>c0011570</td>
</tr>
<tr>
<td>increase my weight</td>
<td>c0043094-Weight gain</td>
<td>c0043094</td>
</tr>
<tr>
<td>gain so much weight</td>
<td>c0043094-Weight gain</td>
<td>c0043094</td>
</tr>
<tr>
<td>fewer hours sleep</td>
<td>c0235161-Sleep loss</td>
<td>c0235161</td>
</tr>
<tr>
<td>feel like need to throw up</td>
<td>c0027497-Nausea</td>
<td>c0917799-Hypersomnia</td>
</tr>
<tr>
<td>just eat, and eat</td>
<td>c0232461-Apetite increase</td>
<td>c0015672-Fatigue</td>
</tr>
<tr>
<td>falling asleep every day</td>
<td>c0541854-Daytime sleepiness</td>
<td>c0917801-Insomnia</td>
</tr>
</tbody>
</table>
Medication abuse monitoring

Users post information about medication abuse on social media

- about to be cracked on adderall to survive today
- i’m just gonna shower and overdose on Seroquel so I’ll sleep until morning.
- popped Adderall tonight hahahah let’s finish this 100 page paper
- an oxycodone high from snorting lasts for one hour, if it is swallowed, your looking at three hour high.
Adderall® vs. oxycodone abuse patterns

Supervised classification to investigate patterns of abuse\(^1\)

Ongoing tasks and resources

Medication case studies

Signal comparison from distinct sources

Resources:

- ADR and drug abuse annotations
  - http://diego.asu.edu/downloads/
- Unlabeled drug related Twitter chatter
- Other published resources
  - e.g., http://diego.asu.edu/Publications/ADRSMReview/ADRSMReview.html
Disclosure & Acknowledgments

- The authors have no competing interests

We thank the NIH/NLM support of this project, award 1R01LM011176-01

Take home message:
- if u took a med, and xperience sum reaction, plz tweet about it ... #psb16 #smmpsb16

Contact: abeed.sarker@asu.edu

Questions?
Thank You.

Questions?

Contact: abeed.sarker@asu.edu
References


