Health monitoring and surveillance from social media

A discussion of past work, progress, easy and difficult tasks

Abeed Sarker (@sarkerabeed)
Research Associate
Department of Biostatistics, Epidemiology and Informatics

03/15/2016
Health Language Processing lab overview

❖ Focus areas
  • Social media
  • Medical literature
  • Electronic health records/expert authored texts

❖ Sample current (and upcoming) projects
  • Deriving knowledge from drug-related chatter in social media
    – Primarily Twitter
  • Safety surveillance of nutritional products
  • Aiding phylogeography of zoonotic viruses
  • Assessing the accuracy of antibiotic prescriptions automatically
  • Text classification, information extraction and normalization of various types of health-related texts

❖ Website: https://healthlanguageprocessing.org/
Presentation overview

- Social media mining for health
  - Why social media?

- Pharmacovigilance from social media

- Easy and difficult tasks in pharmacovigilance from social media
  - Annotation, classification, extraction, normalization, signal generation

- Future possibilities in pharmacovigilance from social media

- Other tasks utilizing social media
  - Prescription medication abuse
  - Cohort identification and monitoring using social media (e.g., pregnant women)
Why social media for health?

Health and internet

- 26% of internet users actively discuss health information. Of that group …
  - 30% *changed behavior* as a result
  - 42% discussed *current* medical conditions

- Abundance of health-related knowledge online

Trends—U.S.A

Trends in social media usage among different age groups since 2005.

Progress and possibilities

- Growing interest - from just over 100 to 2000 publications including “social media” or “social network” in PubMed over the last 10 years:
  - Early approaches used keywords and hashtags—e.g., for monitoring flu spread (Szomszor et al., 2010; Lampos et al., 2010)
  - Language analysis of social media and lexicon-based approaches (Schwartz et al., 2013; Leaman et al., 2010)
  - NLP-intensive/learning-based approaches (pharmacovigilance, user behavior analysis, detection of topics)

- Possibilities in epidemiology: how do we get observations over time for specific groups that share particular characteristics?

- Combining social media with other sources
  - e.g., in flu outbreaks, pharmacovigilance
Projects/studies on social media mining

- Pharmacovigilance
- Toxicovigilance/addictovigilance
- Nutritional supplements’ safety assessment
- User sentiment assessment for medications
- Cohort identification and monitoring
- Medication safety assessment during pregnancy
Pharmacovigilance from social media

- The activities relating to the detection, assessment, understanding and prevention of adverse effects attributable to prescription drugs

- Pharmacovigilance begins during clinical trials and continues after the drug is released into the market

- Due to limitations of clinical trials, not possible to fully assess the consequences of taking a specific drug prior to its release
  - e.g., Vioxx®—between 88,000 and 140,000 cases of serious heart disease

- Public health problem
  - deaths and hospitalizations numbering in millions (up to 5% hospital admissions, 28% emergency visits, and 5% hospital deaths), and associated costs of about seventy-five billion dollars annually
Pharmacovigilance from social media

- What do systematic reviews tell us?
  - Under-reporting is a problem in current surveillance systems. (37 studies from 12 countries) showed median under-reporting rate was 94% (82-98%). For serious/severe, 85%.

  - Abundant reports in SM. (29 studies that compared SM to other sources) showed a higher frequency of adverse events was found in social media and that this was particularly true for ‘symptom’ related and ‘mild’ adverse events.

  - Patient reporting brings different perspective, more info. (34 studies) Patient reporting brings novel information, more detail, info on severity and impact of ADRs in daily life.
Challenges

- **Data collection**
  - Drug names are often misspelled
  - Other misspellings

- **Noise**
  - Data imbalance (majority of posts contain no ADRs)
  - Adverse reactions are often expressed creatively
  - Posts lack context

- **Availability and incompleteness**
  - Complete data about individual cases may not be available
  - High dropout rate

- **Real world problems are either too easy or too difficult to solve** – Mark Johnson (Macquarie University)
Twitter ADR lingo

- HA! Not if you're on #Seroquil. EXTREMELY vivid dreams that stay in conscious memory. Very #Freaky! Any idea why?

- I'm def suing cymbalta. I can't wait until its out of my system. Get out!!!!!!! Nowwww!!!!!! You turn peaceful people into the hulk!

- Apparently, Baclofen greatly exacerbates the "AD" part of my ADHD. Average length of focus today: about 30 seconds.

- 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
Sources for pharmacovigilance research

- **Twitter**
  - Large user base (319 monthly active users now)
  - Publicly available API with large dataset available
  - Possible to share and distribute data
  - Noisy chatter
  - Increasing bots

- **DailyStrength**
  - Online health community with over 500 support groups
  - Targeted and detailed posts
  - Smaller user base
  - Not possible to redistribute data for research
Social media mining pipeline

- Social Media
  - Data Collection
- Raw Data
  - Classification
- Filtered Data
  - ADR Extraction
- Drug-ADR Pairs
  - Statistical Analysis
  - Generalizing ADR mentions
    - What’s the denominator?
    - Other epidemiological questions.
    - How do we find rare ADRs?

Other resources (e.g. lexicons, topics)

Needs mapping/normalization
Data collection and annotation

- Phonetic spelling variants for capturing misspelled medication names
  (http://diego.asu.edu/Publications/ADRSpell/ADRSpell.html)
  - Seroquel -> seraquil, seroquil etc.

- Binary and full ADR annotations
  - Mostly available publicly

- Multiple trained annotators + pharmacology expert to resolve annotation disagreements
  - Many disagreements, many fights!
  - Iterations!
Annotation example

... works to calm mania or depression but zonks me and scares me about diabetes issues reported.

Other: diabetes

Indication: crying (C0010399)

Indication: depression (C001157)

ADR: drowsiness (C0013144)

ADR: emotional indifference (C0001726)

... stops me from crying most of the time, blocks most of my feelings
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/adaptation pays off
  - SVMs work great!

- Data and resources available
  - [http://diego.asu.edu/Publications/ADRCClassify.html](http://diego.asu.edu/Publications/ADRCClassify.html)
  - [https://bitbucket.org/asarker/adrbinaryclassifier](https://bitbucket.org/asarker/adrbinaryclassifier) or [https://bitbucket.org/pennhlp/adrbinaryclassifier](https://bitbucket.org/pennhlp/adrbinaryclassifier)

- PSB 2016 shared task; possibly AMIA 2017
  - [http://diego.asu.edu/psb2016/task1data.html](http://diego.asu.edu/psb2016/task1data.html)
## Classification performances

Leave-one-out classification scores over the three data sets showing how accuracies and ADR F-scores are affected as one feature is removed from the set.

<table>
<thead>
<tr>
<th>Features</th>
<th>TW</th>
<th>DS</th>
<th>ADE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy</td>
<td>ADR F-score</td>
<td>Accuracy</td>
</tr>
<tr>
<td>All</td>
<td>86.2</td>
<td>0.538</td>
<td>83.6</td>
</tr>
<tr>
<td>N-grams</td>
<td>80.7</td>
<td>0.424</td>
<td>82.6</td>
</tr>
<tr>
<td>UMLS STs and CUIs</td>
<td>85.7</td>
<td>0.505</td>
<td>82.8</td>
</tr>
<tr>
<td>Syn-set expansions</td>
<td>86.1</td>
<td>0.545</td>
<td>84.0</td>
</tr>
<tr>
<td>Change phrases</td>
<td>87.1</td>
<td>0.521</td>
<td>83.9</td>
</tr>
<tr>
<td>ADR lexicon match</td>
<td>86.1</td>
<td>0.492</td>
<td>83.5</td>
</tr>
<tr>
<td>Sentitword score</td>
<td>86.2</td>
<td>0.530</td>
<td>82.8</td>
</tr>
<tr>
<td>Topics</td>
<td>86.1</td>
<td>0.535</td>
<td>83.7</td>
</tr>
<tr>
<td>Other features</td>
<td>86.9</td>
<td>0.534</td>
<td>83.6</td>
</tr>
</tbody>
</table>

### TW Actual vs. Predicted F-score

- TW

### DS Actual vs. Predicted F-score

- DS

### TW with additional features

- TW + ADE
- TW + DS

### DS with additional features

- DS + ADE
- DS + TW

| TW + ADE | 0.538 |
| TW + DS  | 0.545 |
| TW + TW  | 0.597*|
| DS + ADE | 0.674 |
| DS + TW  | 0.704*|
ADR extraction: ADRMine

- Automatically extract exact ADR mentions (which can be mapped)
  - Long term goal: automatically extract health information

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

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

- Conditional random fields (CRF) classifier

- Word/phrase generalization using standard tools (e.g., MetaMap) performs poorly

- Clustering similar terms/phrases help in capturing non-standard expressions

I had the side effect of a bloody nose\textsuperscript{ADR} and hated it.
Made me feel numb\textsuperscript{ADR} and apathetic\textsuperscript{ADR} to pretty much everything ... made me gain about 40 lbs\textsuperscript{ADR}.
Working well no side effects from this besides cotton mouth\textsuperscript{ADR}. 
## Sample clusters

<table>
<thead>
<tr>
<th>Cluster#</th>
<th>Topic</th>
<th>Examples of clustered words</th>
</tr>
</thead>
<tbody>
<tr>
<td>c₁</td>
<td>Drug</td>
<td>abilify, adderall, ambien, ativan, aspirin, citalopram, effexor, paxil, …</td>
</tr>
<tr>
<td>c₂</td>
<td>Signs/Symptoms</td>
<td>hangover, headache, rash, hive, …</td>
</tr>
<tr>
<td>c₃</td>
<td>Signs/Symptoms</td>
<td>anxiety, depression, disorder, ocd, mania, stabilizer, …</td>
</tr>
<tr>
<td>c₄</td>
<td>Drug dosage</td>
<td>1000mg, 100mg, .10, 10mg, 600mg, 0.25, .05, …</td>
</tr>
<tr>
<td>c₅</td>
<td>Treatment</td>
<td>anti-depressant, antidepressant, drug, med, medication, medicine, treat, …</td>
</tr>
<tr>
<td>c₆</td>
<td>Family member</td>
<td>brother, dad, daughter, father, husband, mom, mother, son, wife, …</td>
</tr>
<tr>
<td>c₇</td>
<td>Date</td>
<td>1992, 2011, 2012, 23rd, 8th, april, aug, august, december, …</td>
</tr>
</tbody>
</table>
## Extraction performance

<table>
<thead>
<tr>
<th>Method</th>
<th>DS</th>
<th></th>
<th></th>
<th>Twitter</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
<td>F</td>
<td>P</td>
<td>R</td>
<td>F</td>
</tr>
<tr>
<td>MetaMap&lt;sub&gt;_ADR_LEXICON&lt;/sub&gt;</td>
<td>0.470</td>
<td>0.392</td>
<td>0.428</td>
<td>0.394</td>
<td>0.309</td>
<td>0.347</td>
</tr>
<tr>
<td>MetaMap&lt;sub&gt;_SEMANTIC_TYPE&lt;/sub&gt;</td>
<td>0.289</td>
<td>0.484</td>
<td>0.362</td>
<td>0.230</td>
<td>0.403</td>
<td>0.293</td>
</tr>
<tr>
<td>Lexicon-based</td>
<td>0.577</td>
<td>0.724</td>
<td>0.642</td>
<td>0.561</td>
<td>0.610</td>
<td>0.585</td>
</tr>
<tr>
<td>SVM</td>
<td>0.869</td>
<td>0.671</td>
<td>0.760</td>
<td>0.778</td>
<td>0.495</td>
<td>0.605</td>
</tr>
<tr>
<td>ADRMine&lt;sub&gt;_WITHOUT_CLUSTER&lt;/sub&gt;</td>
<td>0.874</td>
<td>0.723</td>
<td>0.791</td>
<td>0.788</td>
<td>0.549</td>
<td>0.647</td>
</tr>
<tr>
<td>ADRMine&lt;sub&gt;_WITH_CLUSTER&lt;/sub&gt;</td>
<td>0.860</td>
<td>0.784</td>
<td>0.821</td>
<td>0.765</td>
<td>0.682</td>
<td>0.721</td>
</tr>
</tbody>
</table>

### CRF Features

<table>
<thead>
<tr>
<th>CRF Features</th>
<th>DS</th>
<th></th>
<th></th>
<th>Twitter</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
<td>F</td>
<td>P</td>
<td>R</td>
<td>F</td>
</tr>
<tr>
<td>All</td>
<td>0.856</td>
<td>0.776</td>
<td>0.814</td>
<td>0.765</td>
<td>0.682</td>
<td>0.721</td>
</tr>
<tr>
<td>All – lexicon</td>
<td>0.852</td>
<td>0.781</td>
<td>0.815</td>
<td>0.765</td>
<td>0.646</td>
<td>0.701</td>
</tr>
<tr>
<td>All – POS</td>
<td>0.853</td>
<td>0.776</td>
<td>0.812</td>
<td>0.754</td>
<td>0.653</td>
<td>0.700</td>
</tr>
<tr>
<td>All – negation</td>
<td>0.854</td>
<td>0.769</td>
<td>0.810</td>
<td>0.752</td>
<td>0.646</td>
<td>0.695*</td>
</tr>
<tr>
<td>All – context</td>
<td>0.811</td>
<td>0.665</td>
<td>0.731*</td>
<td>0.624</td>
<td>0.498</td>
<td>0.554*</td>
</tr>
<tr>
<td>All – cluster</td>
<td>0.851</td>
<td>0.745</td>
<td>0.794*</td>
<td>0.788</td>
<td>0.549</td>
<td>0.647*</td>
</tr>
<tr>
<td>Context + cluster</td>
<td>0.860</td>
<td>0.784</td>
<td>0.821*</td>
<td>0.746</td>
<td>0.628</td>
<td>0.682*</td>
</tr>
</tbody>
</table>
Errors

False Negative ADRs

- Too descriptive/vague-- explained with general words: loved it, except for [not being able to be woken up at night].
- Lack of context (too short phrases or irrelevant context): [Ecstasy] side effects
- Misclassified to indications: I have terrible [pain in joints]
- Annotation guideline: Used to work, [does not anymore]
- Spelling error: ... Started [hallucinating] ... NOT cool !!!
- Idiomatic expressions: didn't work I [pack the fat on] too; My [hair seems to be shedding], ...

False Positive ADRs

- Indications/beneficial effects: He is no longer in pain and [vomiting] all the time.
- Negative modifiers: Finding the right dose is a [nightmare]; Its [annoying] but the benefits are worth it.
- Non-ADR general clinical terms: [Tired] of the side effects; I am very [chemical sensitive].
- Non-ADR symptom descriptions: I have really bad spasms that [keep me up all times] of day and night.
Data, resources and tools

- PSB 2016 shared task: 
  [http://diego.asu.edu/psb2016/task2data.html](http://diego.asu.edu/psb2016/task2data.html)
- Data, tools and resources at: 
  [http://diego.asu.edu/Publications/ADRMine.html](http://diego.asu.edu/Publications/ADRMine.html)
Concept normalization

- Good performance for classification and extraction tasks, but extracted ADRs need to be mapped to standard IDs (or grouped together) to generate reliable signals.

- Good accuracy required for reliable signal generation.

- Example

<table>
<thead>
<tr>
<th>Increase my weight</th>
<th>Weight gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>gain so much weight</td>
<td>(c0043094)</td>
</tr>
<tr>
<td>made me heavy</td>
<td></td>
</tr>
<tr>
<td>gaining weight</td>
<td></td>
</tr>
<tr>
<td>weight gain</td>
<td></td>
</tr>
<tr>
<td>making me fatter</td>
<td></td>
</tr>
</tbody>
</table>
Normalization approach

- **Exact match**
- **Definition match**
- **Semantic match**

**Semantic relatedness kernels**
- PMI
- LSA
- NGD
- Gloss Vector
- Lesnik

**Resources**
- PubMed Dental Journals
- Yahoo Search
- Google Search
- PubMed Clinical Journals
- PubMed Nursing Journals

**MSR+Resource Matcher**
- **Feature Calculator**

**Training/Testing pairs set**
- word1, word2, expected1
- word3, word4, expected2
- ...

**SVMLight**
- **Evaluator**

**Testing Mode**
- word1, word2, feature1, feature2, ..., expected1
- word3, word4, feature1, feature2, ..., expected2
- ...

**Training Mode**
- word1, word2, expected1, predicted1
- word3, word4, expected2, predicted 2
- ...

**SVM model**
## Normalization performance

<table>
<thead>
<tr>
<th>Annotated Phrase</th>
<th>Expected</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>depressed</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>
ADR signal generation

- **Safety signal**— reported information on a causal relationship between an adverse reaction and a drug (WHO)

- The main approach for identifying drug safety signals from reported data (e.g., in FAERS) is to detect the *disproportionality* of reports about a given drug’s adverse events

- Popular methods include:
  - Proportional reporting ratios (PRR)
  - Reporting odds ratios (ROR)
  - Lift
  - Bayesian Confidence Propagation Neural Network (BCPNN)
  - Relative Risk (RR)
## Direct disproportionality measures

### Table II. A 2 × 2 Table for Disproportionality Calculation

<table>
<thead>
<tr>
<th></th>
<th>Reports with ADE $j$</th>
<th>Reports Without ADE $j$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reports with drug $i$</td>
<td>$n_{i,j}$</td>
<td>$n_i - n_{i,j}$</td>
<td>$n_i$</td>
</tr>
<tr>
<td>Reports without drug $i$</td>
<td>$n_j - n_{i,j}$</td>
<td>$n - n_i - n_j + n_{i,j}$</td>
<td>$n - n_i$</td>
</tr>
<tr>
<td>Total</td>
<td>$n_j$</td>
<td>$n - n_j$</td>
<td>$n$</td>
</tr>
</tbody>
</table>

$$P_{RR} = \frac{n_{i,j}/n_i}{(n_j - n_{i,j})/(n - n_i)}.$$  
$$\ln(P_{RR}) \pm 1.96 e^\sqrt{\frac{1}{n_{i,j}} - \frac{1}{n_i} + \frac{1}{n_j - n_{i,j}} - \frac{1}{n - n_i}}.$$  

$$R_{OR} = \frac{n_{i,j}/(n_j - n_{i,j})}{(n_i - n_{i,j})/(n - n_i - n_j + n_{i,j})}.$$  
$$\ln(R_{OR}) \pm 1.96 e^\sqrt{\frac{1}{n_{i,j}} + \frac{1}{n_i - n_{i,j}} + \frac{1}{n_j - n_{i,j}} + \frac{1}{n - n_i - n_j + n_{i,j}}}.$$
- Easy to interpret

- Widely used (e.g., the EMA\(^1\))

- For PRR
  - >= 2 indicator of disproportionality
  - >= two times the upper limit of CI – indicator of disproportionality
  - Validation is performed on a case by case basis

- Very sensitive (particularly for ADRs/drugs with low numbers of reports)
  - .... also why correct normalization is important

- Our approach:
  - Large set of tweets mentioning medications as denominator
Sample PRR charts

- Too big to fit in a slide…
- Fluoxetine (Prozac)
  - SIDER:
    - [http://sideeffects.embl.de/drugs/3386/](http://sideeffects.embl.de/drugs/3386/)
- Adalimumab (Humira)
  - EverydayHealth (no SIDER entry)
    - [http://www.everydayhealth.com/drugs/humira](http://www.everydayhealth.com/drugs/humira)
Some medications may be more suitable for social media based ADR monitoring

Certain age-groups are more likely to be represented in the data

Significant information available regarding
- Tolerability
- Tolerance
- Potential for abuse

Cohorts can be identified over social media and their health-related information tracked
- e.g., pregnant women
Pregnancy cohort monitoring via SM

TW

Querying + Collection of posts

Manual annotation of announcements

Supervised classification of announcements

Timeline collection

DS

User detection

Timeline collection

Medication intake classification

Medication intake annotation

Deriving associations

Health-information analysis

Outcome detection + analysis

Health-information filtering

Timeline collection
Pregnancy announcements

- Queries (16 in total; ~60% true positives)
  - ‘I .* m .* x .* [weeks\|months] .* pregnant’
  - ‘having.*baby’
  - ‘my .* pregnancy’

- Samples
  - I think I could hit this woman in the head with my pregnancy belly and she still wouldn't offer me her seat #ttcproblems #TTC #Toronto
  - Tummy's flat as f*** all day then eat a slice of toast and I'm like 6 months pregnant
  - Everyone in my house is so inconsiderate sometimes. I'm 8 and a half months pregnant, I'm sick and I'm trying to sleep. Shut the f*** up
  - I'm having a baby JB day and it's killing me. I love him so much @justinbieber
  - My sister is five weeks and three days pregnant. I’m going to be an auntie oh my god
  - Girls will be two days pregnant already posting pictures talking bout “I’m getting big.”
Announcement classification + other tasks

- Recycled SVMs with basic features
- ~15k annotations, F-score: 0.88
- Collection of timelines for pregnant women
- Currently cohort is at ~100k; implemented pipeline should collect more
- Pilots studies:
  - Trimester detection
  - Medication mentions at each trimester
- Top 10 medication mentions (sample cohort of 15k)
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 hahahahah 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-related tweets (only ~6k Tweets)
- Low accuracy for classification, but Adderall® pattern matches manual analysis
Contact

- abeed@upenn.edu
- Twitter: @sarkerabeed
- Lab website: https://healthlanguageprocessing.org