RSDD-Time: Temporal Annotation of Self-Reported Mental Health Diagnoses

Sean MacAvaney¹, Bart Desmet²,¹, Arman Cohan¹, Luca Soldaini¹, Andrew Yates³,¹, Ayah Zirikly⁴ and Nazli Goharian¹

1. Georgetown University
2. Ghent University
3. Max Planck Institute for Informatics
4. National Institutes of Health
Motivation

• We use social media data to study the connection between language and mental health

• Various studies use self-reported diagnoses to collect data at scale (De Choudhury et al. 2014, Coppersmith et al. 2015, Yates et al. 2017)

[-] ExampleUser 3 points 28 days ago
In high school, I got diagnosed with depression and anxiety. I’ve come a long way since then, and have been off the meds for almost 2 years now.
Motivation

Temporal aspects of diagnoses have thus far been ignored

- Reported conditions might not be current anymore
  - Labels for current-condition detection might be noisy
- How old is the diagnosis?
  - Study language use as a function of how old a condition is

ExampleUser 3 points 28 days ago

In high school, I got diagnosed with depression and anxiety. I’ve come a long way since then, and have been off the meds for almost 2 years now.
Dataset

RSDD-Time

• 598 Reddit posts (2006-2016) that report a depression diagnosis
• Based on Reddit Self-reported Depression Diagnosis (RSDD) dataset (Yates et al. 2017)
• Developed temporal annotation based on TIMEX3
• Guidelines and data available with data usage agreement
I was diagnosed with depression
Temporal annotation

When I was sixteen, After the divorce, I was diagnosed with depression today 4 months ago
Temporal annotation

When I was sixteen, I was diagnosed with depression.

After the divorce, today 4 months ago

Diagnosis

• Recency
  • very recent (<2 months)
  • >2 months – 1 year
  • >1 year – 3 years
  • >3 years
  • unspecified
  • unspecified but past
Temporal annotation

When I was sixteen,
After the divorce,

I was diagnosed with depression today
4 months ago

Diagnosis

• Condition state
  • current
  • don’t know
  • past
When I was sixteen, After the divorce, I was diagnosed with depression today 4 months ago.

Diagnosis

• Condition state
  • current
  • don’t know
  • past
Temporal annotation

When I was sixteen, after the divorce, I was diagnosed with depression today. 4 months ago.

Diagnosis

- Condition state
  - current
  - don’t know
  - past
Temporal annotation

When I was sixteen, After the divorce, today

I was diagnosed with depression 4 months ago

but I am all better now

Diagnosis

• Condition state
  • current
  • don’t know
  • past
Temporal annotation

When I was sixteen, I was diagnosed with depression. After the divorce, today 4 months ago

Time expressions

• Help to determine diagnosis recency
Temporal annotation

When I was sixteen, I was diagnosed with depression today.

After the divorce, 4 months ago

Time expressions

• Help to determine diagnosis recency

• Inferencing:
  • explicit
Temporal annotation

When I was sixteen, I was diagnosed with depression today.
After the divorce, 4 months ago.

Time expressions

• Help to determine diagnosis recency
• Inferencing:
  • explicit
  • inferable from age of the poster
Corpus annotation

• Double annotation by 6 trained annotators
• Disagreements are resolved into a consensus annotation
• Inter-annotator agreement moderate to perfect

<table>
<thead>
<tr>
<th>span</th>
<th>attribute</th>
<th>%</th>
<th>( \kappa )</th>
</tr>
</thead>
<tbody>
<tr>
<td>diagnosis</td>
<td>condition state</td>
<td>0.68</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>diagnosis recency</td>
<td>0.66</td>
<td>0.64</td>
</tr>
<tr>
<td>time</td>
<td>explicit</td>
<td>0.91</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>inferable from age</td>
<td>0.93</td>
<td>0.82</td>
</tr>
</tbody>
</table>
Corpus analysis
Corpus analysis

- Current
- Probably current
- Don't know
- Probably past
- Past

- Unspecified
- Unspecified-past
- <2m
- 2m-1y
- 1y-3y
- >3y
Baseline experiments

- Classifying the **recency** and **condition state** of a given diagnosis
- Features: tf-idf weighted bag-of-character ngrams of length 2-5
Baseline experiments: classification

<table>
<thead>
<tr>
<th>Model</th>
<th>Diagnosis Recency</th>
<th>Condition State</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
</tr>
<tr>
<td>Logistic Reg.</td>
<td>0.47</td>
<td>0.35</td>
</tr>
<tr>
<td>Linear SVM</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td>Gradient-Boosted Trees</td>
<td><strong>0.56</strong></td>
<td>0.42</td>
</tr>
</tbody>
</table>
Baseline experiments: time expressions detection

Detecting temporal expressions that bear on the diagnosis time using off-the-shelf SUTime library (Chang & Manning 2012)
Baseline experiments: time expressions detection

Detecting **temporal expressions** that bear on the diagnosis time using off-the-shelf SUTime library (Chang & Manning 2012)

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>R</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUTime</td>
<td>0.17</td>
<td>0.59</td>
<td>0.26</td>
</tr>
</tbody>
</table>
Detecting temporal expressions that bear on the diagnosis time using off-the-shelf SUTime library (Chang & Manning 2012)

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>R</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUTime</td>
<td>0.17</td>
<td>0.59</td>
<td>0.26</td>
</tr>
<tr>
<td>+ age rules</td>
<td>0.20</td>
<td>0.81</td>
<td>0.32</td>
</tr>
</tbody>
</table>
Baseline experiments: time expressions detection

Detecting temporal expressions that bear on the diagnosis time using off-the-shelf SUTime library (Chang & Manning 2012)

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>R</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUTime</td>
<td>0.17</td>
<td>0.59</td>
<td>0.26</td>
</tr>
<tr>
<td>+ age rules</td>
<td>0.20</td>
<td>0.81</td>
<td>0.32</td>
</tr>
<tr>
<td>+ closest heuristic</td>
<td>0.33</td>
<td>0.51</td>
<td>0.40</td>
</tr>
</tbody>
</table>
Baseline experiments: time expressions detection

 Detecting **temporal expressions** that bear on the diagnosis time using off-the-shelf SUTime library (Chang & Manning 2012)

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>R</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUTime</td>
<td>0.17</td>
<td>0.59</td>
<td>0.26</td>
</tr>
<tr>
<td>+ age rules</td>
<td>0.20</td>
<td>0.81</td>
<td>0.32</td>
</tr>
<tr>
<td>+ closest heuristic</td>
<td>0.33</td>
<td>0.51</td>
<td>0.40</td>
</tr>
<tr>
<td>+ closest heuristic + age rules</td>
<td><strong>0.44</strong></td>
<td><strong>0.69</strong></td>
<td><strong>0.53</strong></td>
</tr>
</tbody>
</table>
Conclusion

• Considerable proportion of self-reported diagnoses may not be current anymore
• Dataset can be used to
  • construct better self-reported diagnosis datasets
  • study language change as a function of diagnosis age
• Benchmarks for future work

http://ir.cs.georgetown.edu/resources