Please sign up for the class presentations!!!!!

Mid-term progress presentations due on Mar 16
Predicting Depression via Social Media
Summary

- Can social media activities and connectedness predict risk to major depressive disorder?
- Recruitment of a sample of Twitter users through a survey methodology over Amazon’s Mechanical Turk
  - ~40% provided access to Twitter data
Summary

• Social engagement
• “Insomnia index” – mean z-score of an individual’s volume of Twitter activity per hour
• Ego-centric social graph – nodal properties (inlinks, outlinks); dyadic properties (reciprocity, interpersonal exchange); neighborhood properties (density, clustering coefficient, two-hop neighborhood, embeddedness, number of ego components)
• Language
  • Depression lexicon – top uni- and bigrams compiled from Yahoo! Answers category on mental health
  • Linguistic style
Summary

Non-depression class

Depression class

\[ f(x) = 0.00070393x^2 + -0.015183x + 0.00 \]
\[ g(x) = -6.9926e-05x^2 + 0.0045778x + 0.00 \]

\[ f(x) = 0.000010968x + 1.5931 \]
\[ g(x) = 0.0013533x + 1.9698 \]

\[ f(x) = -0.00015787x + 0.24889 \]
\[ g(x) = 0.00014436x + 0.3314 \]

\[ f(x) = 4.8122e-06x + 0.003255 \]
\[ g(x) = 1.5179e-06x + 0.0012 \]

\[ f(x) = 8.2818e-05x + 0.043383 \]
\[ g(x) = 1.986e-05x + 0.0033 \]
Summary

<table>
<thead>
<tr>
<th>Egonetwork measures</th>
<th>Depres. class</th>
<th>Non-depres. class</th>
</tr>
</thead>
<tbody>
<tr>
<td>#followers/inlinks</td>
<td>26.9 (σ=78.3)</td>
<td>45.32 (σ=90.74)</td>
</tr>
<tr>
<td>#followees/outlinks</td>
<td>19.2 (σ=52.4)</td>
<td>40.06 (σ=63.25)</td>
</tr>
<tr>
<td>Reciprocity</td>
<td>0.77 (σ=0.09)</td>
<td>1.364 (σ=0.186)</td>
</tr>
<tr>
<td>Prestige ratio</td>
<td>0.98 (σ=0.13)</td>
<td>0.613 (σ=0.277)</td>
</tr>
<tr>
<td>Graph density</td>
<td>0.01 (σ=0.03)</td>
<td>0.019 (σ=0.051)</td>
</tr>
<tr>
<td>Clustering coefficient</td>
<td>0.02 (σ=0.05)</td>
<td>0.011 (σ=0.072)</td>
</tr>
<tr>
<td>2-hop neighborhood</td>
<td>104 (σ=82.42)</td>
<td>198.4 (σ=110.3)</td>
</tr>
<tr>
<td>Embeddedness</td>
<td>0.38 (σ=0.14)</td>
<td>0.226 (σ=0.192)</td>
</tr>
<tr>
<td>#ego components</td>
<td>15.3 (σ=3.25)</td>
<td>7.851 (σ=6.294)</td>
</tr>
</tbody>
</table>
Quantifying Mental Health Signals on Twitter
Summary

- Overcome challenges of solicited survey data – use self-identified mental health diagnoses on Twitter for obtaining ground truth.

![Summary Diagram]
Summary

Figure 2: ROC curves for separating diagnosed from control users, compared across disorders: bipolar in red, depression in blue, PTSD in purple, SAD in orange. The precision (diagnosed, correctly labeled) for each disorder at false alarm (control, labeled as diagnosed) rates of 10% and 20% are shown to the right of the ROC curve. Chance performance is indicated by the dotted black line.

Figure 3: ROC curves of performance of individual analytics for each disorder: LIWC in blue, pattern of life in yellow, CLM in red, ULM in green, all in black. Chance performance is indicated by the dotted black line.

The receiver operating characteristic (ROC) curves in Figures 2 and 3 demonstrate performance of the various classifiers at the task of separating diagnosed from control groups. In all cases, the correct detections (or hits) are on the y-axis and the false detections (or false alarms) are on the x-axis. Figure 2 compares performance across diagnoses, one line per disorder. Figure 3 shows one plot per mental health condition, with the performance of the various analytics, individually and in concert as individual ROC curves. A few trends emerge – 1) All analytics show some ability to separate the classes, indicating they are finding useful signals. 2) The LMs provide superior performance to the other analytics, indicating there are more signals present in the language than are captured by LIWC and pattern-of-life analytics. For readability we do not show the performance of all combinations of analytics, but they perform as expected: any set of them perform equal to or better than their individual components. Taken together, this indicates that there is information relevant to separating diagnosed users from controls in all the analytics discussed here. Furthermore, this highlights that there remains significant signals to be uncovered and understood in the language of social media. These trends also allow us to compare the disorders as manifest in language usage, though this...
What are the differences you observed between the two studies? What are the strengths and limitations of those differences?
A consistent challenge in many prediction tasks like these, is gathering gold standard information (or ground truth). What could be different ways to get at this problem?
In many ways, predictive models are never 100% perfect. How can social media platforms leverage the predictive methodologies outlined in the two papers?
Both papers used Twitter as the data source of study – in ways it is a nice platforms where an individual can make their profile however they wish it to be. Would the same findings hold on a platform that enforces real identities, like Facebook?
Depression is not an online condition, but one that spans both the online and the offline life. The papers do not take offline attributes into their models.

Is there a way to that into account? What would be the most significant offline attributes to consider?
Anurag brings up the concern of “awareness contamination”. To what extent do you think this would impact a predictive model like the one proposed in the papers going forward? How would you combat that?
The analyses in Coppersmith et al. indicate that depression, PTSD, and bipolar disorder have high mutual correlation, but correlation is low between SAD and the others. Why do you think that is the case – is there anything in the model or the data that account for it?
In the first paper, the ground truth was obtained from Amazon mechanical turk workers. Anything odd or specific about this population that may have affected the findings? What would be alternative ways of recruiting people?
Coppersmith et al shows how $n$-gram language models can provide better performance over LIWC features. What is the downside of using language models though?