Communication in Online Support Groups

Alison Berkowitz\textsuperscript{1}, Alexis Larry\textsuperscript{2}, Venessa Roshni Tauro\textsuperscript{3}, Stephanie Yang\textsuperscript{4}, Jessica Yang\textsuperscript{5}

\textsuperscript{1} - Olin College; \textsuperscript{2,4} - CMU; \textsuperscript{3} - IIIT; \textsuperscript{5} - Yale
Why do we care?

- Social support is correlated with positive individual outcomes
- 80% of patients experience significant psychological distress during initial treatment
- Support groups can improve psychological well-being; decreasing depression and reducing stress
Types of Social Support

Informational
- Information and advice about the disease and treatment
- ...call and ask if you can put ice on your port site before you go in

Emotional
- Expressions of concern, sympathy and reassurance
- ...I am truly inspired by you. Big cyber hugs and good wishes to you :>
140k members
- Users differ in their propensity to give different types of support
- No correlation between likelihood of giving emotional/informational support
it's been a very long few months.....
Now waiting for oncologist appt to be scheduled... Sometimes it can be very overwhelming!

...I'm tired and because I cannot get out as much anymore this can be very lonely. Are there others there out there to get virtually no sleep and have absolutely no 1 to speak to at night!? The nights are the worst!

Hi honey, I am so sorry for how your feeling Paulette. Yes, it can be very overwhelming. It seems nothing ends. And, during this time it does get lonely at times. That is when I turned to here for support from the ladies. And, Facebook. Ha. And, I turned to God for comfort. It helped alot. And, still does. Honey, all these emotions arise, which are normal. As time goes on, and, you get closer to the end of your treatments, things will start getting better for you. Remember, to be strong thru this, and, stay positive. Feel free to ask questions. I am here.
Research question

How can we identify valuable members early on, based on their interaction on the forum during their first month?
Methodology

- Gathered posting statistics & data
  - Users active for at least 110 days
  - Initial: posts by user in the first month, replies to those posts in the first month
  - Future: amount of emotional support in posts after the first month
Machine Learning

- tool: LightSIDE

Generated key features

Example on ‘praying’

- 63% accuracy
- Kappa = 0.094

<table>
<thead>
<tr>
<th>Actual \ Predicted</th>
<th>Low support</th>
<th>High support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low support</td>
<td>981</td>
<td>27</td>
</tr>
<tr>
<td>High support</td>
<td>598</td>
<td>72</td>
</tr>
</tbody>
</table>
The data set was divided into two parts randomly.

- 90% for training and development
- 10% for testing
### Results

<table>
<thead>
<tr>
<th>Feature Set</th>
<th>Accuracy</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline (no features)</td>
<td>60%</td>
<td>0</td>
</tr>
<tr>
<td>unigrams (5000 features)</td>
<td>65%</td>
<td>0.265</td>
</tr>
<tr>
<td>determination-words</td>
<td>63%</td>
<td>0.199</td>
</tr>
<tr>
<td>get-through-it-words</td>
<td>62%</td>
<td>0.1</td>
</tr>
<tr>
<td>prayer-words</td>
<td>66%</td>
<td>0.245</td>
</tr>
<tr>
<td>strength-words</td>
<td>62%</td>
<td>0.124</td>
</tr>
<tr>
<td>support-words</td>
<td>61%</td>
<td>0.175</td>
</tr>
<tr>
<td>Description</td>
<td>Accuracy</td>
<td>Kappa</td>
</tr>
<tr>
<td>----------------------------------------------------------------------------</td>
<td>----------</td>
<td>-------</td>
</tr>
<tr>
<td>sum of language features</td>
<td>66%</td>
<td>0.251</td>
</tr>
<tr>
<td># replies per post, emotional support in replies</td>
<td>64%</td>
<td>0.162</td>
</tr>
<tr>
<td>sum of features + # replies per post + emotional support in replies</td>
<td>67%</td>
<td>0.269</td>
</tr>
</tbody>
</table>
Most significant feature

- sum of all word-class features
- number of replies to each post
- emotional support provided in replies

Accuracy: 71.6%
Kappa: 0.38

Performance comparable to unigram model with 5000-6000 features!
Counterintuitive results

- Kappa = 0:
  - Medical terms
  - Family relations
  - Collective pronouns (we, our, us)
  - Length of post
Conclusions

The language used and responses to a user’s initial posts can indicate how much emotional support she will later contribute to the online community.
Next steps

Larger dataset e.g. including thread titles

More features, that would be better predictors
Learning points

- Learned a bit about ML
- We should apply to grad school
Thanks

Bob Kraut
Carolyn Rose
Miaomiao Wen
YiChia Wang
OurCS Staff
Works Cited

http://community.breastcancer.org/forum/6/topic/812825