#Datagrant project: presentation UT
text analysis café
Dong Nguyen & Tijs van den Broek

2-6-2016
February 4, 2014

Twitter announces to allow a handful of researchers access to its historical data
Cancer Early Detection Campaigns on Twitter

<table>
<thead>
<tr>
<th>Annual Cancer Deaths</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Per Year</td>
<td>6.202.080</td>
</tr>
<tr>
<td>Per Month</td>
<td>516.840</td>
</tr>
<tr>
<td>Per Day</td>
<td>16.992</td>
</tr>
<tr>
<td>Per Hour</td>
<td>708</td>
</tr>
<tr>
<td>Per Minute</td>
<td>12</td>
</tr>
</tbody>
</table>
Cancer Early Detection Campaigns on Twitter

<table>
<thead>
<tr>
<th>Annual Cancer Deaths</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Per Year</td>
<td>6.202.080</td>
</tr>
<tr>
<td>Per Month</td>
<td>516.840</td>
</tr>
<tr>
<td>Per Day</td>
<td>16.992</td>
</tr>
<tr>
<td>Per Hour</td>
<td>708</td>
</tr>
<tr>
<td>Per Minute</td>
<td>12</td>
</tr>
</tbody>
</table>

One third of these deaths could have been prevented
How Cancer Deaths can be Prevented

1. Doing research
2. Regular screening and early diagnosis
3. Vaccination
4. Lifestyle
How Cancer Deaths can be Prevented

1. Doing research
2. Regular screening and early diagnosis
3. Vaccination
4. Lifestyle
How Cancer Deaths can be Prevented

1. Doing research
2. Regular screening and early diagnosis
3. Vaccination
4. Lifestyle
How Cancer Deaths can be Prevented

1. Doing research
2. Regular screening and early diagnosis
3. Vaccination
4. Lifestyle
How Cancer Deaths can be Prevented

1. Doing research
2. Regular screening and early diagnosis
3. Vaccination
4. Lifestyle

[SunSmart logo]
The data grant proposal aims to study the diffusion process and effectiveness of cancer early detection campaigns on Twitter.
Twitter #DataGrant

Congratulations @UTwente with @Twitter #DataGrants selection!
blog.twitter.com/2014/twitter-d...
Results

Classifying countries
Classifying Countries

Classifying motivations

Classifying countries

Role of celebrities

Network structure vs Altruism

Reactions?

Role of identity
Results

Classifying motivations
Classifying Motivations
Results

Role of celebrities
Role of Celebrities
Results

Network structure vs Prosocial behavior
Network Structure vs. Altruism
Results

Role of identity
Role of Identity

Classifying motivations
Classifying countries
Role of celebrities
Network structure vs Altruism
Reactions?
Role of identity
Work in progress

- Impact of tweets and offline fundraising events on fundraising performance
- Multilevel study in US counties (linking data about ratio prostate and testicular cancer)
- Motivation of campaign members over time
- The effectiveness of Movember teams and their leaders
Identifying Motivations to Participate in Online Health Campaigns

Knowing individual motivations helps to explain the amount of campaign donations raised by participants.
Social Identity Model of Collective Action (van Zomeren et al., 2008)

• **Injustice**: A shared emotion that includes both affective (e.g., anger) and cognitive perceptions (ideology) of an unfair situation
Social Identity Model of Collective Action (van Zomeren et al., 2008)

• **Injustice**: A shared emotion that includes both affective (e.g., anger) and cognitive perceptions (ideology) of an unfair situation

  “I had testicular cancer”

  “my dad”

  “because men’s health is important to me”
Social Identity Model of Collective Action (van Zomeren et al., 2008)

- **Injustice**: A shared emotion that includes both affective (e.g., anger) and cognitive perceptions (ideology) of an unfair situation

- **Social identity**: A sense of belonging together that emerges out of common attributes, experiences and external labels
Social Identity Model of Collective Action (van Zomeren et al., 2008)

• **Injustice**: A shared emotion that includes both affective (e.g., anger) and cognitive perceptions (ideology) of an unfair situation

• **Social identity**: A sense of belonging together that emerges out of common attributes, experiences and external labels

“my friends asked me again to join them”

“a great excuse to grow a stache”
Social Identity Model of Collective Action (van Zomeren et al., 2008)

- **Injustice**: A shared emotion that includes both affective (e.g., anger) and cognitive perceptions (ideology) of an unfair situation

- **Social identity**: A sense of belonging together that emerges out of common attributes, experiences and external labels

- **Collective efficacy**: The shared belief that ones group is capable of resolving its grievances through a campaign
Social Identity Model of Collective Action (van Zomeren et al., 2008)

• **Injustice**: A shared emotion that includes both affective (e.g., anger) and cognitive perceptions (ideology) of an unfair situation

• **Social identity**: A sense of belonging together that emerges out of common attributes, experiences and external labels

• **Collective efficacy**: The shared belief that one's group is capable of resolving its grievances through a campaign

  “this campaign can make a difference!”
Nick Wheeler

My Motivation

Because I have just turned 50 and have become aware how many people this disease affects. Selling shirts is not enough. It is more important to beat this terrible cancer.

- Nick Wheeler
Linking profiles

In total 5,519 users linked. 2,108 were manually annotated for their motivation.
## Dataset statistics

<table>
<thead>
<tr>
<th></th>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td># Participants</td>
<td>1,494</td>
<td>614</td>
</tr>
<tr>
<td>% US / UK</td>
<td>54.8/45.2</td>
<td>53.3/46.7</td>
</tr>
<tr>
<td>% Injustice</td>
<td>37.6</td>
<td>40.2</td>
</tr>
<tr>
<td>% Social identity</td>
<td>48.7</td>
<td>46.9</td>
</tr>
<tr>
<td>% Collective efficacy</td>
<td>36.1</td>
<td>35.0</td>
</tr>
</tbody>
</table>
Annotation agreement

<table>
<thead>
<tr>
<th></th>
<th>Cohen’s Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Injustice</td>
<td>0.71</td>
</tr>
<tr>
<td>Social identity</td>
<td>0.67</td>
</tr>
<tr>
<td>Collective efficacy</td>
<td>0.47</td>
</tr>
</tbody>
</table>

Based on 200 double annotations
Automatic classification of Movember profiles

- Logistic Regression
- Unigrams, bigrams, topics, text length, country

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Injustice</td>
<td>0.816</td>
</tr>
<tr>
<td>Social Identity</td>
<td>0.788</td>
</tr>
<tr>
<td>Collective efficacy</td>
<td>0.627</td>
</tr>
</tbody>
</table>

Final system
Feature analysis

<table>
<thead>
<tr>
<th>Injustice</th>
<th>Social Identity</th>
<th>Collective Efficacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDA topic&lt;sup&gt;a&lt;/sup&gt;</td>
<td>fun</td>
<td>LDA topic&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>cancer</td>
<td>team</td>
<td>beat</td>
</tr>
<tr>
<td>friend</td>
<td>moustache</td>
<td>and family</td>
</tr>
<tr>
<td>lost</td>
<td>mo</td>
<td>change</td>
</tr>
<tr>
<td>father</td>
<td>grow</td>
<td>yourself</td>
</tr>
<tr>
<td>had</td>
<td>mustache</td>
<td>all of</td>
</tr>
<tr>
<td>survivor</td>
<td>LDA topic&lt;sup&gt;c&lt;/sup&gt;</td>
<td>awareness</td>
</tr>
<tr>
<td>prostrate</td>
<td>fuzz</td>
<td>for movember</td>
</tr>
<tr>
<td>for my</td>
<td>movement</td>
<td>awareness of</td>
</tr>
<tr>
<td>my</td>
<td>look</td>
<td>last</td>
</tr>
</tbody>
</table>

Table 4: Top-weighted features for free-text motivation experiments.

<sup>a</sup>topic about family/friends who had cancer
<sup>b</sup>topic about raising funds for research
<sup>c</sup>topic about the Movember campaign
Automatic classification of Twitter profiles

- Logistic regression
- Unigrams, bigrams, topics, text length, country, behavior, urls, user mentions,..

Low performance, why?
- Few tweets per user
- Nature of Twitter influences content

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Injustice</td>
<td>0.458</td>
</tr>
<tr>
<td>Social Identity</td>
<td>0.531</td>
</tr>
<tr>
<td>Collective efficacy</td>
<td>0.399</td>
</tr>
</tbody>
</table>

**Final system**
Findings

- Campaign participants with an **injustice** motivation raise significantly ($p < 0.001$) more money.
- Participants that are part of a **team** raise significantly more money ($p < 0.001$).
- Participants with a **social identity** motivation are more often part of a team.

<table>
<thead>
<tr>
<th></th>
<th>Injustice ($)</th>
<th>Identity ($)</th>
<th>Efficacy ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK</td>
<td>203.74</td>
<td>128.36</td>
<td>123.39</td>
</tr>
<tr>
<td>US</td>
<td>234.47</td>
<td>156.07</td>
<td>169.03</td>
</tr>
</tbody>
</table>

$n=90,484$
Summary

• Explored machine learning methods to automatically identify the motivations of Movember participants

• We found a strong link between motivations and donations, and motivations and team membership
Thank you!