Using Comments to Improve Services: Text mining Electronic Health Records

Harold Baize, PhD.
San Francisco Department of Public Health
Behavioral Health Services
Text as Data

• Behavioral Health generates and collects vast amounts of text:
  Surveys
  Assessments
  Progress Notes

• Any text field in an Electronic Health Record is a source of data for analysis.

• Recent advances in data science have given us powerful tools to derive useful information from text.
Satisfaction Survey Comments

- Satisfaction surveys typically have one or more text boxes that invite written comments.

- Usually the instructions ask for feedback about services.

- Although not explicitly stated, there is an implied promise that this information will be shared with service providers, if not it could not be “feedback.”

- Sometimes the instructions for comments address specific questions such as how to improve services, or what was helpful about services (YSS / YSS-F).

- Respondent often write far more than the space allows, in margins and the back of the page.

- Additional comments are made to explain Likert item responses.
Steps in the process

• Check comments immediately for potential threat to self/other
• Enter English comments with program code
• Scan and Translate non-English languages
• De-identify comments
• Data cleaning and spell check { 😊 ⭐️❤️ } 
• Apply text analyses
• Return comments and analysis to programs
• Engage program staff in a process of data reflection
Treatment Perceptions Survey Comments - Fall 2017
There were 623 surveys with comments, with a total of 8,844 words.
Languages: English 611, Spanish 12
Applied Comment Text Analysis

• Sentiment Analysis
  Relative proportion of negative or positive evaluative words

• Emotive content
  Relative proportion of eight basic emotions:
  Joy Trust Anticipation Surprise Anger Disgust Fear Sadness

• Bag of words (with negation)

• Lexicon based
Lexicons

- The lexicons available for text analysis are not specific to behavioral health.
  NRC EmoLex 14,182 words

- Most were created to analyze tweets.

- If you search for behavioral health lexicons what you find is not an evaluative dictionary but a glossary.

- To realize the potential of behavioral health text analysis we will need to develop context specific lexicons

- Cheng, Ramos, Bitsch, Jonas, Ix, See, Wehrle (2016) Psychologist in a Pocket: Lexicon Development and Content Validation of a Mobile-Based App for Depression Screening

<table>
<thead>
<tr>
<th>Word</th>
<th>Sent</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>anger</td>
<td>-3</td>
<td>3</td>
</tr>
<tr>
<td>best</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>frustration</td>
<td>-2</td>
<td>1</td>
</tr>
<tr>
<td>help</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>improved</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>mad</td>
<td>-3</td>
<td>1</td>
</tr>
<tr>
<td>prison</td>
<td>-2</td>
<td>1</td>
</tr>
<tr>
<td>ruin</td>
<td>-2</td>
<td>1</td>
</tr>
<tr>
<td>selfish</td>
<td>-3</td>
<td>1</td>
</tr>
<tr>
<td>support</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>upset</td>
<td>-2</td>
<td>1</td>
</tr>
</tbody>
</table>

Sentiment Score: -14
Positive Sentiment of Comments

- Proportion of evaluative words with positive meaning (yellow): 76.8%
- Proportion of evaluative words with negative meaning (gray): 23.2%

Emotion Content of Comments

- Joy: 21.4%
- Trust: 29.3%
- Anticipation: 16.8%
- Surprise: 7.6%
- Anger: 5.1%
- Disgust: 3.9%
- Fear: 9.0%
- Sadness: 6.9%

Proportion of words with emotion content that fall within each category
The {name} methadone clinic staff are very helpful and professional. They are the very best and nicest people that I have ever worked with for any type treatment services.

Love this place!

I am very **grateful** for your janitors keeping the building clean. Thank you

Hell, things are good everyone can come on out!!

{name} is the best worker and inspiring to the staff is very supportive. {name} and the others very professional help me with my life.

NO complaints. Great staff.
Data Reflection

- Reports are not posted publicly
- Programs are encouraged to review comments
- Comments do result in change
- Programs are encouraged to engage in a process of data reflection with staff
- Provider meeting survey

“Yes. We have taken those comments to the all staff meeting to address it and work on the issues that we need to improve.”
R is a free and open source programming environment for analysis and data science.

It includes many packages for text mining:
- openNLP
- tm
- tidytext
- Syuzhet

https://www.tidytextmining.com/
https://www.youtube.com/watch?v=oP3c1h8v2ZQ
A Portrait of the Artist as a Young Man
by James Joyce

Then the season of pleasure came to an end

He saw clearly too his own futile isolation

“Give me a kiss, she said”

Shoved in the Ditch
Father Arnel
Christmas Break
Uncertainty

Oh, Hell!
Further Text Mining Clinical Records

- Could be applied to the therapeutic trajectory within episode using progress notes.
- Limited by existing lexicons and lack of standardization
- Need for Behavioral Health specific lexicons
- Standard format for progress notes and assessments
A very negative progress note

THIS NOTE IS TO SERVE AS 8 MONTH PSYCHIATRIC REASSESSMENT OF PATIENT

General: (x) none; ( ) Weight Change, ( ) Fatigue, ( ) Weakness, ( ) Fever, ( ) Chills,

Skin (skin, hair, nail changes): (x) none; ( ) Itching, ( ) Rashes, ( ) Sores, ( ) Lumps

Head: (x) none; ( ) Head injury; ( ) Headache location and frequency

Eyes: (x) none; ( ) Glasses, ( ) Contact lenses, ( ) visual change

Ears: (x) none; ( ) Hearing loss, ( ) tinnitus, ( ) dizziness, ( ) earache

Nose/Sinuses: (x) none; ( ) running nose, ( ) stuffiness, ( ) nose bleeding

Mouth/throat/neck: (x) none; ( ) bleeds gums, ( ) sore throat,

Breasts: (x) none; ( ) skin changes, ( ) mass/lumps, ( ) pain, ( ) discharge, ( ) self exams

Cardiac: (x) none; ( ) High blood pressure: , ( ) abnormal heart sound, ( ) chest pain,

( ) irregular heartbeat, ( ) short of breath while walking, ( ) dizziness:,

( ) short of breath at night time, ( ) last ECG

Respiratory: (x) none; ( ) short of breath, ( ) chest tightness, ( ) cough,

( ) cough with blood, ( ) asthma,

GI: (x) none; ( ) appetite, ( ) nausea, ( ) vomiting, ( ) constipation,

( ) diarrhea, ( ) abdominal pain

Urinary: (x) none; ( ) Urine change, ( ) incontinence
Topic modeling

Intertopic Distance Map (via multidimensional scaling)

Top-30 Most Relevant Terms for Topic 39 (2.1% of tokens)

Marginal topic distribution