Automatically Identifying Patient Needs from Online Health Communities

Mark Greenwood
Patient Needs

- no strict definition
- uses
  - practice
  - research
Online Health Discussion

• social networking platforms
  – blogs
  – microblogs
  – forums
  – Facebook, etc...

• patients sharing
  – information
  – support
  – experiences
Online Health Discussion
Who’s taking part?

<table>
<thead>
<tr>
<th>Rank</th>
<th>Condition</th>
<th>Authors</th>
<th>Prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cardiovascular Disease</td>
<td>65k</td>
<td>81m</td>
</tr>
<tr>
<td>2</td>
<td>Arthritis</td>
<td>38k</td>
<td>52m</td>
</tr>
<tr>
<td>3</td>
<td>Asthma</td>
<td>45k</td>
<td>24m</td>
</tr>
<tr>
<td>4</td>
<td>COPD</td>
<td>12k</td>
<td>24m</td>
</tr>
<tr>
<td>5</td>
<td>Type 2 Diabetes</td>
<td>24k</td>
<td>16m</td>
</tr>
</tbody>
</table>
Potential

• self-reported experiences
  – unprompted

• healthcare professionals and researchers
  – patient experience
  – patient satisfaction
  – patient needs
Research Question

Can patient blogs be used to discover the needs of the COPD patient population?

- COPD
  - Chronic Obstructive Pulmonary Disease
  - reduction of airflow in the lungs
  - not fully reversible
An example...

“My husband doesn’t drive so if I have an exacerbation it would mean phoning for an Ambulance, the steroids I’m on are a five day course what I need to know is how often can I expect an exacerbation?”
Challenges

• large
  – expensive to manually process

• unstructured

• noisy
  – spelling/typography errors
  – shorthand
  – etc...
Text Mining

• automatic knowledge discovery from unstructured text

• high throughput

• extract trends
General overview

Approach overview

Blog post collection

Documents

Pre-processing

term feature extraction

sentiment feature extraction

Mining

trend analysis

‘need’ classification

Patient concerns

Patient needs

'need' classification
Data collection

- **hand-picked blogs**
  - 12 COPD blogs
  - ~370 posts
  - ~8k sentences
  - 165k words
- **blog search**
  - wider community
  - noise filter
- **annotation**
  - training/test sets
Pre-processing step

- term feature extraction
- sentiment feature extraction
Pre-processing: Term Extraction

- **MetaMap** – National Library of Medicine (US)
  - identifies medical terms in text
  - dictionary-based
  - inc. semantic network

“I was diagnosed with chronic obstructive pulmonary disease (also known as COPD)…”

<table>
<thead>
<tr>
<th>Term</th>
<th>Concept</th>
<th>Semantic Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronic obstructive pulmonary disease</td>
<td>C0024117 Chronic obstructive Airway Disease</td>
<td>T047 Disease or Syndrome</td>
</tr>
<tr>
<td>COPD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnosed</td>
<td>C0011900 Diagnosis</td>
<td>T033 Finding</td>
</tr>
</tbody>
</table>
Pre-processing: Sentiment Mining

- emotional polarity
- emotional taxonomy

Positive
- ‘Love’

Negative
- ‘Hate’

Neutral
- ‘Supervised’

Positive emotion

Love

Emotion

Positive-expectation

Anticipation

‘adorably’

‘lovingly’

‘fondly’
Mining data

- trend analysis
- ‘need’ classification
Progress: Trend Analysis

• ‘relatedness’

• probabilistic measure
  – pointwise mutual information

\[
PMI(x, y) = \frac{P(x, y)}{P(x)P(y)}
\]
Progress: Classifying ‘needs’

• formally defining need
  – literature review

• final test/training data needed
  – manual annotation
  – crowd-sourcing
Next Steps...

• annotation
  – final training/test sets

• ‘need’ classifier

• evaluation
While trying to prepare myself for the possibility of having a double lung transplant, I was told that one of the requirements for lung recipients at the Toronto General was to be physically fit with a very strong heart. They would expect me to be on a strict exercise program, I was aghast, how could I withstand the rigors of exercise when I couldn't walk from room to room in my home without being breathless?