Closing in on open-ended patient questionnaires with text mining

Prof. Irena Spasić
i.spasic@cs.cardiff.ac.uk

Cardiff School of Computer Science & Informatics
http://www.cs.cf.ac.uk

Team

- Dr Kate BUTTON Healthcare Sciences
- Prof. Andy SMITH Psychology
- Mr David OWEN Computer Science
Outline

- measurement of self-care
  - closed-ended vs. open-ended questionnaire
  - quantitative vs. qualitative analysis
- qualitative analysis based on text mining
  - ordinal text classification against a fixed, discrete rating scale, e.g. not at all < somewhat < a lot
  - sentiment analysis, e.g. negative < neutral < positive
  - named entity recognition
- results and next steps
Self–care

- all actions taken by people to recognise, treat and manage their own health
- independently or in partnership with the healthcare system

"We need to fully understand ... what support people need for good self-care." NHS

Knee injury & Osteoarthritis Outcome Score

- an instrument for assessing patients' opinion about their knee and associated problems
- 42–item self-administered questionnaire http://www.koos.nu/
- ... to assess 5 outcomes: (1) pain, (2) other symptoms, (3) daily living, (4) sport & recreation, (5) quality of life

**KOOS**

- extensively used for clinical/research purposes
- used to monitor groups & individuals over time
- intended to be used short–term
  - to assess changes from week to week induced by treatment
- ... or long–term
  - over years following a primary injury or OA

**Pros & cons**

- **pros**
  - user friendly
  - relatively quick to complete (approx. 10min)

- **cons**
  - closed–ended questions
  - answers limited to few options provided
  - 5–point Likert scale
Limitations

- example

Q2. Have you modified your lifestyle to avoid potentially damaging activities to your knee?

- Not at all
- Mildly
- Moderately
- Severely
- Totally

Q3. How much are you troubled with lack of confidence in your knee?

- Not at all
- Mildly
- Moderately
- Severely
- Extremely

- we get answers to what?
- ... but not to how? and why?

Limitations

- very limited insight into patient's opinion

- quantitative vs. qualitative research
Qualitative vs. quantitative analysis

- data collection
  - interviews
  - focus groups
  - questionnaires
- data analysis
  - content analysis
  - thematic analysis
- cons
  - sampling bias
  - low statistical power

Qualitative research
Text analytics

- emerged as a potential solution to the problems of
  - information overload
  - information overlook
- uses techniques from
  - natural language processing
  - knowledge management
  - data mining

to efficiently process large text collections

Our approach

- qualitative analysis... on steroids
- take advantage of modern technology to facilitate remote interaction with patients on a regular basis
- ... something traditional focus groups and interviews cannot provide easily
- patients report on their experience online with regular e-mail prompts
- ... then crunch the data using text analytics
Data collection

- KLOG – a self-administered open-ended questionnaire
  - 10 questions
  - based on KOOS
- accessibility
  - secure online platform
  - Internet-enabled device
  - responsive mode

Example: closed-ended question(s)

Symptoms
These questions should be answered thinking of your knee symptoms during the last week.

S1. Do you have swelling in your knee?
- Never
- Rarely
- Sometimes
- Often
- Always

S2. Do you feel grinding, hear clicking or any other type of noise when your knee moves?
- Never
- Rarely
- Sometimes
- Often
- Always

S3. Does your knee catch or hang up when moving?
- Never
- Rarely
- Sometimes
- Often
- Always

S4. Can you straighten your knee fully?
- Always
- Often
- Sometimes
- Rarely
- Never

S5. Can you bend your knee fully?
- Always
- Often
- Sometimes
- Rarely
- Never
Example: open-ended question

Can you describe any knee **symptoms** you have experienced over the past week?

**Answer:**

I am having trouble getting full extension as well, guess its more accurate to say, it takes a while. At first when I stretch in the bath around the bottom of the knee cap after a few minutes the knee gets more extension and then after about 5 minutes it goes flat like the other one. This is new for me and appeared after I started walking and started doing some heel props on a chair at work to improve my extension. Surgeon says it was good but at a few degrees. This was all about 2 weeks ago, I have no swelling, just some bruising on the bottom of knee cap and a real sore spot to touch on the bottom inside of my knee just to the inside of the patella tendon. It was not like this when the surgeon examined me, now when physio asks me to straighten let it take a few minutes for it to stretch out and it hurts around the bottom of knee like I said.

<table>
<thead>
<tr>
<th>ID</th>
<th>Question</th>
<th>Coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>What knee <strong>condition</strong> are you currently receiving treatment for?</td>
<td>what? where?</td>
</tr>
<tr>
<td>Q2</td>
<td>What treatment are you currently using for your knee?</td>
<td>what?</td>
</tr>
<tr>
<td>Q3</td>
<td>Have there been any <strong>changes</strong> to your knee condition over the past week?</td>
<td>↓ worse &lt; same &lt; better ↑</td>
</tr>
<tr>
<td>Q4</td>
<td>How <strong>confident</strong> do you feel about looking after your knee?</td>
<td>↓ no &lt; fairly &lt; fully ↑</td>
</tr>
<tr>
<td>Q5</td>
<td>Can you describe any knee <strong>stiffness</strong> you have experienced over the past week?</td>
<td>↓ severe &lt; some &lt; none ↑</td>
</tr>
<tr>
<td>Q6</td>
<td>Can you describe any knee <strong>pain</strong> you have experienced over the past week?</td>
<td>↓ severe &lt; some &lt; none ↑</td>
</tr>
<tr>
<td>Q7</td>
<td>What knee <strong>symptoms</strong> other than stiffness and pain you have experienced over the past week?</td>
<td>↓ severe &lt; some &lt; none ↑</td>
</tr>
<tr>
<td>Q8</td>
<td>Has your knee condition limited your ability to carry out your <strong>day to day tasks</strong> over the past week?</td>
<td>↓ a lot &lt; somewhat &lt; no ↑</td>
</tr>
<tr>
<td>Q9</td>
<td>Has your knee condition limited your ability to carry out your <strong>work, hobbies or exercise</strong> over the past week?</td>
<td>↓ a lot &lt; somewhat &lt; no ↑</td>
</tr>
<tr>
<td>Q10</td>
<td>Do you have any other comments?</td>
<td>↓ neg &lt; neutral &lt; pos ↑</td>
</tr>
</tbody>
</table>
Q9: Yes, kneeling in particular, gardening, getting in bottom cupboards to try and find shoes.

Advantages

- therapeutic benefits (writing therapy)
- listening to patients
- clinical decision making
## Closing in open-ended questions

<table>
<thead>
<tr>
<th>ID</th>
<th>Question</th>
<th>Coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>What knee condition are you currently receiving treatment for?</td>
<td>what? where?</td>
</tr>
<tr>
<td>Q2</td>
<td>What treatment are you currently using for your knee?</td>
<td>what?</td>
</tr>
<tr>
<td>Q3</td>
<td>Have there been any changes to your knee condition over the past week?</td>
<td>↓ worse &lt; same &lt; better ↑</td>
</tr>
<tr>
<td>Q4</td>
<td>How confident do you feel about looking after your knee?</td>
<td>↓ no &lt; fairly &lt; fully ↑</td>
</tr>
<tr>
<td>Q5</td>
<td>Can you describe any knee stiffness you have experienced over the past week?</td>
<td>↓ severe &lt; some &lt; none ↑</td>
</tr>
<tr>
<td>Q6</td>
<td>Can you describe any knee pain you have experienced over the past week?</td>
<td>↓ severe &lt; some &lt; none ↑</td>
</tr>
<tr>
<td>Q7</td>
<td>What knee symptoms other than stiffness and pain you have experienced over the past week?</td>
<td>↓ severe &lt; some &lt; none ↑</td>
</tr>
<tr>
<td>Q8</td>
<td>Has your knee condition limited your ability to carry out your day to day tasks over the past week?</td>
<td>↓ a lot &lt; somewhat &lt; no ↑</td>
</tr>
<tr>
<td>Q9</td>
<td>Has your knee condition limited your ability to carry out your work, hobbies or exercise over the past week?</td>
<td>↓ a lot &lt; somewhat &lt; no ↑</td>
</tr>
<tr>
<td>Q10</td>
<td>Do you have any other comments?</td>
<td>↓ neg &lt; neutral &lt; pos ↑</td>
</tr>
</tbody>
</table>
Unified Medical Language System (UMLS)

Semantic types
### Features

**Q9:** Has your knee condition limited your ability to carry out your work, hobbies or exercise over the past week?

**A:** Not been able to carry out my normal job at work due to my injury as I am a commercial vehicle mechanic. Limited exercise and hobbies. I am very frustrated that I haven't been able to play squash due to the injury and golf as I can't twist my knee fully yet.

**C:** a lot < somewhat < no?
Features

Q9: Has your knee condition limited your ability to carry out your **work, hobbies or exercise** over the past week?

A: Not been able to carry out my normal **job at work** due to my injury as I am a commercial **vehicle mechanic**. Limited **exercise** and **hobbies**. I am very frustrated that I haven't been able to play **squash** due to the injury and **golf** as I can't twist my knee fully yet.

C: a lot < somewhat < no?

---

Features

Q9: Has your knee condition limited your **ability** to carry out your **work, hobbies or exercise** over the past week?

A: Not been able to carry out my normal **job at work** due to my injury as I am a commercial **vehicle mechanic**. Limited **exercise** and **hobbies**. I am very frustrated that I haven't been able to play **squash** due to the injury and **golf** as I can't twist my knee fully yet.

C: a lot < somewhat < no?
Features

Q9: Has your knee condition limited your ability to carry out your work, hobbies or exercise over the past week?

A: Not been able to carry out my normal job at work due to my injury as I am a commercial vehicle mechanic. Limited exercise and hobbies. I am very frustrated that I haven't been able to play squash due to the injury and golf as I can't twist my knee fully yet.

C: a lot < somewhat < no?
Features

- concepts
- frequency
- context
  - negation
  - sentiment

... methods

- ontology mapping
- dependency parsing
- sentiment analysis

... tools

- MetaMap
- Core NLP
- Core NLP

Supervised machine learning
## Cross–validation results

<table>
<thead>
<tr>
<th>Question</th>
<th>Topic</th>
<th>Method</th>
<th>Features</th>
<th>P (%)</th>
<th>R (%)</th>
<th>F (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>condition</td>
<td>MetaMap</td>
<td>N/A</td>
<td>95.3</td>
<td>87.8</td>
<td>91.4</td>
</tr>
<tr>
<td>Q2</td>
<td>treatment</td>
<td>MetaMap</td>
<td>N/A</td>
<td>82.4</td>
<td>63.0</td>
<td>71.4</td>
</tr>
<tr>
<td>Q3</td>
<td>changes</td>
<td>complement naive Bayes</td>
<td>8</td>
<td>88.5</td>
<td>85.3</td>
<td>86.9</td>
</tr>
<tr>
<td>Q4</td>
<td>confidence</td>
<td>random tree</td>
<td>8</td>
<td>68.1</td>
<td>68.8</td>
<td>68.4</td>
</tr>
<tr>
<td>Q5</td>
<td>stiffness</td>
<td>logistic model tree</td>
<td>8</td>
<td>85.9</td>
<td>81.3</td>
<td>83.5</td>
</tr>
<tr>
<td>Q6</td>
<td>pain</td>
<td>complement naive Bayes</td>
<td>10</td>
<td>64.8</td>
<td>62.5</td>
<td>63.6</td>
</tr>
<tr>
<td>Q7</td>
<td>symptoms</td>
<td>naive Bayes</td>
<td>5</td>
<td>78.6</td>
<td>80.6</td>
<td>79.6</td>
</tr>
<tr>
<td>Q8</td>
<td>daily</td>
<td>ordinal class classifier</td>
<td>14</td>
<td>71.3</td>
<td>61.3</td>
<td>65.9</td>
</tr>
<tr>
<td>Q9</td>
<td>work</td>
<td>J48 pruned tree</td>
<td>14</td>
<td>84.1</td>
<td>80.6</td>
<td>82.3</td>
</tr>
<tr>
<td>Q10</td>
<td>other</td>
<td>Stanford Core NLP</td>
<td>N/A</td>
<td>70.2</td>
<td>70.9</td>
<td>70.5</td>
</tr>
</tbody>
</table>
Conclusion

- Can we measure self-care with text analytics?
  - Absolutely!
- next steps
  1. get more training/test data
  2. re-train and properly evaluate on unseen data
  3. embed automated measurement into patient dashboard
- future work
  1. validate as an instrument
  2. explore speech-to-text for data collection