Extractive Evidence Based Medicine
Summarisation Based on Sentence-Specific Statistics

Abeed Sarker\textsuperscript{1}  Diego Mollá\textsuperscript{1}  Cécile Paris\textsuperscript{2}

\textsuperscript{1}Centre for Language Technology, Macquarie University, Sydney
\textsuperscript{2}CSIRO ICT Centre, Sydney

CBMS 2012, Rome
Contents

Background
   Evidence Based Medicine

Method
   Corpus
   Generation of Statistics

Evaluation
Contents

Background
    Evidence Based Medicine

Method
    Corpus
    Generation of Statistics

Evaluation
Contents

Background
Evidence Based Medicine

Method
Corpus
Generation of Statistics

Evaluation
Evidence Based Medicine

EBM and Natural Language Processing


NLP tasks

- Question analysis and classification
- Information Retrieval
- Classification and re-ranking
- Information extraction
- Question answering
- Summarisation
Contents

Background
Evidence Based Medicine

Method
Corpus
Generation of Statistics

Evaluation
General Approach

In a Nutshell

1. Gather statistics from the best 3-sentence extracts.
   ▶ Exhaustive search to find these best extracts.

2. Build three classifiers, one per sentence in the final extract.
   ▶ Classifier 1 based on statistics from best 1st sentence.
   ▶ Classifier 2 based on statistics from best 2nd sentence.
   ▶ Classifier 3 based on statistics from best 3rd sentence.
Contents

Background

Evidence Based Medicine

Method

Corpus

Generation of Statistics

Evaluation
Which treatments work best for hemorrhoids?

Evidence-based answer

Excision is the most effective treatment for thrombosed external hemorrhoids of non-confluent thrombosed hemorrhoids (SOR: A, systematic reviews). Of nonoperative techniques, rubber band ligation produces the lowest rate of recurrence (SOR: A, systematic review).

Evidence summary

External hemorrhoids originate below the dentate line and become acutely painful with thrombosis. They can cause perianal pruritus and excoriation because of interference with perianal hygiene. Internal hemorrhoids become symptomatic when they bleed or prolapse (Table).

For thrombosed external hemorrhoids, surgery works best. Few studies have evaluated the best treatment for thrombosed external hemorrhoids. A retrospective study of 231 patients treated conservatively or surgically found that the 48.5% of patients treated surgically had a lower recurrence rate than the conservative group (number needed to treat [NNT]=2 for recurrence at mean follow-up of 8.6 months) and earlier resolution of symptoms (average 3.9 days compared with 24 days for conservative treatment).

Another retrospective analysis of 340 patients who underwent outpatient excision of thrombosed external hemorrhoids under local anesthesia reported a lower recurrence rate of 6.5% at a mean follow-up of 17.3 months.

A prospective, randomized controlled trial (RCT) of 98 patients treated nonsurgically found improved pain relief with a combination of topical nitric oxide 0.3% and lidocaine 1.5% compared with lidocaine alone. The NNT for complete pain relief at 7 days was 3.

Conventional hemorrhoidectomy beats stapling

Many studies have evaluated the best treatment for prolapsed hemorrhoids. A Cochrane systematic review of 12 RCTs that compared conventional hemorrhoidectomy with stapled hemorrhoidectomy in patients with grades I to III hemorrhoids found a lower rate of recurrence (follow-up ranged from 6 to 39 months) in patients who had conventional hemorrhoidectomy (NNT=144). Conventional hemorrhoidectomy showed a nonsignificant trend in decreased bleeding and decreased incontinence.

A second systematic review of 25 studies, including some that were of lower quality, showed a higher recurrence rate at 1 year with stapled hemorrhoidectomy than with conventional surgery.

Nonoperative techniques? Consider rubber band ligation

A systematic review of 3 poor-quality trials comparing rubber band ligation with excisional hemorrhoidectomy in patients with grade III hemorrhoids found that excisional hemorrhoidectomy produced better long-term symptom control but more immediate postoperative complications of anal stenosis and hemorrhage.

Rubber band ligation had the lowest recurrence rate at 12 months compared with the other nonoperative techniques of sclerotherapy and infrared coagulation.

Fiber supplements help relieve symptoms

A Cochrane systematic review of 7 RCTs involving a total of 378 patients with grade I or II hemorrhoids evaluated the effect of fiber supplements on pain, itching, and bleeding. Persistent hemorrhoid symptoms decreased by 53% in the group receiving fiber.

When surgical hemorrhoidectomy is recommended

The American Society of Colon and Rectal Surgeons recommends adequate fluid and fiber intake for all patients with symptomatic hemorrhoids. For grade II or III hemorrhoids, the society states that banding is usually most effective. When officed treatments fail, the society recommends surgical hemorrhoidectomy (SOR: B).

The society recommends excision of thrombosed hemorrhoids less than 72 hours old and expectant treatment with hemorrhoids that present early. Surgical hemorrhoidectomy should be reserved for when conservative treatment fails and for patients with symptomatic grade III and IV hemorrhoids.

References


The XML Contents I

<record id="7843">
  <url>http://www.jfponline.com/Pages.asp?AID=7843&amp;issue=September_2009&amp;UID=</url>
  <question>Which treatments work best for hemorrhoids?</question>
  <answer>
    <snip id="1">
      <snipertext>Excision is the most effective treatment for thrombosed external hemorrhoids.</snipertext>
    </snip>
    <sor type="B">retrospective studies</sor>
  </answer>
  <long id="1_1">
    <longtext>A retrospective study of 231 patients treated conservatively or surgically found that the 48.5% of patients treated surgically had a lower recurrence rate than the conservative group (number needed to treat [NNT]=2 for recurrence at mean follow-up of 7.6 months) and earlier resolution of symptoms (average 3.9 days compared with 24 days for conservative treatment).</longtext>
  </long>
  <long id="1_2">
    <longtext>A retrospective analysis of 340 patients who underwent outpatient excision of thrombosed external hemorrhoids under local anesthesia reported a low recurrence rate of 6.5% at a mean follow-up of 17.3 months.</longtext>
  </long>
</record>
The XML Contents II


A prospective, randomized controlled trial (RCT) of 98 patients treated nonsurgically found improved pain relief with a combination of topical nifedipine 0.3% and lidocaine 1.5% compared with lidocaine alone. The NNT for complete pain relief at 7 days was 3.

Corpus Statistics

Size

- 456 questions ("records").
- Over 1,100 distinct answers ("snips").
- 3,036 text explanations ("longs").
- 2,707 references.
Summarisation Using This Corpus

Input

▶ Question.
▶ Document Abstract.

Output

▶ Extractive summary that answers the question.
▶ Target summary is the annotated evidence text (“long”).
▶ Evaluated using ROUGE-L.
Contents

Background
Evidence Based Medicine

Method
Corpus
Generation of Statistics

Evaluation
The Statistics Gathered

1. Source sentence position.
2. Sentence length.
3. Sentence similarity.
4. Sentence type.
1. Source Sentence Position

- Compute relative positions.
- Create normalised frequency histograms $f_1, f_2, \ldots, f_{10}$.
- Score all relative positions of bin $i$ with its bin frequency: $S_{pos}(i) = f_{bin(i)}$. 

![Frequency Distributions for Relative Sentence Positions](image1.png)

![Frequency Distributions for Relative Sentence Positions](image2.png)

![Frequency Distributions for Relative Sentence Positions](image3.png)
2. Sentence Length

Reward larger sentences and penalise shorter sentences:

**Normalised sentence length**

\[
S_{len}(i) = \frac{l_s - l_{avg}}{l_d}
\]

- \(l_s\): sentence length
- \(l_{avg}\): average sentence length in the corpus
- \(l_d\): document length
3. Sentence Similarity

Sentence Similarity

- Lowercase, stem, remove stop words.
- Build vector of $tf.idf$ with remaining words and UMLS semantic types.
- $\text{CosSim}(X, Y) = \frac{X \cdot Y}{||X|| ||Y||}$

Maximal Marginal Relevance (Carbonell & Goldstein, 1998)

Reward sentences similar to the query and penalise those similar to other summary sentences.

$$\text{MMR} = \lambda(\text{CosSim}(S_i, Q)) - (1 - \lambda)\max_{S_j \in S} (\text{CosSim}(S_i, S_j))$$
4. PIBOSO (Kim et al. 2011) I

1. Classify all sentences into PIBOSO types (a variant of PICO).
2. Generate normalised frequency histograms of resulting PIBOSO types.
4. PIBOSO (Kim et al. 2011) II

Position independent

\[ S_{PIPS}(i) = \frac{P_{best}}{P_{all}} \]

Position dependent

\[ S_{PDPS}(i) = \frac{P_{pos}}{P_{best}} \]

- \( P_{best} \): proportion of this PIBOSO type among all best summary sentences.
- \( P_{all} \): proportion of this PIBOSO type among all sentences.
- \( P_{pos} \): proportion of this PIBOSO type among at best summary sentences at this position.
Classification

Edmunsonian Formula

\[ S_{S_i} = \alpha S_{rpos_i} + \beta S_{len_i} + \gamma S_{PIPS_i} + \delta S_{PDPS_i} + \epsilon S_{MMR_i} \]

- MMR is replaced with cosine similarity for first sentence.
- In case of ties, the sentence with greatest length is chosen.
- Parameters are fine-tuned through exhaustive search using training set.

\[ \alpha = 1.0, \beta = 0.8, \gamma = 0.1, \delta = 0.8, \epsilon = 0.1, \lambda = 0.1. \]
Contents

Background
- Evidence Based Medicine

Method
- Corpus
- Generation of Statistics

Evaluation
Percentile-based Evaluation (Ceylan et al. 2010)

We compare against all possible 3-sentence extracts in the \textit{test set}.

1. Bin all possible three-sentence combinations of each abstract.
   - 1,000 bins.
2. Normalise the resulting histograms.
3. Combine all histograms.
   - convolution.
4. The result approximates the probability density distribution of all three-sentence summaries in all abstracts.
Percentile-based Evaluation (Ceylan et al. 2010) II
Systems

L3 Last three sentences.
O3 Last three PIBOSO outcome sentences.
R Random.
O All outcome sentences.
PI Sentence position independent.

PD Sentence position dependent (our proposal).
## Results

<table>
<thead>
<tr>
<th>System</th>
<th>F-Score</th>
<th>95% CI</th>
<th>Percentile (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L3</td>
<td>0.159</td>
<td>0.155–0.163</td>
<td>60.3</td>
</tr>
<tr>
<td>O3</td>
<td>0.161</td>
<td>0.158–0.165</td>
<td>77.5</td>
</tr>
<tr>
<td>R</td>
<td>0.158</td>
<td>0.154–0.161</td>
<td>50.3</td>
</tr>
<tr>
<td>O</td>
<td>0.159</td>
<td>0.155–0.164</td>
<td>60.3</td>
</tr>
<tr>
<td>PI</td>
<td>0.160</td>
<td>0.157–0.164</td>
<td>69.4</td>
</tr>
<tr>
<td>PD</td>
<td>0.166</td>
<td>0.162–0.170</td>
<td>97.3</td>
</tr>
</tbody>
</table>
Questions?

Background
Evidence Based Medicine

Method
Corpus
Generation of Statistics

Evaluation

Further Information
http://web.science.mq.edu.au/~diego/medicalnlp/