A Corpus for Evidence Based Medicine Summarisation

Diego Mollá

Centre for Language Technology,
Macquarie University

ALTA, 10 Dec 2010
Contents

Evidence Based Medicine and Summarisation

A Corpus for Summarisation

Summarisation Experiments
Contents

Evidence Based Medicine and Summarisation

A Corpus for Summarisation

Summarisation Experiments
Evidence Based Medicine

EBM and Natural Language Processing

The Five Steps of EBM

1. Frame Patient Scenario into a Clinical Question
2. Systematically retrieve best evidence available
3. Critically appraise evidence
4. Apply results to patient
5. Evaluate decision making

EBM and Natural Language Processing

The Five Steps of EBM

1. Frame Patient Scenario into a Clinical Question
2. Systematically retrieve best evidence available
3. Critically appraise evidence
4. Apply results to patient
5. Evaluate decision making

NLP Tasks

- Question analysis and classification

EBM and Natural Language Processing

The Five Steps of EBM

1. Frame Patient Scenario into a Clinical Question
2. Systematically retrieve best evidence available
3. Critically appraise evidence
4. Apply results to patient
5. Evaluate decision making

NLP Tasks

- Question analysis and classification
- Information retrieval

EBM and Natural Language Processing

The Five Steps of EBM

1. Frame Patient Scenario into a Clinical Question
2. Systematically retrieve best evidence available
3. Critically appraise evidence
4. Apply results to patient
5. Evaluate decision making

NLP Tasks

- Question analysis and classification
- Information retrieval
- Information extraction

EBM and Natural Language Processing

The Five Steps of EBM

1. Frame Patient Scenario into a Clinical Question
2. Systematically retrieve best evidence available
3. Critically appraise evidence
4. Apply results to patient
5. Evaluate decision making

NLP Tasks

- Question analysis and classification
- Information retrieval
- Information extraction
- Classification and re-ranking

EBM and Natural Language Processing

**The Five Steps of EBM**

1. **Frame Patient Scenario into a Clinical Question**
2. Systematically retrieve best evidence available
3. Critically appraise evidence
4. Apply results to patient
5. Evaluate decision making

**Librarian Centred**

**NLP Tasks**

- Question analysis and classification
- Information retrieval
- Information extraction
- Classification and re-ranking
- Question answering

EBM and Natural Language Processing

The Five Steps of EBM

1. Frame Patient Scenario into a Clinical Question
2. Systematically retrieve best evidence available
3. Critically appraise evidence
4. Apply results to patient
5. Evaluate decision making

Librarian Centred

NLP Tasks

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

Where’s the Corpus for Summarisation?

Systems

- CENTRIFUSER/PERSIVAL: Developed and tested using user feedback (iterative design)
- SemRep: Evaluation based on human judgement
- Demner-Fushman & Lin: ROUGE on original paper abstracts
- Fiszman: Factoid-based evaluation
Where’s the Corpus for Summarisation?

Systems

- CENTRIFUSER/PERSIVAL: Developed and tested using user feedback (iterative design)
- SemRep: Evaluation based on human judgement
- Demner-Fushman & Lin: ROUGE on original paper abstracts
- Fiszman: Factoid-based evaluation

Corpora

- Several corpora of questions/answers available
- Answers lack explicit pointers to primary literature
- Medical doctors want to know the primary sources
Contents

Evidence Based Medicine and Summarisation

A Corpus for Summarisation

Summarisation Experiments
Which treatments work best for hemorrhoids?

**Evidence-based answer**

Excision is the most effective treatment for thrombosed external hemorrhoids (strength of recommendation [SOR]: A, systematic reviews). Of nonoperative techniques, rubber band ligation produces the lowest rate of recurrence (SOR: A, systematic reviews).

**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]: 24 for recurrence at mean follow-up of 7.6 months) and earlier resolution of symptoms (average 3.9 days compared with 2.4 days for conservative treatment).1

Another 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.2 A prospective, randomized controlled trial (RCT) of 98 patients treated nonsurgically found improved pain relief with a combination of topical nefedipine 0.3% and lidocaine 1.5% compared with lidocaine alone. The NNT for complete pain relief at 7 days was 3.3

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 hemorrhoidopexy 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=14).4 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.3

**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.4 Rubber band ligation had the lowest recurrence rate at 12 months compared with the other nonoperative techniques of sclerotherapy and infrared coagulation.7

**Fiber supplements help relieve symptoms**

A Cochrane systematic review of 7 RCTs enrolling a total of 378 patients with grade I to III hemorrhoids evaluated the effect of fiber supplements on pain, itching, and bleeding. Persistent hemorrhoid symptoms decreased by 53% in the group receiving fiber.6

**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 I to III hemorrhoids, the society states that banding is usually most effective. When office 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 analgesia and sitz baths for thrombosed hemorrhoids present for longer than 72 hours (SOR:B).

The American Gastroenterological Association recommends excision of symptomatic thrombosed external hemorrhoids that present early. Surgical hemorrhoidectomy should be reserved for when conservative treatment fails and for patients with symptomatic grade III and IV hemorrhoids.11

**References**


Which treatments work best for hemorrhoids?

**Evidence-based answer**
Excision is the most effective treatment for thrombosed external hemorrhoids (strength of recommendation [SOR]: B, retrospective studies). For prolapsed internal hemorrhoids, the best definitive treatment is traditional hemorrhoidectomy (SOR: A, systematic reviews). Of nonoperative techniques, rubber band ligation produces the lowest rate of recurrence (SOR: A, systematic reviews).

**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.

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...
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 7.6 months) and earlier resolution of symptoms (average 3.9 days compared with 24 days for conservative treatment).\(^1\)

Another retrospective analysis of 340 patients who underwent outpatient excision of thrombosed external hemorrhoids under local anesthesia re-

midodine 0.5% and lidocaine 1.5% compared with lidocaine alone. The NNT for complete pain relief at 7 days was 3.\(^3\)

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=14).\(^4\) Conventional hemorrhoidectomy showed a nonsignificant trend in decreased bleeding and decreased incontinence.

A second systematic review of 25 studies, including some that were of
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 enrolling a total of 378 patients with grade I to III 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 I to III hemorrhoids, the society recommends hemorrhoid ligation is usually most of 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**


An extract of our corpus

<question>Which treatments work best for hemorrhoids?</question>

<Answer>Excision is the most effective treatment for thrombosed external hemorrhoids

<long>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). <ref ID="15486746" />

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. <ref ID="12972967" />

For prolapsed internal hemorrhoids, the best definitive treatment is traditional hemorrhoidectomy.

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 (\(NNT=14\)). Conventional hemorrhoidectomy showed a nonsignificant trend in decreased bleeding and decreased incontinence. <ref ID="17054255" />

A systematic review of 25 studies showed a higher recurrence rate at 1 year with stapled hemorrhoidectomy than with conventional surgery. <ref ID="17380367" />

... </answer>
Components of the Corpus

Components

Question  direct extract from the source
Answer   split from the source and manually checked
Evidence extracted from the source
Additional text manually extracted from the source and massaged
References PMID looked up in PubMed (automatic and manual procedure)

Planned Size

- 496 questions
- 3,000 references (a very rough estimate)
Status

Done

- All data converted from source to intermediate format
- All questions automatically extracted and split
- All evidence types automatically extracted
- All reference IDs automatically looked up
- Annotation tool functional
Status

Done

▶ All data converted from source to intermediate format
▶ All questions automatically extracted and split
▶ All evidence types automatically extracted
▶ All reference IDs automatically looked up
▶ Annotation tool functional

To Do

▶ Manually check questions and evidence types
▶ Manually extract and massage text
▶ Manually check reference IDs
### JFP Corpus Annotation Tool

<table>
<thead>
<tr>
<th>SNIP ID</th>
<th>SNIP TEXT</th>
<th>SOR TYPE</th>
<th>SOR BASES</th>
<th>REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Excision is the most effective treatment for thrombosed external hemorrhoids.</td>
<td>B</td>
<td>retrospective studies</td>
<td>None</td>
</tr>
<tr>
<td>1_1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+Long</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>For prolapsed internal hemorrhoids, the best definitive treatment is traditional hemorrhoidectomy.</td>
<td>A</td>
<td>systematic reviews</td>
<td>None</td>
</tr>
</tbody>
</table>
Contents

Evidence Based Medicine and Summarisation

A Corpus for Summarisation

Summarisation Experiments
Summarisation Framework

- Single document summarisation
- Use ROUGE on the target text
- Pilot corpus fragment
  - 12 questions
  - 73 references
Straight Baselines

Systems

Last Return the last $n$ sentences

Outcomes Return the output of NLM’s outcome extractor
Straight Baselines

Systems

Last  Return the last $n$ sentences

Outcomes  Return the output of NLM’s outcome extractor

Results

<table>
<thead>
<tr>
<th>System</th>
<th>$n$</th>
<th>Avg $F$</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Last</td>
<td>3</td>
<td>0.183</td>
<td>[0.159–0.206]</td>
</tr>
<tr>
<td>Outcomes</td>
<td>3</td>
<td>0.181</td>
<td>[0.158–0.205]</td>
</tr>
</tbody>
</table>
Query-based Baselines

**Simple** Return the last $n$ sentences that share any non-stop words with the question

**UMLS C** Return the last $n$ sentences that share any UMLS concepts with the question

**UMLS G** Return the last $n$ sentences that have the greatest graph similarity with the question (random walks on UMLS relations using Eneko Agirre’s system)
## Query-based Baseline Results

<table>
<thead>
<tr>
<th>System</th>
<th>n</th>
<th>Avg F</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Last</td>
<td>3</td>
<td>0.183</td>
<td>[0.159–0.206]</td>
</tr>
<tr>
<td>Outcomes</td>
<td>3</td>
<td>0.181</td>
<td>[0.158–0.205]</td>
</tr>
<tr>
<td>Simple</td>
<td>3</td>
<td>0.180</td>
<td>[0.157–0.203]</td>
</tr>
<tr>
<td>UMLS C</td>
<td>3</td>
<td>0.185</td>
<td>[0.161–0.209]</td>
</tr>
<tr>
<td>UMLS G</td>
<td>3</td>
<td>0.172</td>
<td>[0.149–0.194]</td>
</tr>
</tbody>
</table>
Using the Abstract Structure

Preselect sentences and then:

1. Map each section to one of: background, setting, design, results, conclusion, evidence, appendix
2. Select the first $n$ sentences of the last "conclusion" section
3. If we have less than $n$ sentences, fill from the first sentences of the previous "conclusion" section, and so on until all "conclusion" sections are used up
4. If we have less than $n$ sentences, fill from the "results" sections
5. If we still have less than $n$ sentences, fill from the "design" sections
6. If the abstract has no structure, return the last $n$ sentences

Abstract

Section 1  S1.1 S1.2
Section 2  S2.1
Section 3  S3.1 S3.2
Section 4  S4.1 S4.2
Section 5  S5.1 S5.2
Section 6  S6.1

Summary
Using the Abstract Structure

Preselect sentences and then:

1. Map each section to one of: background, setting, design, results, conclusion, evidence, appendix

<table>
<thead>
<tr>
<th>Abstract</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background</td>
<td>S1.1 S1.2</td>
</tr>
<tr>
<td>Design</td>
<td>S2.1</td>
</tr>
<tr>
<td>Results</td>
<td>S3.1 S3.2</td>
</tr>
<tr>
<td>Conclusion</td>
<td>S4.1 S4.2</td>
</tr>
<tr>
<td>Conclusion</td>
<td>S5.1 S5.2</td>
</tr>
<tr>
<td>Appendix</td>
<td>S6.1</td>
</tr>
</tbody>
</table>
Using the Abstract Structure

Preselect sentences and then:

1. Map each section to one of: background, setting, design, results, conclusion, evidence, appendix

2. Select the first $n$ sentences of the last “conclusion” section

3. If we have less than $n$ sentences, fill from the first sentences of the previous “conclusion” section, and so on until all “conclusion” sections are used up

4. If we have less than $n$ sentences, fill from the “results” sections

5. If we still have less than $n$ sentences, fill from the “design” sections

6. If the abstract has no structure, return the last $n$ sentences

---

Abstract

<table>
<thead>
<tr>
<th>Section</th>
<th>S1.1</th>
<th>S1.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Design</td>
<td>S2.1</td>
<td></td>
</tr>
<tr>
<td>Results</td>
<td>S3.1</td>
<td>S3.2</td>
</tr>
<tr>
<td>Conclusion</td>
<td>S4.1</td>
<td>S4.2</td>
</tr>
<tr>
<td>Conclusion</td>
<td>S5.1</td>
<td>S5.2</td>
</tr>
<tr>
<td>Appendix</td>
<td>S6.1</td>
<td></td>
</tr>
</tbody>
</table>

Summary

| S5.1 | S5.2 |
Using the Abstract Structure

Preselect sentences and then:

1. Map each section to one of: background, setting, design, results, conclusion, evidence, appendix
2. Select the first $n$ sentences of the last “conclusion” section
3. If we have less than $n$ sentences, fill from the first sentences of the previous “conclusion” section, and so on until all “conclusion” sections are used up

Abstract
Background S1.1 S1.2
Design S2.1
Results S3.1 S3.2
Conclusion S4.1 S4.2
Conclusion S5.1 S5.2
Appendix S6.1

Summary
S5.1 S5.2 S4.1 S4.2
Using the Abstract Structure

Preselect sentences and then:

1. Map each section to one of: background, setting, design, results, conclusion, evidence, appendix
2. Select the first $n$ sentences of the last “conclusion” section
3. If we have less than $n$ sentences, fill from the first sentences of the previous “conclusion” section, and so on until all “conclusion” sections are used up
4. If we have less than $n$ sentences, fill from the “results” sections

Abstract

| Background | S1.1 S1.2 |
| Design    | S2.1      |
| Results   | S3.1 S3.2 |
| Conclusion| S4.1 S4.2 |
| Conclusion| S5.1 S5.2 |
| Appendix  | S6.1      |

Summary

S5.1 S5.2 S4.1 S4.2 S3.1
Using the Abstract Structure

Preselect sentences and then:

1. Map each section to one of: background, setting, design, results, conclusion, evidence, appendix
2. Select the first $n$ sentences of the last “conclusion” section
3. If we have less than $n$ sentences, fill from the first sentences of the previous “conclusion” section, and so on until all “conclusion” sections are used up
4. If we have less than $n$ sentences, fill from the “results” sections
5. If we still have less than $n$ sentences, fill from the “design” sections

<table>
<thead>
<tr>
<th>Abstract</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background</td>
<td>S5.1 S5.2 S4.1 S4.2 S3.1</td>
</tr>
<tr>
<td>Design</td>
<td>S2.1</td>
</tr>
<tr>
<td>Results</td>
<td>S3.1 S3.2</td>
</tr>
<tr>
<td>Conclusion</td>
<td>S4.1 S4.2</td>
</tr>
<tr>
<td>Conclusion</td>
<td>S5.1 S5.2</td>
</tr>
<tr>
<td>Appendix</td>
<td>S6.1</td>
</tr>
</tbody>
</table>
Using the Abstract Structure

Preselect sentences and then:

1. Map each section to one of: background, setting, design, results, conclusion, evidence, appendix
2. Select the first \( n \) sentences of the last “conclusion” section
3. If we have less than \( n \) sentences, fill from the first sentences of the previous “conclusion” section, and so on until all “conclusion” sections are used up
4. If we have less than \( n \) sentences, fill from the “results” sections
5. If we still have less than \( n \) sentences, fill from the “design” sections
6. If the abstract has no structure, return the last \( n \) sentences

Abstract

Background S1.1 S1.2
Design S2.1
Results S3.1 S3.2
Conclusion S4.1 S4.2
Conclusion S5.1 S5.2
Appendix S6.1

Summary

S5.1 S5.2 S4.1 S4.2 S3.1
Abstract Structure Results

<table>
<thead>
<tr>
<th>System</th>
<th>n</th>
<th>Avg F</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Last</td>
<td>3</td>
<td>0.183</td>
<td>[0.159–0.206]</td>
</tr>
<tr>
<td>Outcomes</td>
<td>3</td>
<td>0.181</td>
<td>[0.158–0.205]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>System</th>
<th>n</th>
<th>Avg F</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>3</td>
<td>0.180</td>
<td>[0.157–0.203]</td>
</tr>
<tr>
<td>UMLS C</td>
<td>3</td>
<td>0.185</td>
<td>[0.161–0.209]</td>
</tr>
<tr>
<td>UMLS G</td>
<td>3</td>
<td>0.172</td>
<td>[0.149–0.194]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>System</th>
<th>n</th>
<th>Avg F</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Overlap</td>
<td>3</td>
<td>0.184</td>
<td>[0.161–0.206]</td>
</tr>
<tr>
<td>Word</td>
<td>3</td>
<td>0.178</td>
<td>[0.154–0.199]</td>
</tr>
<tr>
<td>UMLS</td>
<td>3</td>
<td>0.185</td>
<td>[0.160–0.209]</td>
</tr>
</tbody>
</table>
Selected Results (samples=720)

The ROUGE results by duplicating all summaries by 10 for the two most differing scores are:

<table>
<thead>
<tr>
<th>System</th>
<th>n</th>
<th>Avg F</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>UMLS Concepts</td>
<td>3</td>
<td>0.185</td>
<td>[0.178–0.193]</td>
</tr>
<tr>
<td>UMLS Graph</td>
<td>3</td>
<td>0.172</td>
<td>[0.165–0.179]</td>
</tr>
</tbody>
</table>
Summary and Further Work

Summary

- Developing a corpus for EBM summarisation
- Initial baseline experiments
Summary and Further Work

Summary

- Developing a corpus for EBM summarisation
- Initial baseline experiments

Further Work

- Complete the corpus
- Repeat the baseline experiments
- Use corpus for multi-document summarisation
Summary and Further Work

Summary

▶ Developing a corpus for EBM summarisation
▶ Initial baseline experiments

Further Work

▶ Complete the corpus
▶ Repeat the baseline experiments
▶ Use corpus for multi-document summarisation

QUESTIONS?