Automated Summarisation for Evidence Based Medicine

Diego Mollá

Centre for Language Technology,
Macquarie University

HAIL, 22 March 2012
Contents

Evidence Based Medicine

Our Corpus for Summarisation
  Structure of our Corpus
  How we Created the Corpus
  Statistics

Applications
  Possible Uses
  Single-document Summarisation
  Evidence Grading
About Us: Research Group on Natural Language Processing of Medical Texts

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

Active Members

Diego Molla  Senior lecturer at Macquarie University.
Cécile Paris  Senior principal research scientist at CSIRO ICT Centre.
Abeed Sarker  PhD student at Macquarie University.
Sara Faisal Shash  Masters student.

Past Members

María Elena Santiago-Martínez  Research programmer.
Patrick Davis-Desmond  Masters student.
Andreea Tutos  Masters student.
Contents

Evidence Based Medicine

Our Corpus for Summarisation
  Structure of our Corpus
  How we Created the Corpus
  Statistics

Applications
  Possible Uses
  Single-document Summarisation
  Evidence Grading
Evidence Based Medicine

EBM and Natural Language Processing

# Evidence Based Medicine

Our Corpus for Summarisation

- Applications

---

## PICO for Asking the Right Question

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient or Problem</strong></td>
<td><strong>Intervention</strong></td>
<td><strong>Comparison Intervention</strong></td>
<td><strong>Outcomes</strong></td>
</tr>
<tr>
<td>Starting with your patient, ask “How would I describe a group of patients similar to mine?”</td>
<td>Ask “Which main intervention am I considering?”</td>
<td>Ask “What is the main alternative to compare with the intervention?”</td>
<td>Ask “What can I hope to accomplish?”, or “What could this exposure really affect?”</td>
</tr>
<tr>
<td><strong>Tips for Building</strong></td>
<td><strong>(a cause, prognostic factor, treatment, etc.)</strong></td>
<td>Again, be specific</td>
<td>Again, be specific</td>
</tr>
<tr>
<td>“In patients with heart failure from dilated cardiomyopathy who are in sinus rhythm...”</td>
<td>“…would adding anticoagulation with warfarin to standard heart failure therapy...”</td>
<td>“…when compared with standard therapy alone...”</td>
<td>“…lead to lower mortality or morbidity from thromboembolism. Is this enough to be worth the increased risk of bleeding?”</td>
</tr>
</tbody>
</table>

---

MACQUARIE UNIVERSITY

EBM Summarisation

Diego Mollá 7/60
Where to search for external evidence?

1. Evidence-based Summaries (Systematic Reviews):
   - EBM Online (http://ebm.bmj.com).
   - UptoDate (http://www.uptodate.com).
   - The Cochrane Library (http://www.thecochranelibrary.com/).
   - ...
Where to search for external evidence?

1. Evidence-based Summaries (Systematic Reviews):
   - EBM Online (http://ebm.bmj.com).
   - UptoDate (http://www.uptodate.com).
   - The Cochrane Library (http://www.thecochranelibrary.com/).
   - . . .

2. Search the Medical Literature:
Searching Cochrane

Use the tabs below to filter your search results:

Search

Cochrane Reviews Colloquia abstracts Methodology Newsletters Policy Manual Registered titles Whole site

Enter your keywords:
sleep apnea

Search

Showing results 1 to 10 of 12

Anti-inflammatory drugs for the treatment of obstructive sleep ...

Pharmacotherapy for hypertension in adults with obstructive ...

Adenotonsillectomy for obstructive sleep apnoea in children | ...
Searching PubMed

Results: 1 to 20 of 24722

1. Importance and management of chronic sleep apnoea in cardiology.
   Jaffe LM, Kjekshus J, Gottlieb SS.
   Eur Heart J. 2012 Mar 15. [Epub ahead of print]
   PMID: 22427382 [PubMed - as supplied by publisher]
   Related citations

2. Association of Inflammatory biomarkers with sleep disorders in hemodialysis patients.
   Razeghi E, Sahraian MA, Heldari R, Bagherzadeh M.
   PMID: 22427289 [PubMed - in process]
   Related citations

3. Exploration of the relationship between sleep position and isolated tongue base or multilevel surgery in obstructive sleep apnea.
   van Maanen JP, Ravesloot MJ, Witte B, Grisseels M, de Vries N.
   Eur Arch Otorhinolaryngol. 2012 Mar 20. [Epub ahead of print]
   PMID: 22427104 [PubMed - as supplied by publisher]
   Related citations

   Barnes ME, Gozal D, Wolfson DL.
Searching the Trip Database

1. Continuous positive airway pressure delivery interfaces for obstructive sleep apnoea
   - COCHRANE DATABASE OF SYSTEMATIC REVIEWS 2011
   - CPD/CME: Developing World?

2. Treatment of obstructive sleep apnoea for chronic cough in children
   - COCHRANE DATABASE OF SYSTEMATIC REVIEWS 2011
   - CPD/CME: Developing World?

3. Anti-inflammatory medications for obstructive sleep apnoea in children
   - COCHRANE DATABASE OF SYSTEMATIC REVIEWS 2011
   - CPD/CME: Developing World?

4. An integrated health-economic analysis of diagnostic and therapeutic strategies in the treatment of moderate-to-severe obstructive sleep apnoea
   - NHS ECONOMIC EVALUATION DATABASE: 2012
   - CPD/CME: Developing World?
Appraising the Evidence

The SORT Taxonomy

**Level A** Consistent and good-quality patient-oriented evidence.

**Level B** Inconsistent or limited-quality patient-oriented evidence.

**Level C** Consensus, usual practise, opinion, disease-oriented evidence, or case series for studies of diagnosis, treatment, prevention, or screening.
# Levels of Evidence

<table>
<thead>
<tr>
<th>Study quality</th>
<th>Diagnosis</th>
<th>Treatment / prevention / screening</th>
<th>Prognosis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level 1:</strong> good-quality patient-oriented evidence</td>
<td>Validated clinical decision rule; SR/meta-analysis of high-quality studies; high-quality diagnostic cohort study</td>
<td>SR/meta-analysis of RCTs with consistent findings; high-quality individual RCT; all-or-none study</td>
<td>SR/meta-analysis of good-quality cohort studies; prospective cohort study with good follow-up</td>
</tr>
<tr>
<td><strong>Level 2:</strong> limited-quality patient-oriented evidence</td>
<td>Unvalidated clinical decision rule; SR/meta-analysis of lower-quality studies or studies with inconsistent findings; lower-quality diagnostic cohort study or diagnostic case-control study</td>
<td>SR/meta-analysis of lower-quality clinical trials or of studies with inconsistent findings; lower-quality clinical trial; cohort study; case-control study</td>
<td>SR/meta-analysis of lower-quality cohort studies or with inconsistent results; retrospective cohort study or prospective cohort study with poor follow-up; case-control study; case series</td>
</tr>
<tr>
<td><strong>Level 3:</strong> other evidence</td>
<td>Consensus guidelines, extrapolations from bench research, usual practice, opinion, disease-oriented evidence (intermediate or physiologic outcomes only), or case series for studies of diagnosis, treatment, prevention, or screening</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Where can NLP Help?

- **Questions:**
  - Help to formulate answerable questions.
  - Question analysis and classification.
Where can NLP Help?

- **Questions:**
  - Help to formulate answerable questions.
  - Question analysis and classification.

- **Search:**
  - Retrieve and rank relevant literature.
  - Extract the evidence-based information.
  - Summarise the results.
Where can NLP Help?

▶ **Questions:**
  ▶ Help to formulate answerable questions.
  ▶ Question analysis and classification.

▶ **Search:**
  ▶ Retrieve and rank relevant literature.
  ▶ Extract the evidence-based information.
  ▶ Summarise the results.

▶ **Appraisal:** Classify the evidence.
Contents

Evidence Based Medicine

Our Corpus for Summarisation
  Structure of our Corpus
  How we Created the Corpus
  Statistics

Applications
  Possible Uses
  Single-document Summarisation
  Evidence Grading
Where’s the Corpus for Summarisation?

**Summarisation Systems**

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

### Summarisation Systems

- CENTRIFUSER/PERSIVAL: Developed and tested using user feedback (iterative design).
- 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

Our Corpus for Summarisation
  Structure of our Corpus
    How we Created the Corpus
  Statistics

Applications
  Possible Uses
  Single-document Summarisation
  Evidence Grading
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 excretion 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 48.3% 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 12 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 reported a low recurrence rate of 6.5% at 1 year 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 indomethacin 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 hemorrhoidectomy in patients with grade 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 delayed 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.5

**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.6 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.8

**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 hemorrhoids that present early. Surgical hemorrhoidectomy should be reserved for when conservative treatment fails and for patients with symptomatic grade III and IV hemorrhoids.9

**References**


<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">
    <snipText>Excision is the most effective treatment for thrombosed external hemorrhoids.</snipText>
    <sor type="B">retrospective studies</sor>
  </snip>
  <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...
The XML Contents II

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 nifedipine 0.3% and lidocaine 1.5% compared with lidocaine alone. The NNT for complete pain relief at 7 days was 3.

Components of the Corpus

**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).
Contents

Evidence Based Medicine

Our Corpus for Summarisation
  Structure of our Corpus
  How we Created the Corpus
  Statistics

Applications
  Possible Uses
  Single-document Summarisation
  Evidence Grading
Annotation of Text Justifications

Goal

- Identify the text justifications.
- Align the text justifications with the answer parts.

Method

- Three annotators (members of the research group).
- Annotation tool contains pre-zoned text:
  - answer summary;
  - body text;
  - recommendations;
  - references.
- Annotators need to copy and paste (and massage) the text.
Annotation Tool I

**JFP Corpus Annotation Tool**

- **Page id**: 1080
- **Title**: What is the most effective treatment for tinea pedis athlete’s foot?
- **Authors**: Tsveti Markova, MD

**ANSWERS**

<table>
<thead>
<tr>
<th>SNIP ID</th>
<th>SNIP TEXT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Topical therapy is effective for tinea pedis. Topical terbinafine has a 70% cure rate, is available over the counter OTC, and requires only 1 to 2 weeks of therapy. Two other OTC topicals, tolnaftate and micronazole, require 2 to 4 weeks to achieve slightly lower cure rates, but are considerably less expensive.</td>
</tr>
<tr>
<td>1_1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>The most effective treatment for tinea pedis is oral terbinafine 250 mg twice a day for 2 weeks 94% clinical cure rate. However, oral terbinafine is expensive and not approved for this indication. Oral therapy may be required for patients with hyperkeratotic soles, severe disease, topical therapy failure, chronic infection or</td>
</tr>
<tr>
<td>2_1</td>
<td></td>
</tr>
</tbody>
</table>

**SOR TYPE SOR BASES REFERENCES**

- **SOR TYPE**: A
- **SOR BASES**: None
- **REFERENCES**: None
Annotation Tool II

SUMMARY

The Cochrane Database of Systemic Reviews, reported 72 placebo-controlled trials of topical agents that yielded the following cure rates: undecenoic acid, 72%; allylamine terbinafine, naftifine, butenafine, 70%; tolnaftate, 64%; azoles miconazole, clotrimazole, ketoconazole, econazole, oxiconazole, 47%. A meta-analysis of 11 RCTs suggests that allylamine are slightly more effective than azoles. (REF:1,2).

Orally administered antifungal agents are expensive and can have systemic side effects. Griseofulvin and ketoconazole are approved for oral therapy, but product labels clearly state that they should be used only after topical agents have failed. Griseofulvin has been used for more than 30 years, is well tolerated, and efficacious in treating dermatomycoses in the range of 60%. Ketoconazole’s cure rate is similar, but its use in cutaneous infections is limited by multiple drug interactions and serious side effects. Three placebo-controlled RCTs of itraconazole of varying doses and duration of treatment suggested favorable clinical cure of tinea pedis 51%-85%. The most effective itraconazole regimen was 200 mg twice daily for 1 week. In a large double-blind multicenter study of all forms of tinea pedis, De Keyser et al compared 2 weeks of terbinafine at 250 mg/day to 2 weeks of itraconazole at 100 mg/day. After 8 weeks they found itraconazole superior to terbinafine for clinical cure 94.1% vs 72.4%. In a single multicenter open study the cure rate for fluconazole 150 mg was 77% when used once weekly for 3 weeks. (REF:3,4).

RECOMMENDATIONS

American Academy of Dermatology Guidelines recommend topical therapy for initial treatment of tinea pedis. Oral therapy may be required to treat patients with hyperkeratotic soles, disabling or extensive disease, topical therapy failure, chronic infection, or immunosuppression. Surgical therapy is not indicated. (REF:5).

REFERENCES

<table>
<thead>
<tr>
<th>ID</th>
<th>PUBMED</th>
<th>CORRECT PUBMED</th>
<th>SOR TYPE</th>
<th>PUB TYPE</th>
<th>CITATION</th>
</tr>
</thead>
</table>
Conventions for text massaging

1. Remove/edit connecting phrases.
2. Remove irrelevant introductory text.
3. If a paragraph has several references, attempt to split the paragraph.
   ▶ May need to massage the text of resulting splits.
4. If a paragraph has no references, attempt to merge with previous or next paragraph.
Finding PubMed IDs

Method

1. Split the reference text into sentences.
2. Remove author and pagination text:
   - Use simple regexps.
3. Perform a sequence of searches with all combinations of sentences.
Example I


- Collins NC.
- Is ice right?
- Does cryotherapy improve outcome for acute soft tissue injury?
## Example II

<table>
<thead>
<tr>
<th>list</th>
<th>search</th>
<th>ID</th>
<th>title</th>
<th>match %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2, 3</td>
<td>Is ice right? Does cryotherapy improve outcome for acute soft tissue injury?</td>
<td>18212134</td>
<td>Is ice right? Does cryotherapy improve outcome for acute soft tissue injury?</td>
<td>92</td>
</tr>
<tr>
<td>1, 2</td>
<td>Is ice right? Does cryotherapy improve outcome for acute soft tissue injury?</td>
<td>18212134</td>
<td>Is ice right? Does cryotherapy improve outcome for acute soft tissue injury?</td>
<td>100</td>
</tr>
<tr>
<td>1</td>
<td>Is ice right?</td>
<td>None</td>
<td>None</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Does cryotherapy improve outcome for acute soft tissue injury?</td>
<td>15496998</td>
<td>Does Cryotherapy Improve Outcomes With Soft Tissue Injury? 78</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>Emerg Med J</td>
<td>None</td>
<td>None</td>
<td>0</td>
</tr>
</tbody>
</table>
Using Amazon Mechanical Turk I

**Mechanics**

- AMT was used to find the correct IDs.
- An AMT hit had 10 references:
  - 2 known references for checking quality of annotation.
- Each hit was assigned to 5 Turkers.
- There was a preliminary training session.
## Approving and rejecting hits

Reject hit if there are two or more “bad” IDs, i.e. one of:

- A known ID is wrong.
- The ID is invalid:
  - Not found in PubMed;
  - No title is returned.
- The title of the ID does not match the title of our reference:
  - threshold: 50% match.
- The ID does not agree with majority.
Using Amazon Mechanical Turk III

Checking validity for final annotation

- Majority wins automatically except when:
  - majority is a “bad” ID;
  - majority is the “nf” ID;
  - the other two are agreeing (“full house”).
- Manual check is done in all other cases.
Contents

Evidence Based Medicine

Our Corpus for Summarisation
  Structure of our Corpus
  How we Created the Corpus
  Statistics

Applications
  Possible Uses
  Single-document Summarisation
  Evidence Grading
Corpus Statistics

Size

- 456 questions ("records").
- 1,396 answers ("snips").
- 3,036 text explanations ("longs").
- 3,705 references:
  - 2,908 unique references.
  - 2,657 XML abstracts from PubMed.
Answers per Question

Avg = 3.06
Answer justifications per answer

![Graph showing Longs per Snip](image)

Avg = 2.17
References per answer justification

Avg = 1.22
References per question

Average = 6.57
Evidence Grade

SOR Types

- A
- B
- C
- D
- No type

number of SORs
References
Contents

Evidence Based Medicine

Our Corpus for Summarisation
  Structure of our Corpus
  How we Created the Corpus
  Statistics

Applications
  Possible Uses
  Single-document Summarisation
  Evidence Grading

EBM Summarisation
Diego Mollá
41/60
Contents

Evidence Based Medicine

Our Corpus for Summarisation
  Structure of our Corpus
  How we Created the Corpus
  Statistics

Applications
  Possible Uses
  Single-document Summarisation
  Evidence Grading
Evidence-based Summarisation

**Single Document Summarisation**

**Input:** Question, reference.

**Target:** Text explanation.
Evidence-based Summarisation

Single Document Summarisation

Target: Text explanation.

Multi-document Summarisation

Input: Question, group of relevant references.
Target: Answer parts (optional: plus text explanation).
Appraisal, Clustering

Text Classification for Appraisal

**Input:** Group of references.

**Target:** Evidence-based grade.
Appraisal, Clustering

Text Classification for Appraisal

**Input:** Group of references.

**Target:** Evidence-based grade.

Clustering

**Input:** Question, group of relevant references.

**Target:** Cluster groupings (optional: plus answer parts).
Retrieval?

Possible task

**Input:** Question.

**Target:** List of references.

However...
- Some of the references are old.
- The references are likely not exhaustive.
Retrieval?

**Possible task**

- **Input**: Question.
- **Target**: List of references.

**However...**

- Some of the references are old.
- The references are likely not exhaustive.
Contents

Evidence Based Medicine

Our Corpus for Summarisation
  Structure of our Corpus
  How we Created the Corpus
  Statistics

Applications
  Possible Uses
  Single-document Summarisation
  Evidence Grading
Input, Output

Input

- Question.
- Document Abstract.

Output

- Extractive summary that answers the question.
- Target summary is the annotated evidence text ("long").
- Evaluated using ROUGE-L with Stemming.
Baselines

- **plain**: Return the last $n$ sentences.
- **keywords**: Return the last $n$ sentences that share any non-stop words with the question.
- **umls**: Return the last $n$ sentences that share any UMLS concepts with the question.

<table>
<thead>
<tr>
<th>System</th>
<th>F</th>
<th>Conf Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline plain</td>
<td>0.193</td>
<td>[0.190–0.196]</td>
</tr>
<tr>
<td>baseline keywords</td>
<td>0.195</td>
<td>[0.192–0.198]</td>
</tr>
<tr>
<td>baseline umls</td>
<td>0.194</td>
<td>[0.190–0.197]</td>
</tr>
</tbody>
</table>
Using the Abstract Structure

Preselect sentences and then:

- 1. Use PubMed's section tags (background, conclusions, methods, objective, results).
- 2. Select the first $n$ sentences of the last "conclusions" section.
- 3. If we have less than $n$ sentences, fill from the first sentences of the previous "conclusions" section, and so on until all "conclusions" 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 "methods" sections.
- 6. If the abstract has no structure, return the last $n$ sentences.
Using the Abstract Structure

Preselect sentences and then:

1. Use PubMed’s section tags (background, conclusions, methods, objective, results).

<table>
<thead>
<tr>
<th>Abstract</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background</td>
<td>S1.1</td>
</tr>
<tr>
<td>Methods</td>
<td>S2.1</td>
</tr>
<tr>
<td>Results</td>
<td>S3.1</td>
</tr>
<tr>
<td>Conclusions</td>
<td>S4.1</td>
</tr>
<tr>
<td>Conclusions</td>
<td>S5.1</td>
</tr>
</tbody>
</table>

EBM Summarisation

Diego Mollá
Using the Abstract Structure

Preselect sentences and then:

1. Use PubMed’s section tags (background, conclusions, methods, objective, results).

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

<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>Methods</td>
<td>S2.1</td>
</tr>
<tr>
<td>Results</td>
<td>S3.1 S3.2</td>
</tr>
<tr>
<td>Conclusions</td>
<td>S4.1 S4.2</td>
</tr>
<tr>
<td>Conclusions</td>
<td>S5.1 S5.2</td>
</tr>
<tr>
<td>Summary</td>
<td>S5.1 S5.2</td>
</tr>
</tbody>
</table>
Using the Abstract Structure

Preselect sentences and then:

1. Use PubMed’s section tags (background, conclusions, methods, objective, results).

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

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

<table>
<thead>
<tr>
<th>Abstract</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background</td>
<td>S5.1 S5.2</td>
</tr>
<tr>
<td>Methods</td>
<td>S4.1 S4.2</td>
</tr>
<tr>
<td>Results</td>
<td>S3.1 S3.2</td>
</tr>
<tr>
<td>Conclusions</td>
<td>S4.1 S4.2</td>
</tr>
<tr>
<td>Conclusions</td>
<td>S5.1 S5.2</td>
</tr>
</tbody>
</table>
Using the Abstract Structure

Preselect sentences and then:

1. Use PubMed’s section tags (background, conclusions, methods, objective, results).
2. Select the first \( n \) sentences of the last “conclusions” section.
3. If we have less than \( n \) sentences, fill from the first sentences of the previous “conclusions” section, and so on until all “conclusions” sections are used up.
4. If we have less than \( n \) sentences, fill from the “results” sections.

<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>Methods</td>
<td>S2.1</td>
</tr>
<tr>
<td>Results</td>
<td>S3.1 S3.2</td>
</tr>
<tr>
<td>Conclusions</td>
<td>S4.1 S4.2</td>
</tr>
<tr>
<td>Conclusions</td>
<td>S5.1 S5.2</td>
</tr>
</tbody>
</table>
Using the Abstract Structure

Preselect sentences and then:

1. Use PubMed’s section tags (background, conclusions, methods, objective, results).
2. Select the first \( n \) sentences of the last “conclusions” section.
3. If we have less than \( n \) sentences, fill from the first sentences of the previous “conclusions” section, and so on until all “conclusions” 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 “methods” sections.

Abstract

<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>Methods</td>
<td>S2.1</td>
</tr>
<tr>
<td>Results</td>
<td>S3.1 S3.2</td>
</tr>
<tr>
<td>Conclusions</td>
<td>S4.1 S4.2</td>
</tr>
<tr>
<td>Conclusions</td>
<td>S5.1 S5.2</td>
</tr>
</tbody>
</table>
Using the Abstract Structure

Preselect sentences and then:

1. Use PubMed’s section tags (background, conclusions, methods, objective, results).
2. Select the first \( n \) sentences of the last “conclusions” section.
3. If we have less than \( n \) sentences, fill from the first sentences of the previous “conclusions” section, and so on until all “conclusions” 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 “methods” sections.
6. If the abstract has no structure, return the last \( n \) sentences.

<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>Methods</td>
<td>S2.1</td>
</tr>
<tr>
<td>Results</td>
<td>S3.1 S3.2</td>
</tr>
<tr>
<td>Conclusions</td>
<td>S4.1 S4.2</td>
</tr>
<tr>
<td>Conclusions</td>
<td>S5.1 S5.2</td>
</tr>
<tr>
<td>S5.1 S5.2 S4.1 S4.2 S3.1</td>
<td></td>
</tr>
</tbody>
</table>
The F is calculated using ROUGE-L with stemming.

<table>
<thead>
<tr>
<th>System</th>
<th>F</th>
<th>Conf Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline plain</td>
<td>0.193</td>
<td>[0.190–0.196]</td>
</tr>
<tr>
<td>baseline keywords</td>
<td>0.195</td>
<td>[0.192–0.198]</td>
</tr>
<tr>
<td>baseline umls</td>
<td>0.194</td>
<td>[0.190–0.197]</td>
</tr>
<tr>
<td>structure plain</td>
<td>0.196</td>
<td>[0.193–0.199]</td>
</tr>
<tr>
<td>structure keywords</td>
<td>0.193</td>
<td>[0.190–0.197]</td>
</tr>
<tr>
<td>structure umls</td>
<td>0.192</td>
<td>[0.189–0.195]</td>
</tr>
</tbody>
</table>
ROUGE-L with Stemming for All 3-Sentence Subsets

**Process**

1. Compute the ROUGE-L of all 3-sentence subsets in each abstract.
2. Find the decile boundaries in each abstract.
3. Find the distribution of decile boundaries.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.094</td>
<td>0.136</td>
<td>0.153</td>
<td>0.164</td>
<td>0.176</td>
<td><strong>0.188</strong></td>
<td>0.200</td>
<td>0.213</td>
<td>0.229</td>
<td>0.249</td>
<td>0.299</td>
</tr>
<tr>
<td>Std Dev</td>
<td>0.060</td>
<td>0.062</td>
<td>0.065</td>
<td>0.067</td>
<td>0.070</td>
<td><strong>0.073</strong></td>
<td>0.076</td>
<td>0.081</td>
<td>0.087</td>
<td>0.094</td>
<td>0.112</td>
</tr>
</tbody>
</table>
ROUGE-L with Stemming for All 3-Sentence Subsets II
Contents

Evidence Based Medicine

Our Corpus for Summarisation
  Structure of our Corpus
  How we Created the Corpus
  Statistics

Applications
  Possible Uses
  Single-document Summarisation
  Evidence Grading
ALTA 2011 Shared Task

The ALTA Shared Tasks

- Competitions where all participants are evaluated on the same data.
- The ALTA 2011 shared task was based on evidence grading.

The Data

- Clusters of abstracts.
- The SOR grade of each cluster.
Data Sample

<table>
<thead>
<tr>
<th>Fragment</th>
<th>41711</th>
<th>B</th>
<th>10553790</th>
<th>15265350</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>53581</td>
<td>C</td>
<td>12804123</td>
<td>16026213</td>
</tr>
<tr>
<td></td>
<td>53583</td>
<td>B</td>
<td>15213586</td>
<td></td>
</tr>
<tr>
<td></td>
<td>52401</td>
<td>A</td>
<td>15329425</td>
<td>9058342</td>
</tr>
</tbody>
</table>
Words as Features

Abstract $n$-grams

- Generated $n$-grams ($n = 1, 2, 3, 4$) for each of the abstracts.
- Replaced specific medical concepts with generic 'sem_type' tags using UMLS.
- Stemmed, lowercased, stop words removed.

Title $n$-grams

- Generated $n$-grams ($n = 1, 2$) for each title.
- Processed in the same way as abstract $n$-grams.
Publication Types as Features I

Distribution of publication types in a different corpus.
Publication Types as Features II

Publication types

- Rule-based classifier to detect publication types.
- Simple regular expressions that identify major publication types.
- Used the publication types marked up by PubMed when available.
- If an article has several possible publication types, choose the one with highest quality.
Cascaded Classification

Process: Cascaded SVMs

1. Default class: B.
2. SVMs with abstract $n$-grams to identify A and C.
3. SVMs with publication types to identify A and C.
4. SVMs with title $n$-grams to identify A and C.

Results

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
<th>Confidence Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority (B)</td>
<td>48.63%</td>
<td>41.5 – 55.83</td>
</tr>
<tr>
<td>Cascaded SVMs</td>
<td>62.84%</td>
<td></td>
</tr>
</tbody>
</table>
Questions?

Evidence Based Medicine

Our Corpus for Summarisation
  Structure of our Corpus
  How we Created the Corpus
  Statistics

Applications
  Possible Uses
  Single-document Summarisation
  Evidence Grading

Further Information

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