Using multimodal mining to drive clinical guidelines development

Emilie Pasche¹, Julien Gobeill², Douglas Teodoro¹, Dina Vishnyakova¹,
Arnaud Gaudinat², Patrick Ruch² and Christian Lovis¹

¹ SIMED, University of Geneva and University Hospitals of Geneva, Switzerland
² Bibliomics and Text Mining Group, University of Applied Sciences, Geneva, Switzerland

MIE 2011 - Oslo - 29th of August
Oral Presentation
Presenter: Emilie Pasche
DebugIT

Detecting and Eliminating Bacteria Using Information Technology

European project FP7 (grant #217139) with 14 partners.

Disclaimer: this presentation reflects solely the views of the authors and no guarantee or warranty is given that it is fit for any particular purpose. The European Commission, Directorate General Information Society and Media, Brussels, is not liable for any use that may be made of the information contained therein.
Why we need to create clinical guidelines?

Problem
Antibiotic resistance is increasing because of inappropriate use of antibiotics

Solution
Development of clinical guidelines can help to regulate antibiotic prescriptions
Objective: help experts to author clinical guidelines

How can we create clinical guidelines?

With KART

Without KART
Methods

How does KART work?

1. Query
   • Pattern-based query creation

2. Text-Mining
   • Rank answers using question-answering

3. Multimodal-Mining
   • Re-rank answers using source clinical data

4. Evaluation
   • Evaluate answers using IR metrics (TREC)
Step 1. Query

Manual creation of a benchmark

HUG Guidelines → Manual translation and normalization → Query

Antibiotic₁, Antibiotic₂, ... → Training set

23x

Query

Antibiotic₁, Antibiotic₂, ...

72x

Evaluation set

49x

Presented by Emilie Pasche
Step 2. Text-Mining

System architecture of Automatic Question Answering

Corpus (Medline) → Information retrieval → Relevant documents (50 docs) → Answers extraction

Search engine (easyIR, PubMed)

Terminologies (WHO-ATC)

Query

Antibiotic_1

Antibiotic_2

…
Step 3. Multimodal-Mining

Multimodal model

Query
Antibiotic_1
Antibiotic_2
...

Re-ranking

Query
Antibiotic_1
Antibiotic_3
...

Costs
(70 subst.)

Resistance profiles
Step 3. Multimodal-Mining

Getting additional features: antibiotic costs

- Prescription data (HUG)
- Data normalization
- Costs (129 prod.)
- Costs (17 subst.)
- Data completion
- Costs (70 subst.)
- Arbitrary value (0 – 100)
- Antibiotic costs

Normalisation

Costs (17 subst.)

Costs (70 subst.)
Step 3. Multimodal-Mining

Getting additional features: HUG’s resistance profiles

- Clinical Data Repository
- Extract antibiogram
- Resistance profiles
- Data completion
- Resistance profiles

SPARQL queries
(species - antibiotic)

Arbitrary value
(0 – 1)
Step 4. Evaluation

Experimental settings

<table>
<thead>
<tr>
<th>Query</th>
<th>Antibiotic_1</th>
<th>Antibiotic_2</th>
<th>…</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benchmark</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query</th>
<th>Antibiotic_1</th>
<th>Antibiotic_3</th>
<th>…</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatically-generated</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Evaluation

TREC-EVAL

Results

Automatically-generated
### Results

**How well does KART perform?**

<table>
<thead>
<tr>
<th></th>
<th>Answers</th>
<th>Top precision</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antibiotic costs:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline (easyIR)</td>
<td>49/49</td>
<td>34.28%</td>
</tr>
<tr>
<td>Baseline (PubMed)</td>
<td>32/49</td>
<td>40.37%</td>
</tr>
<tr>
<td>Costs (easyIR)</td>
<td>49/49</td>
<td>43.31%</td>
</tr>
<tr>
<td>Costs (PubMed)</td>
<td>32/49</td>
<td>40.28%</td>
</tr>
<tr>
<td><strong>Resistance profile:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline (easyIR)</td>
<td>49/49</td>
<td>39.86%</td>
</tr>
<tr>
<td>Baseline (PubMed)</td>
<td>32/49</td>
<td>56.41%</td>
</tr>
</tbody>
</table>

- EAGLi/easyIR: + 9%
- PubMed: - 0.1%
- EAGLi/easyIR: + 5.5%
- PubMed: + 16%
Limits and future works

Costs

- Currently based on a limited set of costs
  - HUG costs list (17/70 substances)
- We could use broader resources
  - Swiss Kompendium (all substances)

Resistance

- Currently based on species-specific antibiograms
  - E.g. 2 antibiograms for S. pyogenes + clindamycin
- We could use aggregated species
  - E.g. 75 antibiograms for all Streptococcus + clindamycin
Conclusion

• Facilitates clinical guidelines development by extracting hypothetical treatments from literature
  – *E.g. Pneumonia and Streptococcus pneumoniae*
    • 4855 publications in MEDLINE
    • 12 proposed antibiotics in KART

• Combining literature-based discovery with clinical data mining can significantly improve authoring of clinical guidelines
  – 56% of top-ranked answers are correct
Acknowledgments

DebugIT, EU-IST-FP7-217139
EAGL, SNF-325230-120758

Infectious disease service (HUG)
• Angela Huttner
• Marina Macedo
• Thomas Haustein
• Stephan Harbarth

Consultant Physician (Australia)
• Garry Lane

KART: http://eagl.unige.ch/KART/
EAGLi: http://eagl.unige.ch/EAGLi/

DebugIT partners
• Agfa Healthcare (Belgium)
• Empirica (Germany)
• Gama Sofia Ltd (Bulgaria)
• INSERM (France)
• IZIP (Czech Republic)
• Linköping University (Sweden)
• TEILAM (Greece)
• University College London (UK)
• HUG (Switzerland)
• Freiburg University (Germany)
• Geneva University (Switzerland)
• Averbis (Germany)
• MDA (Czech Republic)
• HES-SO (Switzerland)
Thank you for your attention

emilie.pasche@unige.ch