A Simple Approach to NTCIR-10 MedNLP Task

Yuka Tateisi
National Institute of Informatics

Takashi Okumura
National Institute of Public Health

• Background
  – Few NLP resources are available to everyone
  – Doctors may have local, customized dictionaries or terminologies

• Purpose
  – Evaluate one of such terminologies against a publicly-available test data
  – Establish a method for NLP non-specialists to test their resources

• Method

Input

2028年11月頃より通院勤務、胸痛出現。
父：脇痛
胸腹部造影CTではその他明らかな腫瘍性病変、リンパ節腫大は認められなかった。

McCab + User Dictionary

Disease Master (1312 Entries)†
Symptom Master (861)
Lab Test Manual Data (2492)‡

Heuristics

Output

2028年11月頃より通院勤務、胸痛出現。
父：脇痛
胸腹部造影CTではその他明らかな腫瘍性病変、リンパ節腫大は認められなかった。

• Test Results
  – Entities (<c> elements) in the training set were further added to the user dictionary

Elements in the Test Data Recognized by the Method

<table>
<thead>
<tr>
<th>Source</th>
<th>Count</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease Master</td>
<td>127</td>
<td>12.7</td>
</tr>
<tr>
<td>Symptom Master</td>
<td>276</td>
<td>27.5</td>
</tr>
<tr>
<td>Lab Test Manual</td>
<td>147</td>
<td>14.7</td>
</tr>
<tr>
<td>Training Set</td>
<td>454</td>
<td>45.2</td>
</tr>
<tr>
<td>Total &lt;c&gt; elements found</td>
<td>1004</td>
<td>100</td>
</tr>
</tbody>
</table>

• Observations
  – Even a relatively small dictionary can help
  – Rule-based word-formation analysis would improve the result

Results on the Training Set

<table>
<thead>
<tr>
<th>Type</th>
<th>Count</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete Match</td>
<td>697</td>
<td>36.3</td>
</tr>
<tr>
<td>Partial Match</td>
<td>352</td>
<td>18.3</td>
</tr>
<tr>
<td>No Match</td>
<td>873</td>
<td>45.4</td>
</tr>
</tbody>
</table>

Failure Cases in the Training Set

• Nonspecific terms
  – anomalies
  – lesion
    e.g. <c>異常所見</c>, <c>病変</c>

• Findings with the modifier that specify the location of bodies (partial match)
  – edema in both legs
    e.g. <c>両下肢浮腫</c>
    *浮腫(edema) in the dictionary

† Compiled by Dr. K. Torigoe, available from I’ROM Holdings Co. Ltd. (http://www.irom-hd.co.jp/) under BSD license.