InteractiveMediaMINE at the NTCIR-11 IMine Search Task

Shohei MINE, Takuma MATSUMOTO, Tomofumi YOSHIDA, Takuya SHINOHARA, Daisuke KITAYAMA
Graduate School of Engineering, Kogakuin University

Background
TaskMine aims to explore the methods of automatically finding subtasks of a given task
※We use Yahoo! Chiebukuro(web-based Q&A service) as our system resource
• In general, Q&A services aim to collect answers that solve the user’s problems
• We expect Yahoo! Chiebukuro to be useful as our system resource for mining tasks

Step 1. Extending Query
1. Executing a morphological analysis for the inputted query by user, and extracts nouns and verbs from the query
2. Joining the extracted nouns, a single-byte blank, and verbs to make a sentence
3. Adding “方法(hoho)” at the end of the query
※方法 is a word that means method or way in Japanese, and it is effective in retrieving pages that include questions about methods used for solving certain problems

Step 2. Retrieving
1. Using Yahoo! Chiebukuro with the extended search query
2. Retrieving the top 10 pages of the search results

Step 3. Extracting Tasks
1. Performing dependency parsing for all the extracted sentences in order to extract a pattern
   Extracted verb end of the sentence are converted into its dictionary form
2. Extracting chunks that contain the following pattern:
   Chunks that depend on noun(a) + noun(b) + “を(wo)”(c)
   + the chunk that includes certain verb(d) + chunks that depend on the verb(e)
3. Extracting sentences as tasks
※Considering “を(wo)” is a Japanese particle pointing to a direct object, it is useful to extract tasks that can solve the user’s problem

Step 4. Ranking
• Based on the supposition that the words that appeared frequently are important in solving problems
• We define the evaluation scores as follows:

\[
\text{Score}(t_i) = \sum_{\text{noun} \in \text{nouns}} \sum_{\text{ans} \in \text{A}} \text{NounFreq}(\text{noun}, \text{ans}) + \sum_{\text{task} \in \text{T}} \text{VerbFreq}(\text{verb}, \text{task})
\]

<table>
<thead>
<tr>
<th>Noun</th>
<th>Frequency</th>
<th>Verb</th>
<th>Frequency</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>米</td>
<td>4</td>
<td>2</td>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td>水</td>
<td>4</td>
<td>1</td>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td>鍋</td>
<td>3</td>
<td>1</td>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td>火</td>
<td>2</td>
<td>1</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>分量</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>10</td>
</tr>
</tbody>
</table>

Experimental
• We mined tasks for the query set of the TaskMine subtask
• Dataset is 50 tasks

Results
nDCG is computed for a variety of cutoff thresholds k(=1,5,10,50)
→InteractiveMediaMINE perform better than the organizer’s baseline method among all the runs

Our system makes it easy to collect answers accurately for ordinary questions
TM-019,歯周病を治療する, has one of the highest scores for all metrics, but extracted task is long sentence as table below
※ “I want to use the laser cutter”, in English
→Yahoo! Chiebukuro is a general Q & A service that is not specialized in any specific field