SML Question-Answering System for World History Essay Exams at NTCIR-13 QALab-3

Team: SML
Yusuke Doi, Takuma Takada, Takuya Matsuzaki, Satoshi Sato
Graduate School of Engineering, Nagoya University
Our Target

Question Types
1. Multiple-choice question
2. Term question
3. Essay question
   • Long essay (in 450-600 characters)
   • Short essay (in 30-120 characters)

Tasks for Essay question
1. End-to-End task
2. Extraction task
3. Summarization task
4. Evaluation method task
Our Target

Question Types
1. Multiple-choice question
2. Term question
3. Essay question
   • Long essay (in 450-600 characters)
   • Short essay (in 30-120 characters)

Tasks for Essay question
1. End-to-End task
2. Extraction task
3. Summarization task
4. Evaluation method task
Question

Describe, in no more than 30 English words, the process by which the polis were formed.

Model Answer

Under the leadership of powerful nobles, various settlements formed coalitions, and people lived together around the Acropolis, forming poleis.

( The University of Tokyo, 2009 )
Phase 2 results

Number of short essay questions is 22

<table>
<thead>
<tr>
<th></th>
<th>Nuggets</th>
<th>ROUGE-1</th>
<th>ROUGE-2</th>
<th>ROUGE-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run-1</td>
<td>7 / 80</td>
<td>0.313</td>
<td>0.088</td>
<td>0.038</td>
</tr>
<tr>
<td>Run-2</td>
<td>7 / 80</td>
<td>0.312</td>
<td>0.091</td>
<td>0.039</td>
</tr>
</tbody>
</table>

All team’s ROUGE-2 scores

1. Forst  0.107
2. SML    0.091
3. KSU    0.072
4. IMTKU  0.052
Our System

**Extraction Module**
- Identify theme and focus
- Extract sentences from glossary

**Compression Module**
- Optimization-based method
- Rule-based method

**Answer**
Q. Describe, in no more than 30 English words, the content of the **Monroe Doctrine**.

**Monroe Doctrine**
- **Definition**: The Monroe Doctrine was a United States ...
- **Content**: It stated that further efforts by European ...
- **Content**: At the same time, the doctrine noted that ...

**Theme**: Monroe Doctrine  
**Focus**: Content
Com Compression Module

**Optimization-based Method**  [Morita et al. 2011, 2013]

Repeatedly add valid subtrees to answer so that gain of objective function is maximized

Objective function \( f(S) = \sum_{w \in \text{words}(S)} \left( qsb(w) \sum_{i=0}^{\text{count}_S(w)-1} d^i \right) + \gamma \cdot \text{reward}(S) \)

**Rule-based Method**

1. Extend extracted sentences by concatenating each pair
2. Sort sentences by QSB score
3. Compress sentence in turn by compression rules

\[
\text{Compress } 1. \quad s1 \quad s3 \quad \rightarrow \quad \text{too long} \\
2. \quad s1 \quad s2 \quad \rightarrow \quad \text{too long} \\
3. \quad s1 \quad \rightarrow \quad \text{OK} \quad \rightarrow \quad \text{Output}
\]

\( qsb(w) \), query relevance score
Phase 2 results

Run-1  Rule-based method
Run-2  Optimization-based method

Number of short essay questions is 22.

<table>
<thead>
<tr>
<th></th>
<th>Nuggets</th>
<th>ROUGE-1</th>
<th>ROUGE-2</th>
<th>ROUGE-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run-1</td>
<td>7 / 80</td>
<td>0.313</td>
<td>0.088</td>
<td>0.038</td>
</tr>
<tr>
<td>Run-2</td>
<td>7 / 80</td>
<td>0.312</td>
<td>0.091</td>
<td>0.039</td>
</tr>
</tbody>
</table>

There was no significant difference between the performances of two compression methods.
Analysis

In 18/22 question, the extracted sentences included none of the nuggets

In most cases, the theme was wrongly identified

Q. During the middle of the Former Han era Confucianism, which up until that point had been merely one of several valid schools of thought, was given a special position of prominence, separate from other schools of thought. Explain, in 30 English words or less, what event led to this.

“Confucianism” was not extracted as theme because it didn’t match the rule for identifying theme
Summary

• We need to detect the theme more accurately.
• There was no significant difference between the performances of two compression methods.
• But the two compression methods may work differently when extraction module is improved.