KIT Dialogue System for NTCIR-13 STC Japanese Subtask

Team KIT16
Kyoto Institute of Technology

Hiroshi Nakatani, Shigenori Nishiumi
Takahiro Maeda, Masahiro Araki
Three Methods

● Retrieval-based method
  1. TF-IDF / word2vec method
     - similar word and related word
  2. Clustering utterance with CRP method
     - relationship between utterance

● Generation-based method
  3. Seq2seq method
     - It can generate replies repository does not have
Retrieval-based method with TF-IDF and word2vec

- Flow of this method

![Flowchart diagram]
Coherence Evaluation

- Filtering based on proper nouns
  - Proper nouns by word2vec based on input proper nouns.
  - If other proper nouns are in reply, then remove it.

- Scoring using TF-IDF
  
  Calculating cosine distance between
  1. [input data] and [reply],
  2. [input data] and [comment + reply]

  \[
  \text{Score} = 1 \times 2
  \]
Context-dependence and Informativeness Evaluation

- **Scoring using adjectives and adverbs**
  - If a comment includes the same adjectives (or adverbs) as input data, then the score is increased.

  
  \[
  \text{Score} \gets \text{Score} + Pt_A + Pt_B
  \]

- **Scoring using nouns**
  - If a reply includes nouns (obtained by word2vec based on input data), then the score is increased.

  \[
  \text{Score} \gets \text{Score} + Pt_A + Pt_{B-1}
  \]
Retrieval-based method with Clustering utterance with CRP

- **Dialogue-act** clustering
  By learning the tendency of dialogue-act in pairs in the repository, we estimate which dialogue-act should be used in response to a new given utterance.

- **Topic** clustering
  By retrieving reply that is same cluster to the input comment, the reply is same topic to the comment.

By using these clustering, a natural response is selected.
## Clustering

<table>
<thead>
<tr>
<th>Feature</th>
<th>Topic Clustering</th>
<th>Dialogue-act Clustering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature</td>
<td>BoW (noun, adjective, verb)</td>
<td>BoW (symbol, emotion, filler, adverb, particle, conjunction, adnominal)</td>
</tr>
<tr>
<td>Method</td>
<td>CRP</td>
<td>CRP</td>
</tr>
<tr>
<td>Number of Clusters</td>
<td>722</td>
<td>74</td>
</tr>
</tbody>
</table>
Scoring

Comment text and Reply text of training data

Chinese Restaurant Process

Topic Clusters \((C_T, R_T)\)

Dialogue-act Clusters \((C_D, R_D)\)

Theme

Category

Scoring system

\[
\text{Score}(c,r) = (\text{effective}(C_T, R_T) + \text{effective}(C_D, R_D)) \times 1.1 \times \text{count Theme}(c,r) \times 1.1 \times \text{flag Category}(c,r)
\]

\[
\text{effective}(C,R) = \frac{\text{count}(C,R)}{|R|}
\]
Generation-based method with seq2seq model

We add the salient **keyword of the news** to the comment.
Keyword selection

- Keyword is a word that express the news.
- Keyword is a **proper noun** that first comes in titles in Yahoo! Topics.

Keyword examples
- **America**
- **YouTuber**
- **Hakone Ekiden**
- **SMAP×SMAP**
- *etc.*
seq2seq model training

Keyword is added to the end of comment.
Evaluation Results

Retrieval-based method with TF-IDF and word2vec achieved the best results in our systems.

<table>
<thead>
<tr>
<th>method</th>
<th>Mean AccL2@1 (Rule-1)</th>
<th>Mean AccL1,L2@1 (Rule-1)</th>
<th>Mean AccL1,L2@1 (Rule-2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOLD</td>
<td>0.4720</td>
<td>0.8980</td>
<td>0.8660</td>
</tr>
<tr>
<td>TF-IDF/word2vec</td>
<td>0.1800</td>
<td>0.8240</td>
<td>0.6320</td>
</tr>
<tr>
<td>CRP</td>
<td>0.0860</td>
<td>0.4660</td>
<td>0.3840</td>
</tr>
<tr>
<td>seq2seq</td>
<td>0.0960</td>
<td>0.6320</td>
<td>0.4680</td>
</tr>
</tbody>
</table>
Conclusion

We addressed on the STC Japanese Subtask from three approaches:

1. **TF-IDF / word2vec** method
2. **Clustering utterance with CRP** method
3. **seq2seq** model method

As a result, Retrieval-based method with **TF-IDF and word2vec** showed the best results.

The major cause of the result is that TF-IDF can select similar replies to the comment.