Scoring of response based on suitability of dialogue-act and content similarity

Team KIT15

Kyoto Institute of Technology
Sota Matsumoto, Masahito Araki
Short Text Conversation

- Repository of pairs consisting of a post and a comment to it
- When an utterance is given, we search for the appropriate utterances as the response in there
Score the utterances of the repository in following 2 items

- **Interactive Functional Suitability**
  Suitability in function of conversation as response to the input utterance

- **Content Similarity**
  Similarity of topics between utterances
  *e.g. soccer, movie, lecture of university, etc.*

\[
\text{Score}(p, t_a) = \text{ifs}(p, t_a) \times \text{csim}(p, t_a)
\]
Interactive Functional Suitability

- **Dialogue-act**
  - The function of utterance in conversation
    e.g. greeting, question, desire, etc.
  - To design it for the domain automatically, we use the **Chinese Restaurant Process (CRP)**
    - one of the unsupervised classification method
    - So, the names of dialogue-acts are those that we named classified clusters later
Interactive Functional Suitability

- Learn the tendency of dialogue-acts used in pairs of the repository as weight table

\[ W[i][j] = \frac{\text{count}(i, j)}{N} \]

- Using weight table, it’s possible to determine the suitability of utterances in the repository

\[ \text{if s}(p, t_a) = W[\text{dae}(p)][\text{dae}(t_a)] \]

\( \text{dae}(*) \) is dialogue-act estimator
Content Similarity

- **Latent Dirichlet Allocation (LDA)**
  estimate the potential topics to which the document belongs

  These words belong to the same topic

- **Inverse Document Frequency (IDF)**
  search documents with common words having high informativeness

\[
csim(p, t_a) = \alpha \ast lsim(p, t_a) + (1 - \alpha) \ast isim(p, t_a)
\]
Experiments

Data

We used the posts to Twitter in training and testing

- Training data: 822,254 posts (411,127 pairs)
- Testing data: 202 posts
Experimental procedures

- Training of models

  - CRP
    - The feature is bag-of-words whose words with a frequency of appearance of more than 1,000 times

  - LDA
    - To train the topic model we used articles of the free web encyclopedia “Hatena Keyword”
      - The feature is bag-of-words for the noun
Experimental procedures

■ Evaluations

- Three-degree evaluation of 0-2 according to the appropriateness of the response
- Calculate evaluation values in 4 types from “Case X-Y” that is a combination of the two conditions of X and Y
  - X is a set of evaluation values to determine that the response is appropriate (2 or 12)
  - Y is the lowest rank number of the candidates evaluated in each utterance (1 or 5)
- Finally the mean values of the evaluation values for each utterance were calculated
Results

- **Clustering of Dialogue-act**
  - Training data were classified into 41 dialogue-acts
  - About half of the data belonged to one cluster
  - About 70% of data belonged to the clusters mainly focusing case particles that are not related to the function of conversation
  - In this experiment, IFS had the function as a filtering special representation
  - As an exception, the cluster that seems to be greeting worked well because it had strong bias of using
**Evaluation Results**

- $\alpha$ is the parameter for adjusting the ratio of LDA and IDF in content similarity
- Higher evaluation in the case of not using the LDA

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>case 2-1</th>
<th>case 2-5</th>
<th>case 12-1</th>
<th>case 12-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.2297</td>
<td>0.2050</td>
<td>0.5589</td>
<td>0.5380</td>
</tr>
<tr>
<td>0.4</td>
<td>0.1817</td>
<td>0.1743</td>
<td>0.4748</td>
<td>0.4535</td>
</tr>
<tr>
<td>0.5</td>
<td>0.1812</td>
<td>0.1660</td>
<td>0.4614</td>
<td>0.4317</td>
</tr>
<tr>
<td>1.0</td>
<td>0.0787</td>
<td>0.0787</td>
<td>0.2114</td>
<td>0.2130</td>
</tr>
</tbody>
</table>
Consideration and Future works

- Including case particles to factor of classification is one of reason that inhibited performance in Interactive Functional Suitability

- We will improve it by filtering those

- We found that the filtering of words other than the noun didn't work well by inadequate performance of the morphological analysis

- By correcting them and increasing the number of dimensions of the topic vector, we will improve the ability to respond to the topic
Thank you