Dr. Chih-Chien Wang
Dr. Min-Yuh Day
Mr. Wei-Jin Gao
Mr. Yen-Cheng Chiu
Ms. Chun-Lian Wu

National Taipei University
Tamkang University
Taipei, Taiwan
Overview

Retrieval based Method

Solr search engine + Similarity

Generative Model

Short Text Generation

Emotion Classification model

Generative Model + General Purpose Response

Generation Purpose Response
Retrieval Based
Search responses from corpus.
Overview of Retrieval-based Method

- We used Solr to index the corpus.
- Before indexing it, we perform word segmentation, text analysis, and remove stop words.
- Then, we complete the Solr index building.
Retrieval-based Method: Search the new post

- When a new post provided, we searched the Solr index, and obtain the fetched potential candidate comments.
- We used all terms (words) from the provided new post one by one to search the Solr.
- If the term appeared in the post of post-comment pair, we fetched the “comment” (rather than post) as potential candidates for generated comments.
- Keep the first 500 search results
We calculated the accumulated inverse term frequency.

We computed the cosine similarity between the new post and the candidate comments.

We multiplied accumulated inverse term frequency by cosine similarity as the relevance score.

The candidate comment that match the assigned emotion and with highest relevance score was treated as the generated comment.
### Evaluation Results

<table>
<thead>
<tr>
<th>Result</th>
<th>Submission</th>
<th>Method</th>
<th>Label 0</th>
<th>Label 1</th>
<th>Label 2</th>
<th>Total</th>
<th>Overall score</th>
<th>Average score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation result</td>
<td>RUN 1</td>
<td>Retrieval</td>
<td>716</td>
<td>200</td>
<td>84</td>
<td>1000</td>
<td>368</td>
<td>0.368</td>
</tr>
</tbody>
</table>
Only 3 teams submit for retrieval based method

**Table 5.** The result of the overall score and average score.

<table>
<thead>
<tr>
<th>Team Name</th>
<th>Label 0</th>
<th>Label 1</th>
<th>Label 2</th>
<th>Total</th>
<th>Overall score</th>
<th>Average score</th>
</tr>
</thead>
<tbody>
<tr>
<td>AINTPU_1</td>
<td>716</td>
<td>200</td>
<td>84</td>
<td>1000</td>
<td>368</td>
<td>0.368</td>
</tr>
<tr>
<td>IMTKU_1</td>
<td>580</td>
<td>248</td>
<td>172</td>
<td>1000</td>
<td>592</td>
<td>0.592</td>
</tr>
<tr>
<td>WUST_1</td>
<td>601</td>
<td>211</td>
<td>188</td>
<td>1000</td>
<td>587</td>
<td>0.587</td>
</tr>
</tbody>
</table>
Weakness of our retrieval method

We do not used semantic analysis before searching

• We used only the terms in the new post to search the results.
• We should also used similar term with similar meaning to search the corpus.

Emotion Categories

• We do not consider the noisy of emotion classification. We realize the precision issue of emotion categories after receiving the evaluation results.
**Evaluations**  | Retrieval-based Method

<table>
<thead>
<tr>
<th>Evaluation Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Result</strong></td>
</tr>
<tr>
<td>Evaluation result</td>
</tr>
</tbody>
</table>

We realize the precision issue of emotion categories after receiving the evaluation results. Only 30% (84/284) response were with correct emotion.

According to the organizers, the accuracy rate for emotion classification was 62% in their NLPCC papers. The actual accuracy rate may be lower than that.
Generative Approach
Automatically generate responses to questions
Generative Approach

Generative Model

- Short Response Generation
- Emotion Classification model
Generative Models

Automatically Generated Response in Short text conversion

We employed an attention-based sequence to sequence (Seq2Seq) network model for the generation-based approach.
Generative Models | Generation-based Method

Generate Short Responses to the Dialogue

Seq2Seq with attention mechanism
Long Short Term Memory (LSTM) as encoder and decoder
Before training the model, we perform word segmentation, text analysis, and remove stop words.
Then, we used an attention-based sequence to sequence (Seq2Seq) network model which take Long Short Term Memory (LSTM) as encoder and decoder to train the model using the provided corpus.
We compared the different methods of MLP/GRU/LSTM/BiGRU/BiLSTM for developing emotion classification.

We performed preprocessing, label indexing, one-hot encoding, and training to train emotion classification model.
Deep learning approach of Emotion Classification model

- MLP, GRU, LSTM, BiGRU, and BiLSTM

Evaluations of all deep learning approaches

<table>
<thead>
<tr>
<th>DL model</th>
<th>Batch size</th>
<th>Dropout</th>
<th>Epochs</th>
<th>Accuracy</th>
<th>Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>BiGRU</td>
<td>256</td>
<td>0.5</td>
<td>15</td>
<td>0.880</td>
<td>0.333</td>
</tr>
<tr>
<td>BiLSTM</td>
<td>256</td>
<td>0.4</td>
<td>10</td>
<td>0.879</td>
<td>0.335</td>
</tr>
<tr>
<td>LSTM</td>
<td>256</td>
<td>0.1</td>
<td>20</td>
<td>0.879</td>
<td>0.335</td>
</tr>
<tr>
<td>GRU</td>
<td>256</td>
<td>0.4</td>
<td>20</td>
<td>0.872</td>
<td>0.356</td>
</tr>
<tr>
<td>MLP</td>
<td>256</td>
<td>0.4</td>
<td>30</td>
<td>0.843</td>
<td>0.451</td>
</tr>
</tbody>
</table>
Confusion matrix for emotion classification

Best Method
Bi-GRU
We computed the cosine similarity between the new post and the generated candidate comments. The candidate comment that with highest cosine similarity with question was treated as the generated comment.
<table>
<thead>
<tr>
<th>Emotion classification</th>
<th>Label0</th>
<th>Label1</th>
<th>Label2</th>
<th>Total</th>
<th>Overall core</th>
<th>Average score</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLP</td>
<td>873</td>
<td>85</td>
<td>42</td>
<td>200</td>
<td>169</td>
<td>0.169</td>
</tr>
<tr>
<td>GRU</td>
<td>855</td>
<td>69</td>
<td>76</td>
<td>1000</td>
<td>221</td>
<td>0.221</td>
</tr>
<tr>
<td>BiGRU</td>
<td>860</td>
<td>72</td>
<td>68</td>
<td>1000</td>
<td>208</td>
<td>0.208</td>
</tr>
<tr>
<td>LSTM</td>
<td>864</td>
<td>65</td>
<td>71</td>
<td>1000</td>
<td>207</td>
<td>0.207</td>
</tr>
<tr>
<td>BiLSTM</td>
<td>857</td>
<td>84</td>
<td>59</td>
<td>1000</td>
<td>202</td>
<td>0.202</td>
</tr>
</tbody>
</table>

Use MLP to automatically generate responses.
**Self-Evaluation Performance**

The emotion precision rate was only around 50%.

<table>
<thead>
<tr>
<th>Emotion classification</th>
<th>Label0</th>
<th>Label1</th>
<th>Label2</th>
<th>Total</th>
<th>Overall core</th>
<th>Average score</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLP</td>
<td>873</td>
<td>85</td>
<td>42</td>
<td>200</td>
<td>169</td>
<td>0.169</td>
</tr>
<tr>
<td>GRU</td>
<td>855</td>
<td>69</td>
<td>76</td>
<td>1000</td>
<td>221</td>
<td>0.221</td>
</tr>
<tr>
<td>BiGRU</td>
<td>860</td>
<td>72</td>
<td>68</td>
<td>1000</td>
<td>208</td>
<td>0.208</td>
</tr>
<tr>
<td>LSTM</td>
<td>864</td>
<td>65</td>
<td>71</td>
<td>1000</td>
<td>207</td>
<td>0.207</td>
</tr>
<tr>
<td>BiLSTM</td>
<td>857</td>
<td>84</td>
<td>59</td>
<td>1000</td>
<td>202</td>
<td>0.202</td>
</tr>
</tbody>
</table>

Use MLP to automatically generate responses.
General Purpose Response
Generate responses when we do not know how to answer the questions
we used General Purpose Response (GPR) to improve the generative-based response performance. About 1500 general purpose responses were created.

The generated comments will be replaced by the GPR at filter stage if the new post and generated comments received a low relevance score computed by cosine similarity (about 30%).
Use MLP plus GPR to automatically generate responses

<table>
<thead>
<tr>
<th>Emotion classification</th>
<th>Label0</th>
<th>Label1</th>
<th>Label2</th>
<th>Total</th>
<th>Overall core</th>
<th>Average score</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLP</td>
<td>808</td>
<td>124</td>
<td>68</td>
<td>1000</td>
<td>260</td>
<td>0.26</td>
</tr>
<tr>
<td>GRU</td>
<td>756</td>
<td>77</td>
<td>167</td>
<td>1000</td>
<td>411</td>
<td>0.411</td>
</tr>
<tr>
<td>BiGRU</td>
<td>727</td>
<td>111</td>
<td>162</td>
<td>1000</td>
<td>435</td>
<td>0.435</td>
</tr>
<tr>
<td>LSTM</td>
<td>749</td>
<td>89</td>
<td>162</td>
<td>1000</td>
<td>413</td>
<td>0.413</td>
</tr>
<tr>
<td>BiLSTM</td>
<td>753</td>
<td>75</td>
<td>172</td>
<td>1000</td>
<td>419</td>
<td>0.419</td>
</tr>
</tbody>
</table>
### With or Without GPR

Use MLP to automatically generate responses

<table>
<thead>
<tr>
<th>Emotion classification</th>
<th>With GPR Average score</th>
<th>Without GPR Average score</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLP</td>
<td>0.26</td>
<td>0.169</td>
<td>+0.091</td>
</tr>
<tr>
<td>GRU</td>
<td>0.411</td>
<td>0.221</td>
<td>+0.190</td>
</tr>
<tr>
<td>BiGRU</td>
<td>0.435</td>
<td>0.208</td>
<td>+0.227</td>
</tr>
<tr>
<td>LSTM</td>
<td>0.413</td>
<td>0.207</td>
<td>+0.216</td>
</tr>
<tr>
<td>BiLSTM</td>
<td>0.419</td>
<td>0.202</td>
<td>+0.217</td>
</tr>
</tbody>
</table>
Overview of Generative based Method

Emotion Classification model

- Corpus
  - Pre-processing
  - Label index
  - One-hot encoding
  - Training

Generation model

- Corpus
  - New post
  - Pre-processing
  - Remove stop word

Text analysis

- Generation model training
  - Well-trained Model (LSTM)

Cosine similarity analysis

- Ranking
  - Candidate results

General Purpose Response

- GPR Corpus
  - Cosine similarity analysis
  - Filter
  - Results
Conclusion

Comparison between methods
• Performance of Retrieval-based model is better than Generative model
  - However, use different approach of deep learning in Emotion Classification model will have different kinds of improvement
  - Further more, use EGPR can make performance more close to retrieval-based model

Evaluation of Emotion Classification model
• BiGRU > BiLSTM > LSTM > GRU > MLP
Future work

1. conversation model
   - use seqGAN as deep learning neural network of generative model
   - try to add topic layer between encoder and decoder of S2S architecture

2. EGPR
   - take more general condition to expand EGPR dataset

3. Emotion Classification model
   - Bidirectional Encoder Representation from Transformers (BERT) to improve the performance of emotion classification model