IMNTPU Dialogue System Evaluation at the NTCIR-16 DialEval-2 Dialogue Quality and Nugget Detection

1Information Management, National Taipei University, New Taipei City, Taiwan
2Zeals Co., Ltd. Tokyo, Japan

Ting-Yun Hsiao1  Yung-Wei Teng1  Pei-Tz Chiu1  Mike Tian-Jian Jiang2  Min-Yuh Day1,*

myday@gm.ntpu.edu.tw

NTCIR-16 Conference, June 14-17, 2022, Tokyo, Japan
Outline

• IMNTPU Research Architecture
• Proposed Approach
  • Tokenization Tricks
  • Fine-tuning Techniques
• Results
• Conclusion
Highlights

- IMNTPU (Information Management at National TaiPei University) submitted one run for Chinese DQ, English DQ and English ND.
- Dialogue System Evaluation at the NTCIR-16 DialEval-2 Dialogue Quality (DQ) and Nugget Detection (ND).
IMNTPU Research Architecture for NTCIR-16 DialEval-2

Pre-trained Model

XLM-RoBERTa

Tokenization Tricks

Fine-tuning Techniques

Discriminative Fine-tuning

One-cycle Policy

Optimization
IMNTPU Research Architecture for NTCIR-16 DialEval-2

XLM-RoBERTa
Conneau et al. (2019)

- Retains subword token letter cases
- Integrates and improves approaches from Cross-lingual Language Model (XLM) pretraining schemes
- Tokenize with unigram-level sentence piece

Dynamical MLM
Input text
XLM  Unigram
Since there is no lunar calendar in the phone's calendar, I installed a new calendar application, but the date displayed is different. @Smartisan Customer Service

Hello, the problem of not displaying the lunar calendar in the view of the built-in calendar month will be updated in the later version. The external version of the calendar cannot display the dynamic icon at present.

I see. Thank you!
Tokenization Tricks

length  position

xxlen 3 <s> xxtrn 1 xxsd customer Since there is no lunar calendar in the phone's calendar, I installed a new calendar application, but the date displayed is different. @Smartisan Customer Service </s> </s> xxtrn 2 xxsd helpdesk Hello, the problem of not displaying the lunar calendar in the view of the built-in calendar month will be updated in the later version. The external version of the calendar cannot display the dynamic icon at present. </s> </s> xxtrn 3 xxsd customer I see. Thank you! </s>
Tokenization Tricks

Since there is no lunar calendar in the phone's calendar, I installed a new calendar application, but the date displayed is different. @Smartisan Customer Service

Hello, the problem of not displaying the lunar calendar in the view of the built-in calendar month will be updated in the later version. The external version of the calendar cannot display the dynamic icon at present.

I see. Thank you!
Fine-tuning Techniques

- Discriminative Fine-tuning
- One-cycle Policy
- Optimization
Fine-tuning Techniques

1 Discriminative Fine-tuning

• Different layers capture different types of information. They should be fine-tuned to different extents.

• The initial layers capture the most general form of information.

• General information of language are common and would require the least changes in their weights.

• The amount of fine-tuning required increases gradually as we move towards the last layer.
Fine-tuning Techniques

One-cycle Policy

- Slanted triangular learning rates
  - Intuition for adapting parameters to task-specific features.
  - The model should converge quickly to a suitable region and then refine its parameters.

Howard & Ruder (2018)
## Results of IMNTPU at NTCIR-16 DialEval-2
Chinese Dialogue Quality

### NTCIR-16 DialEval-2 Chinese Dialogue Quality (DQ) Test set

<table>
<thead>
<tr>
<th>Model</th>
<th>A-score</th>
<th>S-score</th>
<th>E-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMNTPU-run0</td>
<td>RSNOD</td>
<td>0.2479</td>
<td>0.1618</td>
</tr>
<tr>
<td>Baseline-run0</td>
<td>0.2301</td>
<td>0.1772</td>
<td>0.1998</td>
</tr>
</tbody>
</table>

### NTCIR-16 DialEval-2 Chinese Dialogue Quality (DQ) Development set

<table>
<thead>
<tr>
<th>Model</th>
<th>A-score</th>
<th>S-score</th>
<th>E-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMNTPU-run0</td>
<td>0.2262</td>
<td>0.1495</td>
<td>0.2076</td>
</tr>
</tbody>
</table>

NTCIR-16 Conference, June 14-17, 2022, Tokyo, Japan
## Results of IMNTPU at NTCIR-16 DialEval-2 English Dialogue Quality

<table>
<thead>
<tr>
<th>NTCIR-16 DialEval-2 English Dialogue Quality (DQ) Test set</th>
<th>A-score</th>
<th>S-score</th>
<th>E-score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
<td>RSNOD</td>
<td>NMD</td>
<td>RSNOD</td>
</tr>
<tr>
<td>IMNTPU-run0</td>
<td>0.2535</td>
<td>0.1654</td>
<td>0.2020</td>
</tr>
<tr>
<td>Baseline-run0</td>
<td>0.2321</td>
<td>0.1780</td>
<td>0.1986</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NTCIR-16 DialEval-2 English Dialogue Quality (DQ) Development set</th>
<th>A-score</th>
<th>S-score</th>
<th>E-score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
<td>RSNOD</td>
<td>NMD</td>
<td>RSNOD</td>
</tr>
<tr>
<td>IMNTPU-run0</td>
<td>0.2102</td>
<td>0.1397</td>
<td>0.1879</td>
</tr>
</tbody>
</table>

NTCIR-16 Conference, June 14-17, 2022, Tokyo, Japan
## Results of IMNTPU at NTCIR-16 DialEval-2 Chinese Nugget Detection

<table>
<thead>
<tr>
<th>Model</th>
<th>NTCIR-16 DialEval-2 Chinese Nugget Detection (ND) Test set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>JSD</td>
</tr>
<tr>
<td>Baseline-run0</td>
<td>0.0585</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>NTCIR-16 DialEval-2 Chinese Nugget Detection (ND) Development set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>JSD</td>
</tr>
<tr>
<td>IMNTPU-run0</td>
<td>2.0670</td>
</tr>
</tbody>
</table>

NTCIR-16 Conference, June 14-17, 2022, Tokyo, Japan
## Results of IMNTPU at NTCIR-16 DialEval-2 English Nugget Detection

<table>
<thead>
<tr>
<th>Model</th>
<th>JSD</th>
<th>RNSS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IMNTPU-run0</strong></td>
<td>0.0601</td>
<td>0.1574</td>
</tr>
<tr>
<td><strong>Baseline-run0</strong></td>
<td>0.0625</td>
<td>0.1722</td>
</tr>
</tbody>
</table>

**NTCIR-16 DialEval-2 English Nugget Detection (ND) Test set**

<table>
<thead>
<tr>
<th>Model</th>
<th>JSD</th>
<th>RNSS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IMNTPU-run0</strong></td>
<td>0.0752</td>
<td>0.1727</td>
</tr>
</tbody>
</table>

**NTCIR-16 DialEval-2 English Nugget Detection (ND) Development set**
Conclusion

- We employed XLM-RoBERTa to extract texts between conversations to learn key information and to cope with text categorization challenges in the DQ and ND subtasks.
Contribution

• The important contribution of this study is that we proposed two critical elements, namely, Tokenization procedures and Fine-Tuning Approaches, to improve the DQ and ND subtasks in dialogue analysis.
IMNTPU Dialogue System Evaluation at the NTCIR-16 DialEval-2
Dialogue Quality and Nugget Detection

1Information Management, National Taipei University, New Taipei City, Taiwan
2Zeals Co., Ltd. Tokyo, Japan

Ting-Yun Hsiao1  Yung-Wei Teng1  Pei-Tz Chiu1  Mike Tian-Jian Jiang2  Min-Yuh Day1,*

myday@gm.ntpu.edu.tw