With the development of natural language understanding and generation technologies, more and more commercial companies have been setting up intelligent dialogue systems, such as helpdesk robots. These dialogue systems could provide wait-free and homogeneous services for their customers, but at the same time suffer from the problems of misunderstanding and generating nonsense or offensive utterances to the system users.

The TUA1 team participates in the Dialogue Quality and Nugget Detection subtasks of NTCIR-16 DialEval-2 task. This work is a continuation of our previous works in text conversation and text emotion analysis.

**Methodology**

The dialogue quality prediction network mainly consisted of a feature extractor and a feedforward network. The two parts are shown on the left, separated by a dotted line in between.

Firstly, in order to better represent the structure of a set of dialogue, we preprocess the input dialogue sequence.

Secondly, we employ pre-trained Transformer networks to extract the hidden representations of the structured dialogue input. Topic information is also extracted as a part of the hidden representation of the input dialogue.

Thirdly, the feedforward network is located above the dotted line, which is arranged as follows: a full connection layer, an activation function, a dropout layer, a linear dimension reduction layer, and a softmax function.

Finally, the network get the probabilities over the quality label set for every quality type. We employ the mean squared error (MSE) loss for evaluating the training loss.

In nugget detection subtask, we used the same model as the dialogue quality task. The difference is, in order to fit the nugget labels, we divide all the dialogue utterances into two parts, named the customer part and the helpdesk part, that is, utterances extracted from either customer or helpdesk are trained separately.

**Experiments**

<table>
<thead>
<tr>
<th>Topic Numbers</th>
<th>Score_sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>15.37</td>
</tr>
<tr>
<td>10</td>
<td>16.27</td>
</tr>
<tr>
<td>20</td>
<td>16.11</td>
</tr>
<tr>
<td>50</td>
<td>15.79</td>
</tr>
</tbody>
</table>

The TUA1 team submits three runs for the Chinese DQ subtask, one run for the English DQ subtask, two runs for the Chinese ND subtask, and one run for the English ND subtask.

Our proposed method reaches the best scores for RSNOD and NMD metrics in both Chinese and English dialogue quality subtasks among all participants. The results indicate that the proposed method is promising in learning a dialogue quality prediction system for generating very close predictions to the human annotators.

Our contributions:

- A novel and effective architecture for Dialogue Quality and Nugget Detection tasks.
- The best scores for RSNOD and NMD metrics in both Chinese and English Dialogue Quality subtasks.

Future works:

- Extend stop words library, such as “123456789”.
- Further study the effect of different Transformer levels.

Acknowledgements:

This work is supported by JSPS KAKENHI Grant Number 19K20345 and 19H04215.