

TUA1 at the NTCIR-15 DialEval Task

Xin Kang*, Yunong Wu**, Fuji Ren*
 *Tokushima University
 **CDXT medical technology

Introduction

- DialEval-1 of NTCIR-15
 - Chinese dialogue quality (DQ) subtask
 - Chinese nugget detection (ND) subtask
- To evaluate customer-helpdesk dialogue automatically
- Challenges:
 - Speaker identities should be learned to promote the dialogue understanding.
 - How to incorporate these speaker identities into the dialogue evaluation?
 - Loss function: regression or others...

Dialogue Quality Prediction Network

Structure

- Pre-trained BERT Network
 - takes in a sequence of tokens and generates a sequence of feature vectors
- Bi-LSTM Network
 - adds the speaker identity embeddings to the BERT output and integrates them bi-directionally
- Self-Attention Network
 - summarize the input feature vectors into several compact feature vectors by different attentions
- Feed-forward Network
 - concat the attentional outputs and generate predictions over the dialogue qualities

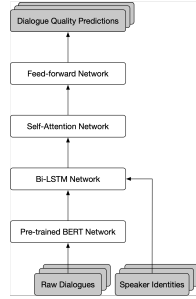


Figure 1: The structure of dialogue quality prediction (DQP) network.

Dialogue Quality Prediction Network

Identifying Speakers in Dialogue

- The dialogues are split by turns.
- Each turn contains one or more utterances.
- We combine multiple utterances into one long utterance.
- Speakers are encoded into 0's and 1's and duplicated into vectors.

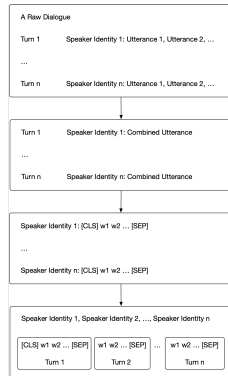


Figure 2: The process of model input for a raw dialogue.

Dialogue Quality Prediction Network

Incorporating Speaker Identities into Network

- Speaker identities are embedded and added to the BERT outputs
- Feed them into Bi-LSTM to integrate the sequential information.
- Self-attention

$$\text{attn_w}^{(l)} = \text{softmax}\left(\frac{W_0^{(l)} + I_{\text{Bi-LSTM}}}{\sqrt{L}}\right)$$

$$\text{attn_v}^{(l)} = \text{attn_w}^{(l)} * I_{\text{Bi-LSTM}}$$

$$O_{\text{Self-Atten}} = \text{Concat}(\text{attn_v}^{(0)}, \dots, \text{attn_v}^{(N)})$$

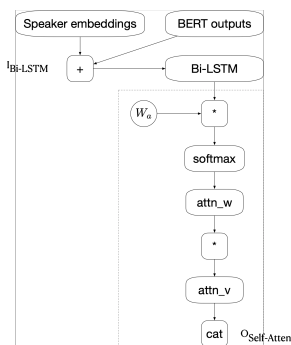


Figure 3: The process of self-attention.

Dialogue Quality Prediction Network

Incorporating Speaker Identities into Network

- Speaker identities are embedded and added to the BERT outputs
- Feed them into Bi-LSTM to integrate the sequential information.
- Self-attention

$$\text{attn_w}^{(l)} = \text{softmax}\left(\frac{W_0^{(l)} + I_{\text{Bi-LSTM}}}{\sqrt{L}}\right)$$

$$\text{attn_v}^{(l)} = \text{attn_w}^{(l)} * I_{\text{Bi-LSTM}}$$

$$O_{\text{Self-Atten}} = \text{Concat}(\text{attn_v}^{(0)}, \dots, \text{attn_v}^{(N)})$$

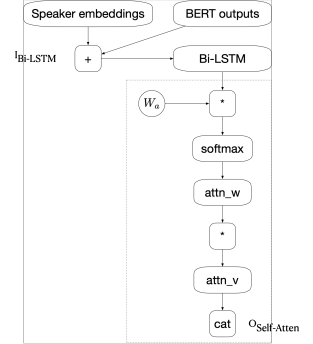


Figure 3: The process of self-attention.

Loss Functions

- Dialogue quality (DQ) subtask

- Mean squared error loss function (\bar{y} for the ground truth distributions and \hat{y} for the model predicted distributions, $\kappa \in \{A, S, E\}$):

$$\text{loss}(\bar{y}, \hat{y}) = \sum_{\kappa \in \{A, S, E\}} \frac{1}{n} \sum_{i=1}^n (\bar{y}_i^{\kappa} - \hat{y}_i^{\kappa})^2$$

- Sinkhorn divergence loss function (y for the set of ground truth labels and \hat{y} for the set of generated samples, $\kappa \in \{A, S, E\}$):

$$\text{loss}(y, \hat{y}) = \sum_{\kappa \in \{A, S, E\}} \frac{1}{n} \sum_{i=1}^n \text{Div}(y_i^{\kappa}, \hat{y}_i^{\kappa})$$

- Nugget detection (ND) subtask

- Mean squared error loss function (y for the ground truth distributions and \hat{y} for the model predicted distributions, $\kappa \in \{\text{CNUGo}, \text{HNUG}, \text{CNUG}, \text{HNUG}^*, \text{CNUG}^*, \text{CNAN}, \text{HNAN}\}$)

$$\text{loss}(y, \hat{y}) = \sum_{\kappa \in \{A, S, E\}} \frac{1}{n} \sum_{i=1}^n (y_i^{\kappa} - \hat{y}_i^{\kappa})^2$$

Experiment

Submitted RUNs

- For dialogue quality (DQ) subtask
 - RUN0: DQP network with the mean squared error loss
 - RUN1: DQP network with the Sinkhorn divergence loss for single-label probabilities
 - RUN2: DQP network with the Sinkhorn divergence loss for multi-label probabilities
- For nugget detection (ND) subtask
 - RUN0: ND network with the mean squared error loss

Results

RUN	Mean RSNOD	Mean NMD
0	0.2136	0.1996 ⁽²⁾
1	0.2484	0.1510
2	0.2102 ⁽¹⁾	0.1412

Table 1: The A-score Results for Chinese Dialogue Quality Prediction.

RUN	Mean RSNOD	Mean NMD
0	0.1615 ⁽¹⁾	0.1144 ⁽¹⁾
1	0.1810	0.1253
2	0.1617	0.1187

Table 3: The E-score Results for Chinese Dialogue Quality Prediction.

RUN	Mean RSNOD	Mean NMD
0	0.2053	0.1322
1	0.2302	0.1397
2	0.2024	0.1310

Table 2: The S-score Results for Chinese Dialogue Quality Prediction.

Run	Mean JSD	Mean RNSS
0	0.0859	0.1892

Table 4: The Results for Chinese Dialogue Nugget Detection.

Conclusions

- Two neural network models for the Chinese dialogue quality (DQ) and nugget detection (ND) subtasks.
- Main contributions:
 - Speaker identities are employed to promote the dialogue understanding.
 - Speaker identities are embedded and incorporated into the network.
 - We propose a new loss function for dialogue quality prediction.
- Future work:
 - Other loss functions to improve the ordinal regression results.