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 dialogue understanding.
- How to incorporate these speaker identities into dialogue evaluation?
- Loss function: regression or others...

Dialogue Quality Prediction Network

Structure
- Pre-trained BERT Network
  - Takes a sequence of tokens and generates a sequence of feature vectors
- Bi-LSTM Network
  - Adjusts the speaker identity embeddings in the BERT output and aggregates them bi-directionally
- Self-Attention Network
  - Summarizes the input feature vectors into several compact feature vectors by different attentions
- Feed-forward Network
  - Consists of the attentional outputs and generates predictions over dialogue qualities

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
  \[ \text{Self-attention}(x) = \text{softmax} \left( \frac{W^T \cdot x + b}{\sqrt{d}} \right) \]

Loss Functions

Dialogue quality (DQ) subtask
- Mean squared error loss function \( \ell \) for the ground truth distribution and \( \hat{\ell} \) for the model predicted distribution, \( \ell \in \{ A, S, E \} \):
  \[ \text{loss}(\theta) = \frac{1}{N} \sum_{i=1}^{N} (\ell(x_i) - \hat{\ell}(x_i))^2 \]

Sinkhorn divergence loss function \( \ell \) for the set of ground truth labels and \( \hat{\ell} \) for the set of generated samples, \( \ell \in \{ A, S, E \} \):
  \[ \text{loss}(\theta) = \frac{1}{N} \sum_{i=1}^{N} | \ell(x_i) - \hat{\ell}(x_i) | \]

Nugget detection (ND) subtask
- Mean squared error loss function \( \ell \) for the ground truth distribution and \( \hat{\ell} \) for the model predicted distribution, \( \ell \in \{ CNUG, HPNX, CNUS, CNUG, HPNX, CNUS \} \):
  \[ \text{loss}(\theta) = \frac{1}{N} \sum_{i=1}^{N} (\ell(x_i) - \hat{\ell}(x_i))^2 \]

Experiment

Submitted RUNs

- For dialogue quality (DQ) subtask
  - RUN: DQP network with the mean squared error loss
  - RUN: DQP network with the Sinkhorn divergence loss for single-label probabilities
  - RUN: DQP network with the Sinkhorn divergence loss for multi-label probabilities

- For nugget detection (ND) subtask
  - RUN: ND network with the mean squared error loss

Results

Conclusions

- Two neural network models for the Chinese dialogue quality (DQ) and nugget detection (ND) subtasks.
- Main contributions:
  - Speaker identities are employed to promote 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.