TUA1 at the NTCIR-15 DialEval Task

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

Contents

Contents

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...

Contents

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 ingetrates 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

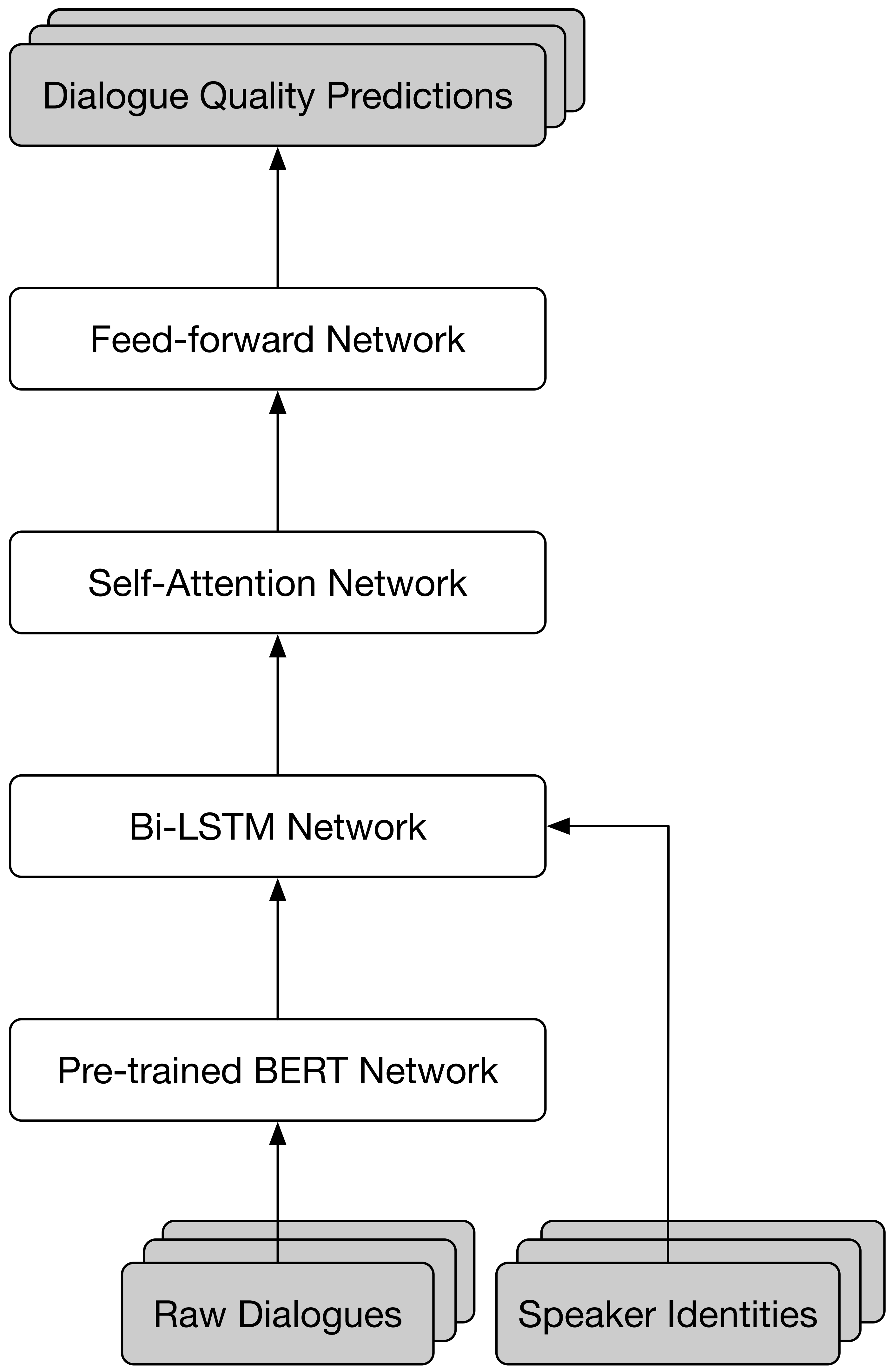

Figure 1: The structure of dialogue quality prediction (DQP) network.

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.

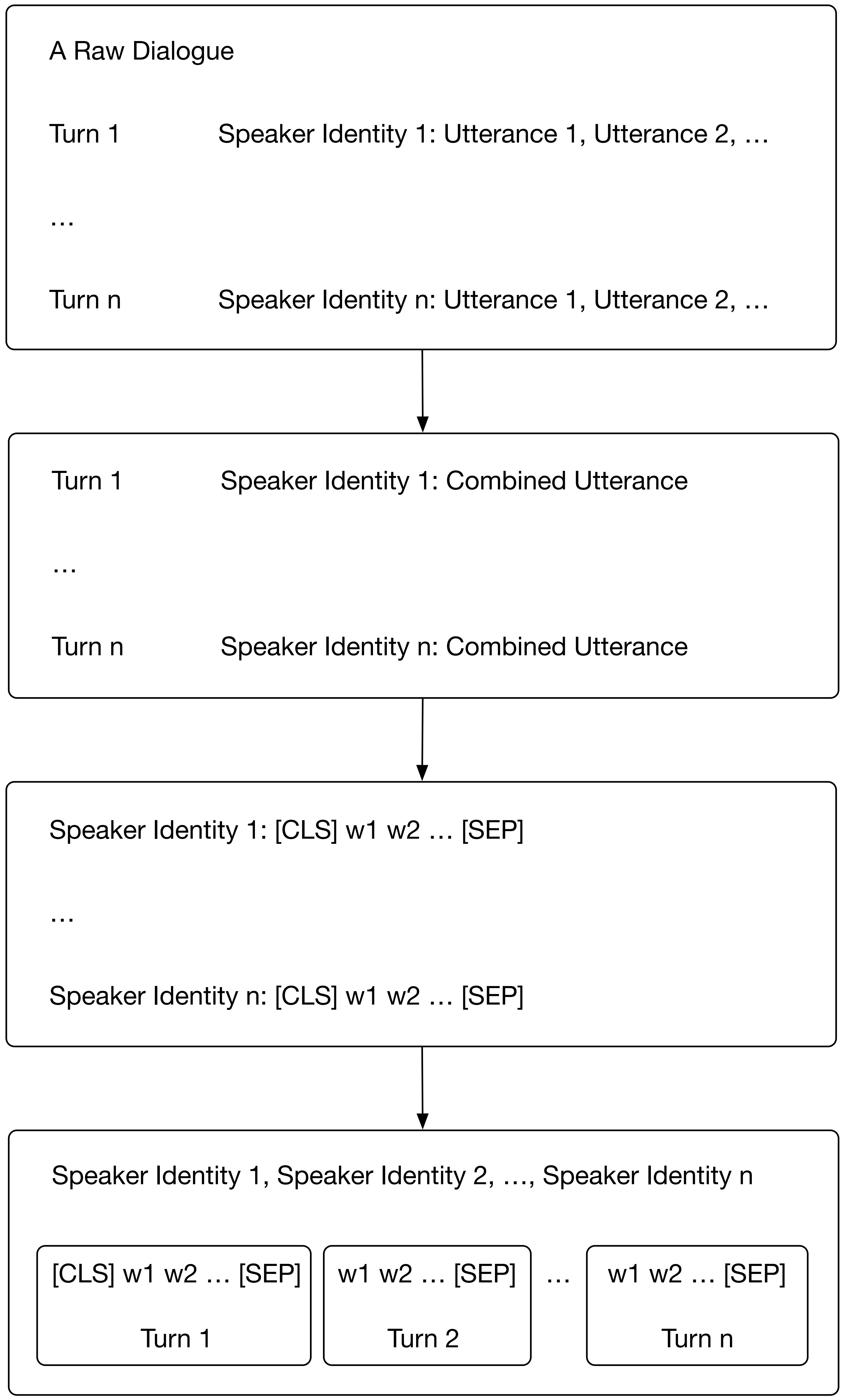

Figure 2: The process of model input for a raw dialogue.

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

$$\text{attn\_w}^{(i)} = \text{softmax}\left(\frac{W_a^{(i)} * \text{I}_{\text{Bi-LSTM}}}{\sqrt{L}}\right)$$

$$\text{attn\_v}^{(i)} = \text{attn\_w}^{(i)} * \text{I}_{\text{Bi-LSTM}}$$

$$\text{O}_{\text{Self-Atten}} = \text{Concat}(\text{attn\_v}^{(0)}, \ldots, \text{attn\_v}^{(h)})$$

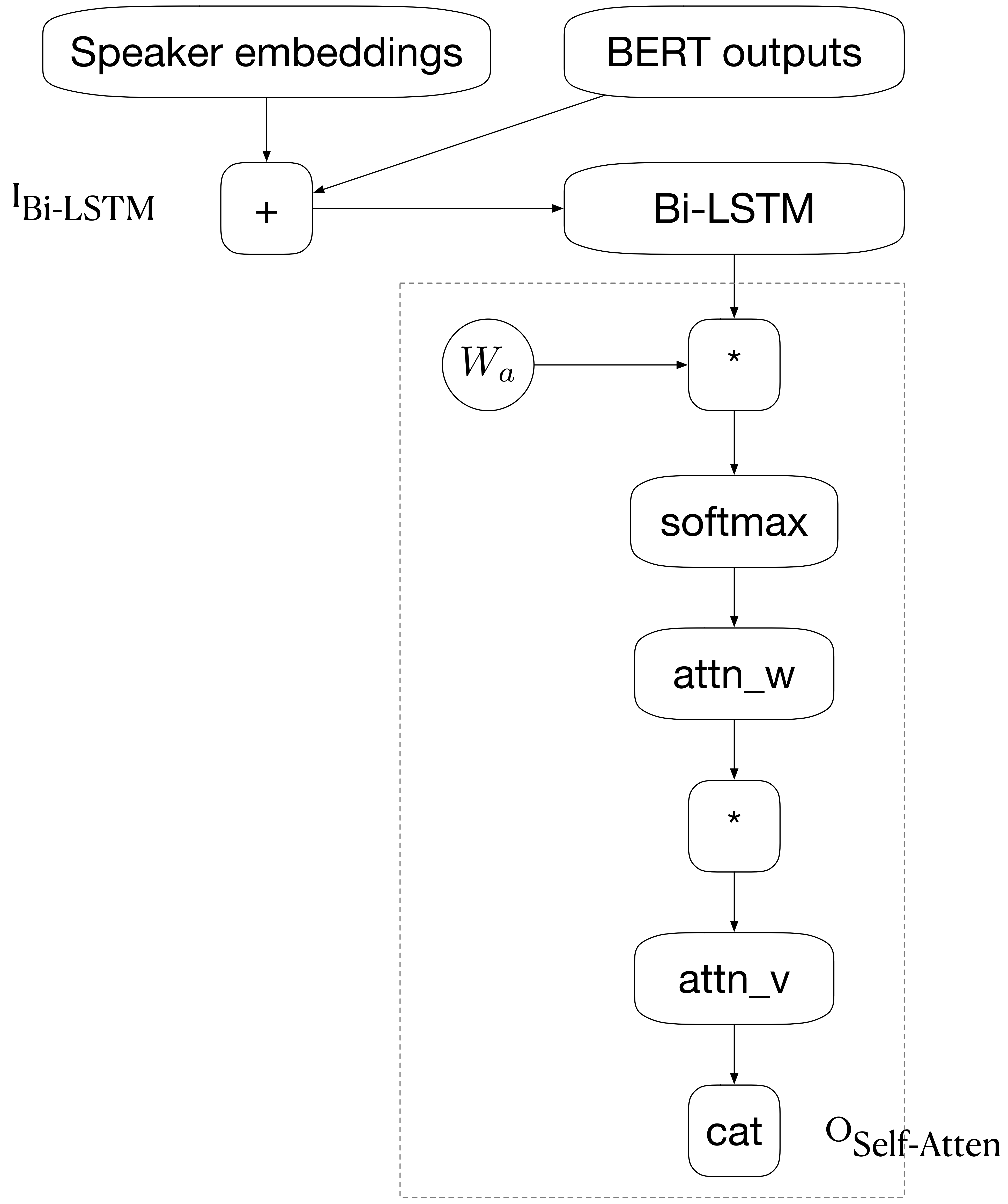

Figure 3: The process of self-attention.

Contents

Nugget Detection Network

Structure

- Pre-trained BERT Network
- Bi-LSTM Network
- Self-Attention Network
- Feed-Forward Network
- Differences to the dialogue quality prediction network:
    - No speaker identifiers in the input
    - Multiple nugget detections in the output

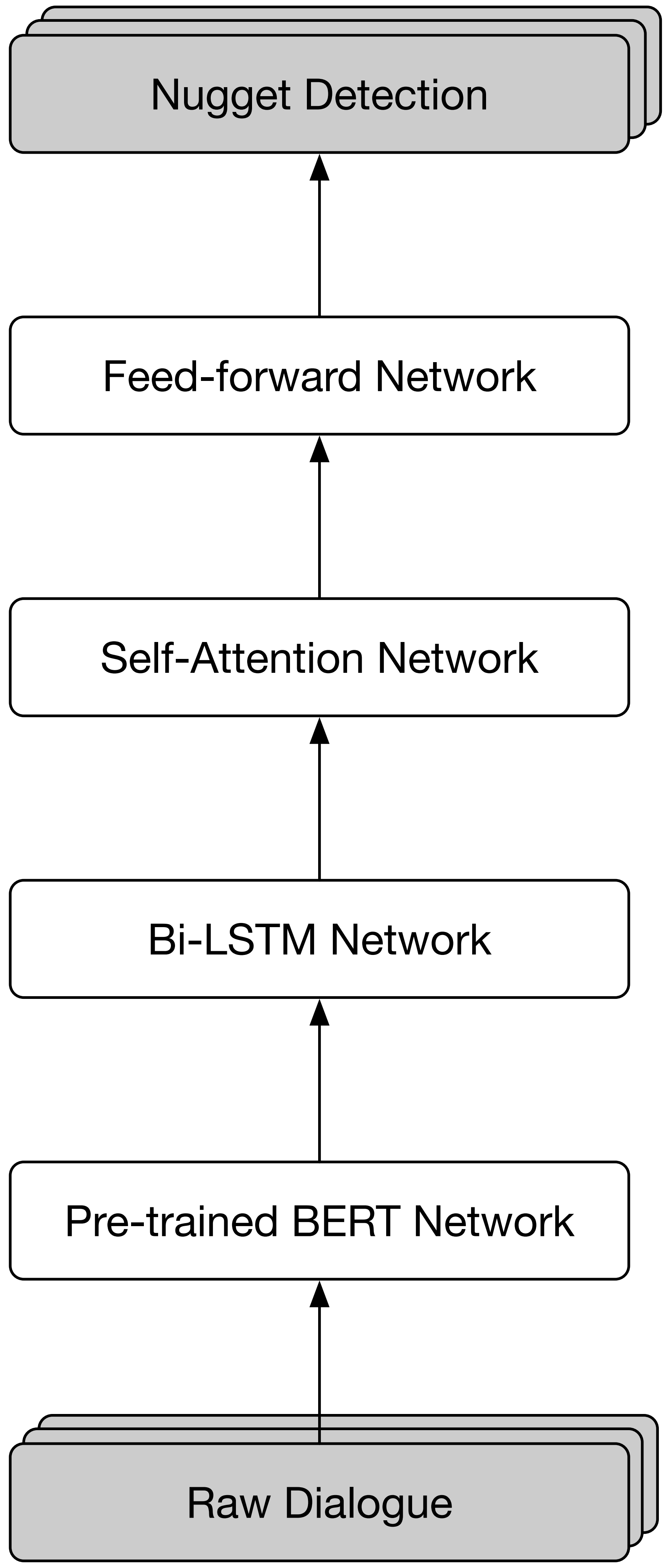

Figure 4: The structure of nugget detection (ND) network.

Contents

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, $\kappa \in \{A, S, E\}$):

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

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

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

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

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

Contents

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

Experiment

Results

RUN	Mean RSNOD	Mean NMD
0	0.2136	**0.1396**$^{(2)}$
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	0.2053	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 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 JSD	Mean RNSS
0	0.0859	0.1892

Table 4: The Results for Chinese Dialogue Nugget Detection.

Contents

- Introduction
- Dialogue Quality Prediction Network
- Nugget Detection Network
- Loss Functions
- Experiment
- <u>Conclusions</u>

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.

Thank you for your attention!