RSLDE at NTCIR-16 DialEval-2 Task

Fan Li, Tetsuya Sakai

Waseda University

The 16th NTCIR (2021 - 2022)
Evaluation of Information Access Technologies
16th, June

Overview

Introduction

- RSLDE at DialEval-2
  - English and Chinese Dialogue Quality (DQ) subtask
  - English and Chinese Nugget Detection (ND) subtask
- Key Challenges
  - DQ task:
    - Representation of the structure of dialogue
  - ND task:
    - Representation of Dialogue Structure
    - Does Dialogue Context Matters?

Sentence Level Model

Idea: Nugget has pattern of words

- CNUG0:
  - ”@ Ctrip Customer Service Please read the details in the picture.”
  - ”@ Smartisan Technology Customer Service ... of the wireless ring of mobile phone ?”
- CNUG*:
  - ”Thank you.”
  - ”A customer service staff called to explain the problem this morning. I’m satisfied with this reply. The staff’s attitude was sincere. I think Unicom is quite good.”
- HNUG*:
  - ”We will redouble our efforts to do better service.”
  - ”You’re welcome. That’s our job.”

Sentence Level Model

Language Model

BERT and XLNet (Permutation AR)

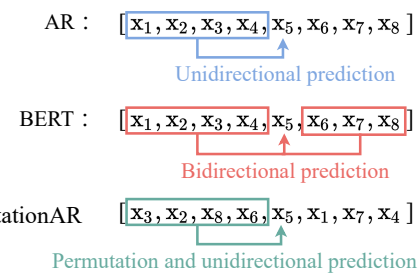


Figure: Comparison between BERT and XLNet

Approache

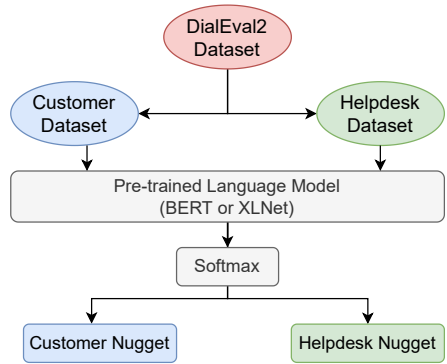
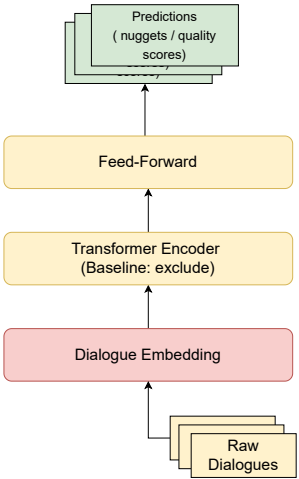
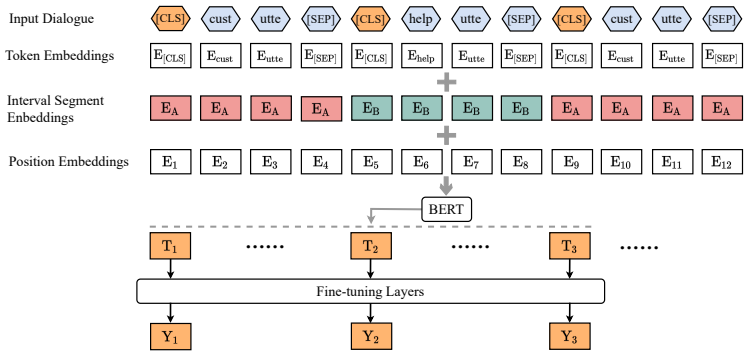


Figure: RSLDE sentence model

Dialogue Level Model





Modified BERT's Interval Segment Embedding to represent the dialogue structure.

Figure: Dialogue Embedding

Experiments

- Loss Function (for all tasks): $\text{Loss} = -\sum_{i=1}^{\text{size}} y_i \cdot \log \widehat{y}_i$
- Submitted Runs:

Task	Language	Run	Model	Batch size
DQ	English	0	Dialogue-BERT+Transformer	12
		1	Dialogue-BERT	
	Chinese	0	Dialogue-BERT+Transformer	
		1	Dialogue-BERT	
ND	English	0	XLNet Baseline	8
		1	BERT Baseline	
		2	Dialogue-BERT+Transformer	
	Chinese	0	XLNet Baseline	
		1	BERT Baseline	
		2	Dialogue-BERT+Transformer	

ND Results

Table: Chinese Nugget Detection Results

Run	Mean JSD	Run	Mean RNSS
RSLDE-run0	0.0560[(1)]	RSLDE-run0	0.1604[(1)]
BL-LSTM	0.0585	BL-LSTM	0.1651
RSLDE-run2	0.0607	RSLDE-run1	0.1712
RSLDE-run1	0.0634	RSLDE-run2	0.1720
BL-popularity	0.1864	BL-popularity	0.2901
BL-uniform	0.2042	BL-uniform	0.3371

Table: English Nugget Detection Results

Run	Mean JSD	Run	Mean RNSS
RSLDE-run0	0.0557[(1)]	RSLDE-run0	0.1615[(2)]
BL-LSTM	0.0625	BL-LSTM	0.1722
RSLDE-run2	0.0676	RSLDE-run2	0.1778
RSLDE-run1	0.0691	RSLDE-run1	0.1853
BL-popularity	0.1864	BL-popularity	0.2901
BL-uniform	0.2042	BL-uniform	0.3371

DQ Results

Table: Chinese Dialogue Quality Results

Run	Mean A-RSNOD	Run	Mean A-NMD
BL-LSTM	0.2301	RSLDE-run0	0.1537
BL-popularity	0.2320	RSLDE-run1	0.1551
RSLDE-run0	0.2438	BL-popularity	0.1577
RSLDE-run1	0.2446	BL-LSTM	0.1772
BL-uniform	0.2767	BL-uniform	0.2500

Run	Mean S-RSNOD	Run	Mean S-NMD
RSLDE-run0	0.1938	RSLDE-run1	0.1229
RSLDE-run1	0.1964	RSLDE-run0	0.1243
BL-LSTM	0.1998	BL-popularity	0.1288
BL-popularity	0.2062	BL-LSTM	0.1523
BL-uniform	0.2959	BL-uniform	0.2565

Run	Mean E-RSNOD	Run	Mean E-NMD
RSLDE-run0	0.1660	RSLDE-run0	$0.1222^{(2)}$
RSLDE-run1	0.1725	RSLDE-run1	0.1286
BL-LSTM	0.1854	BL-LSTM	0.1579
BL-uniform	0.2496	BL-popularity	0.1710
BL-popularity	0.2569	BL-uniform	0.2106

Table: English Dialogue Quality Results

Run	Mean A-RSNOD	Run	Mean A-NMD
BL-popularity	0.2320	BL-popularity	0.1577
BL-LSTM	0.2321	BL-LSTM	0.1780
RSLDE-run0	0.2615	RSLDE-run1	0.1896
RSLDE-run1	0.2725	RSLDE-run0	0.1957
BL-uniform	0.2767	BL-uniform	0.2500

Run	Mean S-RSNOD	Run	Mean S-NMD
BL-LSTM	0.1986	BL-popularity	0.1288
BL-popularity	0.2062	RSLDE-run0	0.1381
RSLDE-run0	0.2078	RSLDE-run1	0.1438
RSLDE-run1	0.2154	BL-LSTM	0.1467
BL-uniform	0.2959	BL-uniform	0.2565

Run	Mean E-RSNOD	Run	Mean E-NMD
BL-LSTM	0.1745	RSLDE-run0	0.1429
RSLDE-run0	0.1832	BL-LSTM	0.1431
RSLDE-run1	0.1889	RSLDE-run1	0.1444
BL-uniform	0.2496	BL-popularity	0.1710
BL-popularity	0.2569	BL-uniform	0.2106

Conclusion

- The XLNet model has an outstanding language understanding capability for customer-helpdesk dialogues.
- Considering the structure and context information of a dialogue is important for the dialogue nugget detection.

The End

Questions? Comments?