# TMUDS at the NTCIR-15 DialEval-1 Task

Yen-Chun Huang, Yi-Hsuan Huang, Yu-Ya Cheng,

Jung-Yi Liao, Yung-Chun Chang

# Author

Yen-Chun Huang
Graduate Institute of Data Science,
Taipei Medical University, Taiwan

Yi-Hsuan Huang
Graduate Institute of Data Science,
Taipei Medical University, Taiwan

Yu-Ya Cheng
Professional Master Program in Data Science,
Taipei Medical University, Taiwan

Jung-Yi Liao
Professional Master Program in Data Science,
Taipei Medical University, Taiwan

Yung-Chun Chang
Graduate Institute of Data Science,
Taipei Medical University, Taiwan

# Outline

1. Introduction
2. Dataset
3. Preprocessing and Feature Extraction
4. Method architecture
5. Experiment and result
6. Experiment and discussion
7. Conclusion and feature works

# Introduction

## DialEval-1

- It's a task that aim to automatically evaluate the quality of the dialogue and which dialogue turns are helpful without a human annotator.
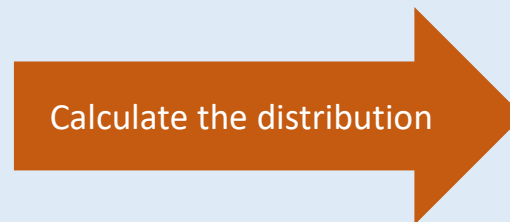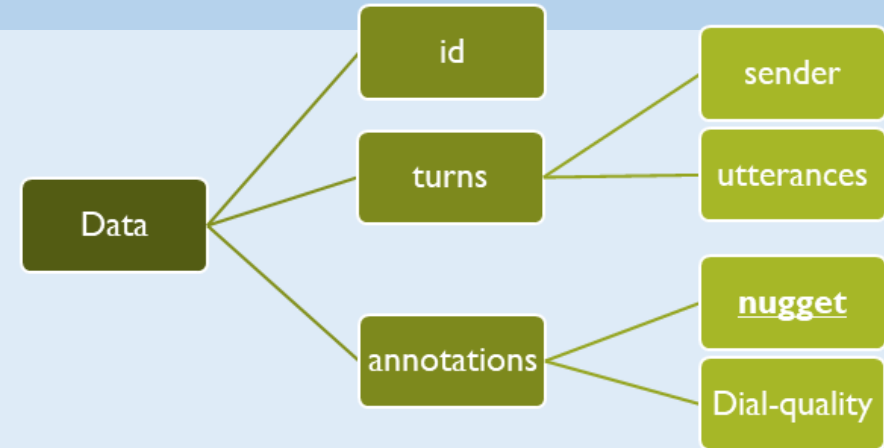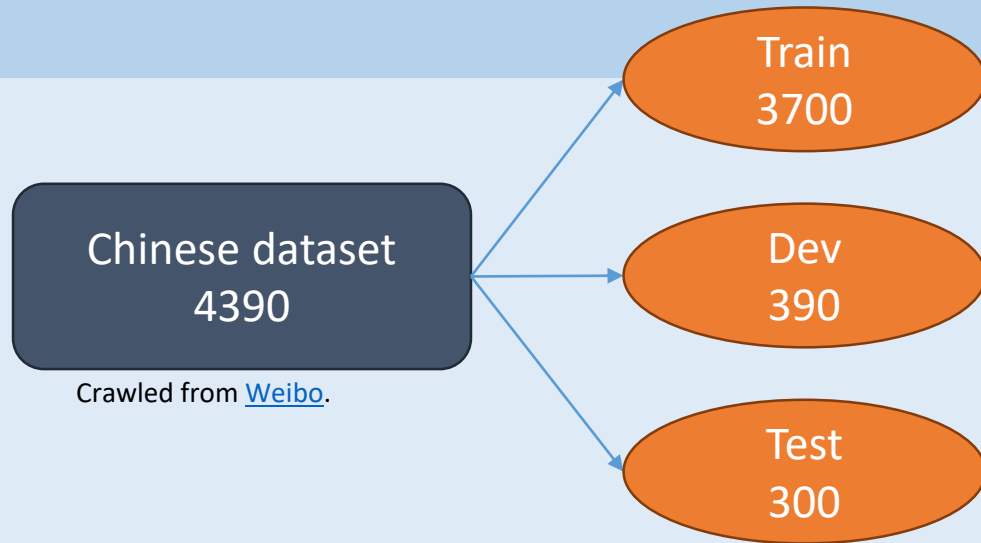
## Nugget Detection

- Given a customer-helpdesk dialogue, return an estimated distribution of labels over nugget types (similar to dialogue acts) for each turn.
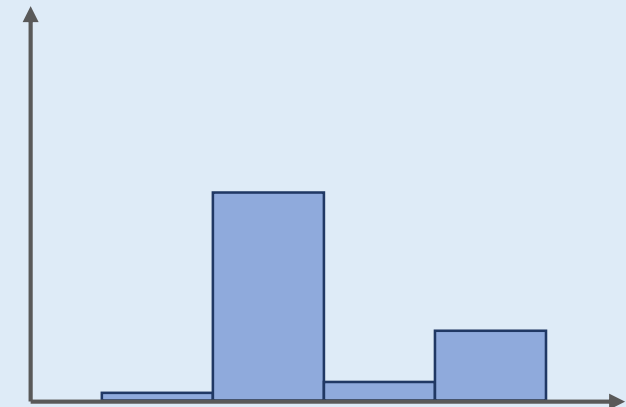
## Dialogue Quality

- Given a customer-helpdesk dialogue, return an estimated distribution of dialogue quality ratings for the entire dialogue.

# Dataset



Chinese dataset
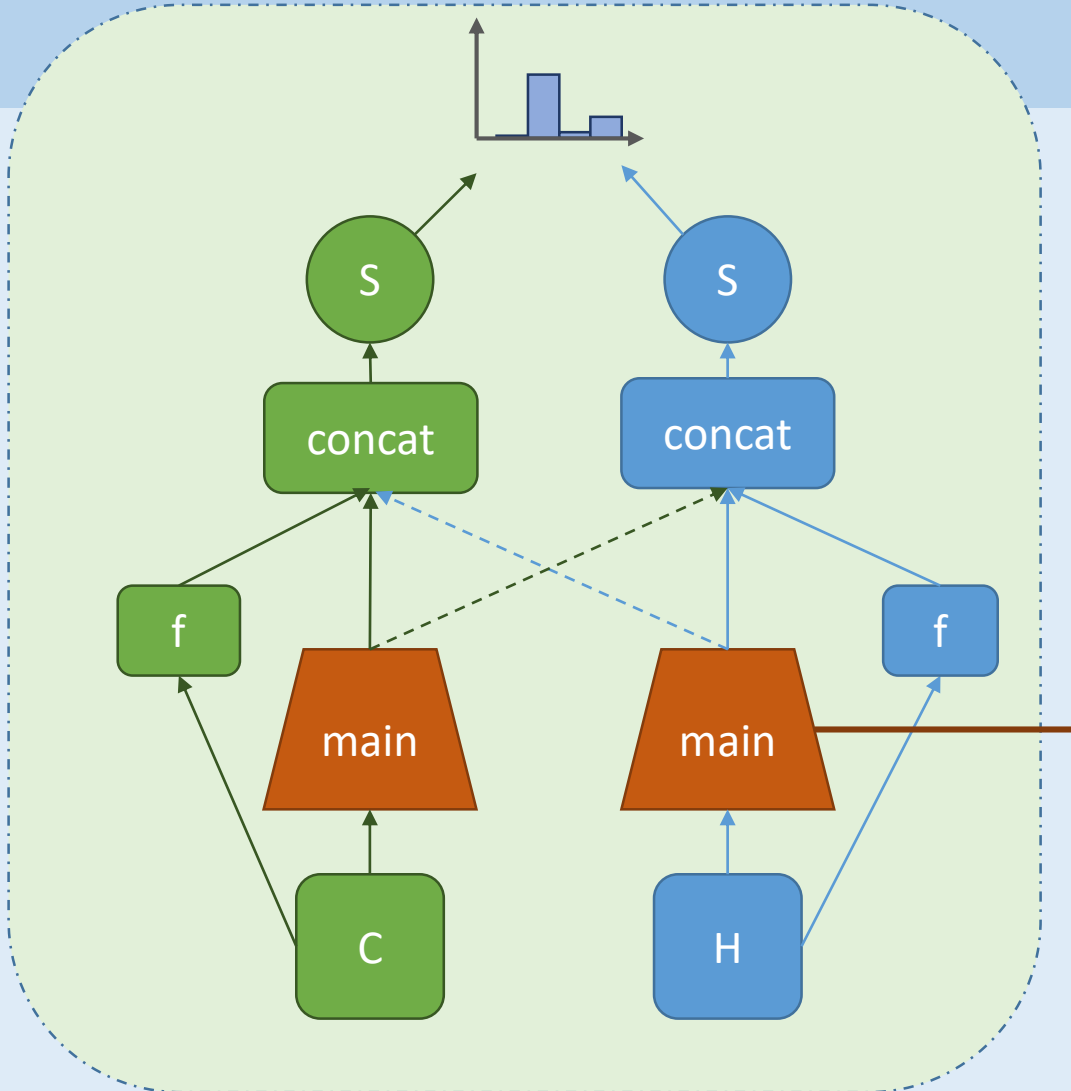4390

Crawled from Weibo.

Train
3700

Dev
390

Test
300

Data
- id
- turns
  - sender
  - utterances
- annotations
  - **nugget**
  - Dial-quality

**19 Nuggets (labels)**

Calculate the distribution

For each utterance

# Preprocessing and Feature Extraction

1. Word segment

2. Round features extraction :

# Method architecture
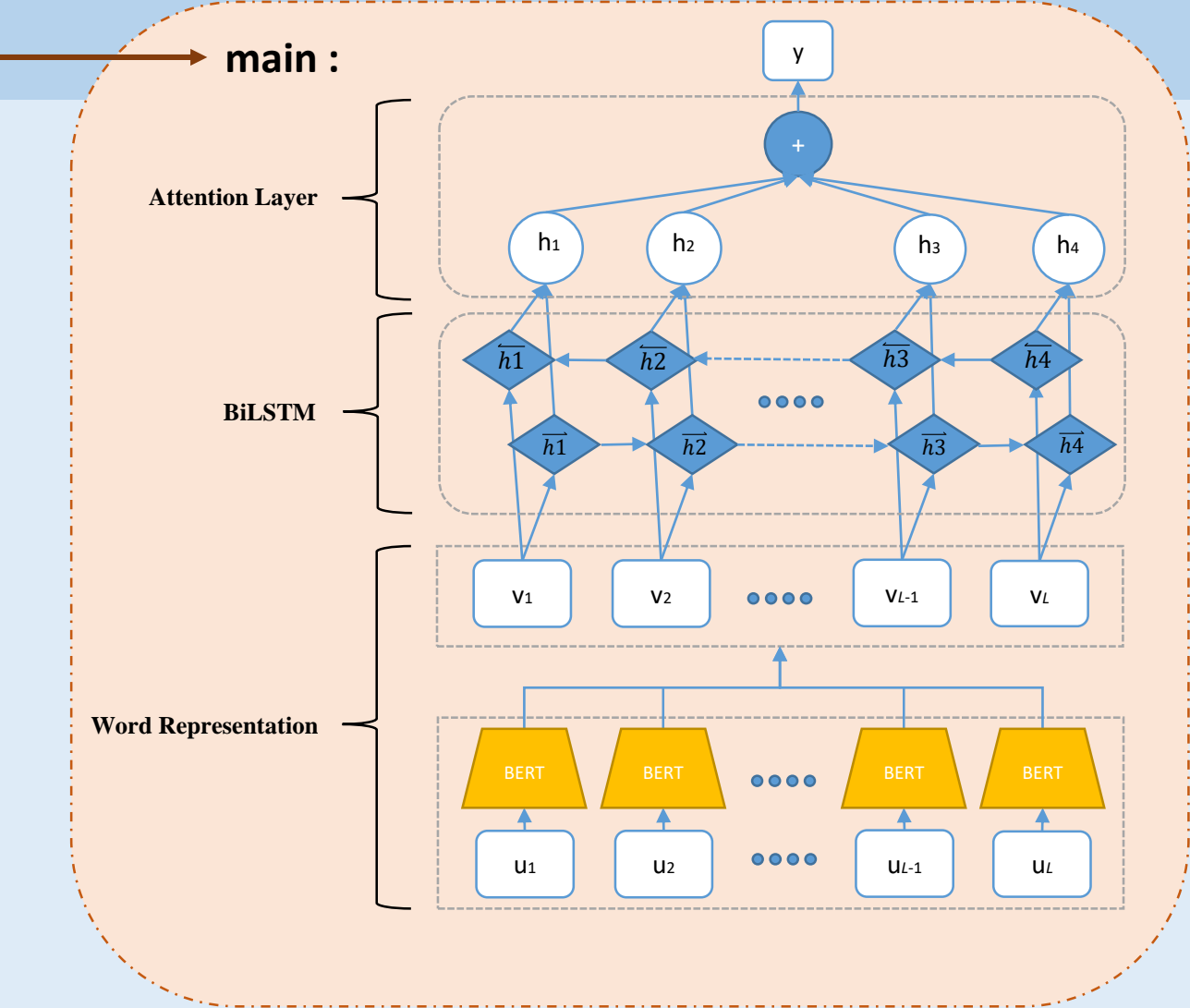
# Experiment and discussion 1

| Hyper-parameter | Value |
|---|---|
| Embedding | Word2Vec (300) Bert (100,768) |
| Learning rate | 1e-3 |
| Drop out | 0.35 |
| Epoch | 50 |
| Batch size | 128 |

| Run | main structure | |
|---|---|---|
| run-0 | | 2 BiLSTM |
| run-1 | Bert emb | 2 BiLSTM + Att |
| run-2 | | 1 BiLSTM + Att |



Results
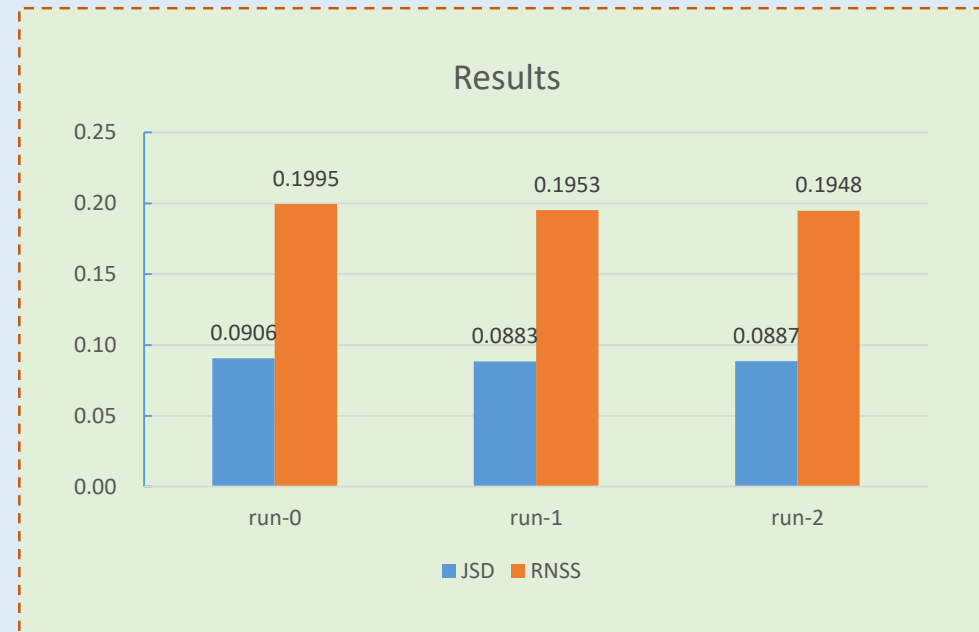
# Experiment and discussion 2

| Method | Local JSD | Local RNSS |
|---|---|---|
| TextCNN | 0.048 | 0.141 |
| LSTM | 0.043 | 0.130 |
| BiLSTM | 0.040 | 0.129 |
| 2 BiLSTM | 0.036 | 0.122 |
| 2 BiLSTM (*) | 0.036 | 0.120 |
| 2 BiLSTM (*/ **) | 0.037 | 0.121 |
| 2 BiLSTM (*/ **) + Att | 0.035 | 0.120 |

- Compared with other basic models, the ability of bidirectional LSTM to process short texts can capture contextual meaning better.

- Based on the statistical results of the training set and the validation set, the nugget label in each round of the dialogue are not uniformly distributed, which has a specific tendency. We separate the first round of dialogue independently from the prediction model.

- We also regard the vector of the previous round as the current training feature and add it to the training.

# Conclusion and feature works

**1**   Extract more features such as dependency and key-word to enhance the effectiveness.

**2**   Adjust the method of position weight from the former utterance.

**3**   Re-modify the model with Transformer BERT.

# Thanks for listening !!