TMUDS at the NTCIR-15 DialEval-1 Task

Yen-Chun Huang#, Yi-Hsuan Huang, Yu-Ya Cheng, Jung-Yi Liao, Yung-Chun Chang*
Graduate Institute of Data Science, Taipei Medical University, Taiwan

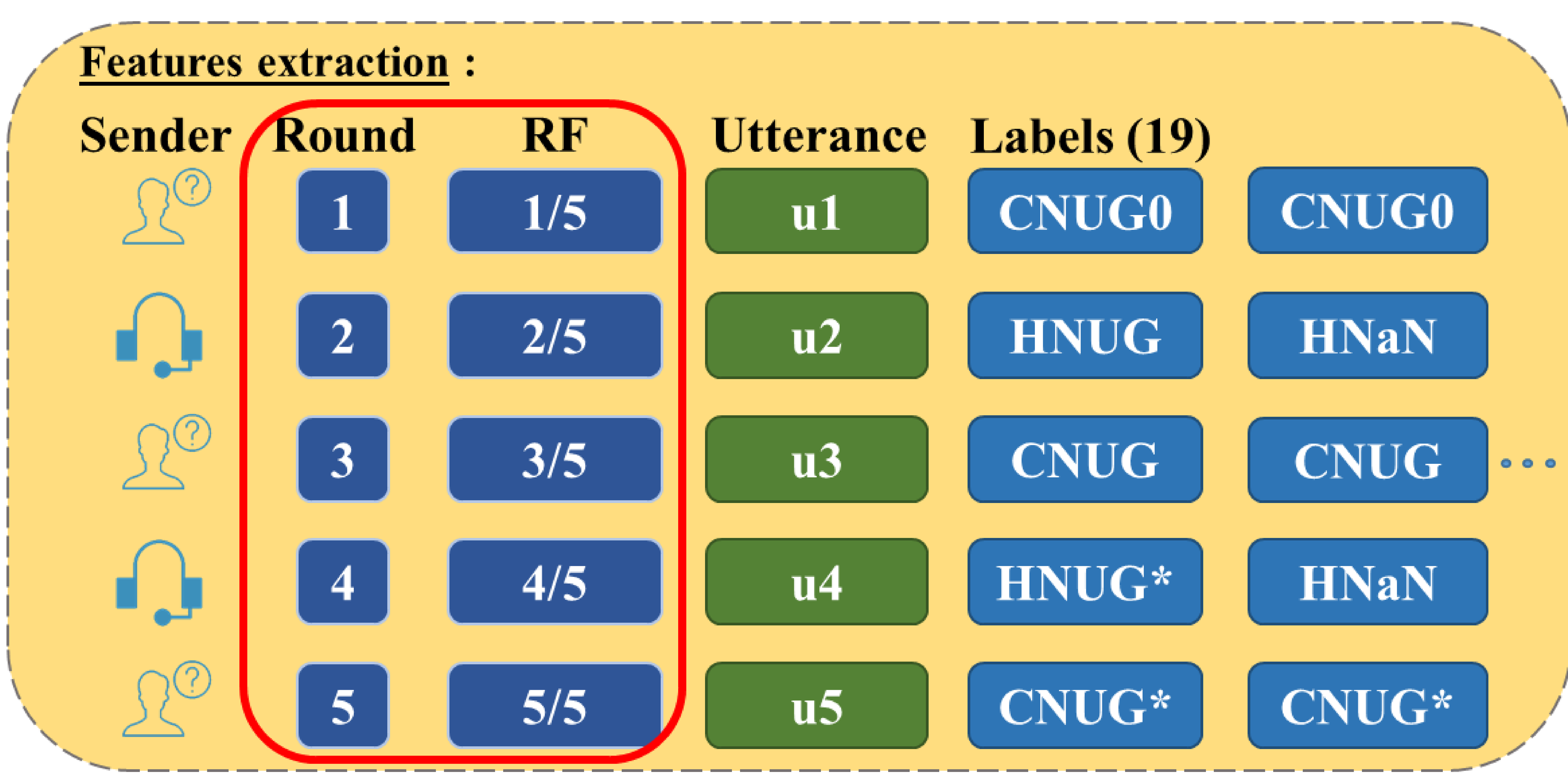
Introduction

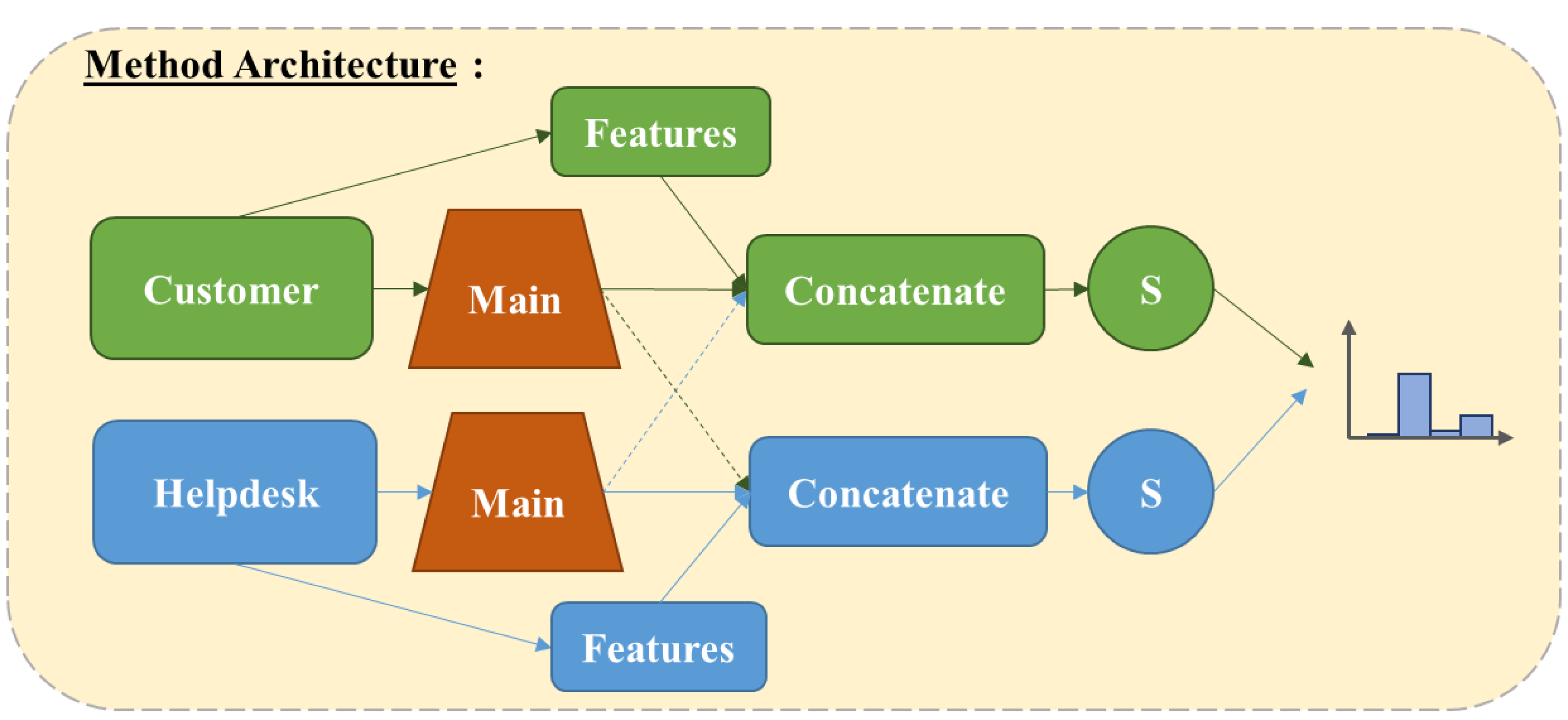
More and more organizations tend to provide online automated customer service to replace traditional human customer service. However, companies must pay additional costs to analyze the increasing amount of customer service dialogue data, so it is imperative to improve the research of automated dialogue system analysis. In this work, we define a dialogue system evaluation model, Bert embedding with Attention for Multi-Turn Dialogue Evaluation (BAMDE): given the multi-turn of dialogue between customers and helpdesks, we can predict the dialogue behavior (donated as nugget) estimated distribution of each sentence and determine whether the task has been successfully achieved.

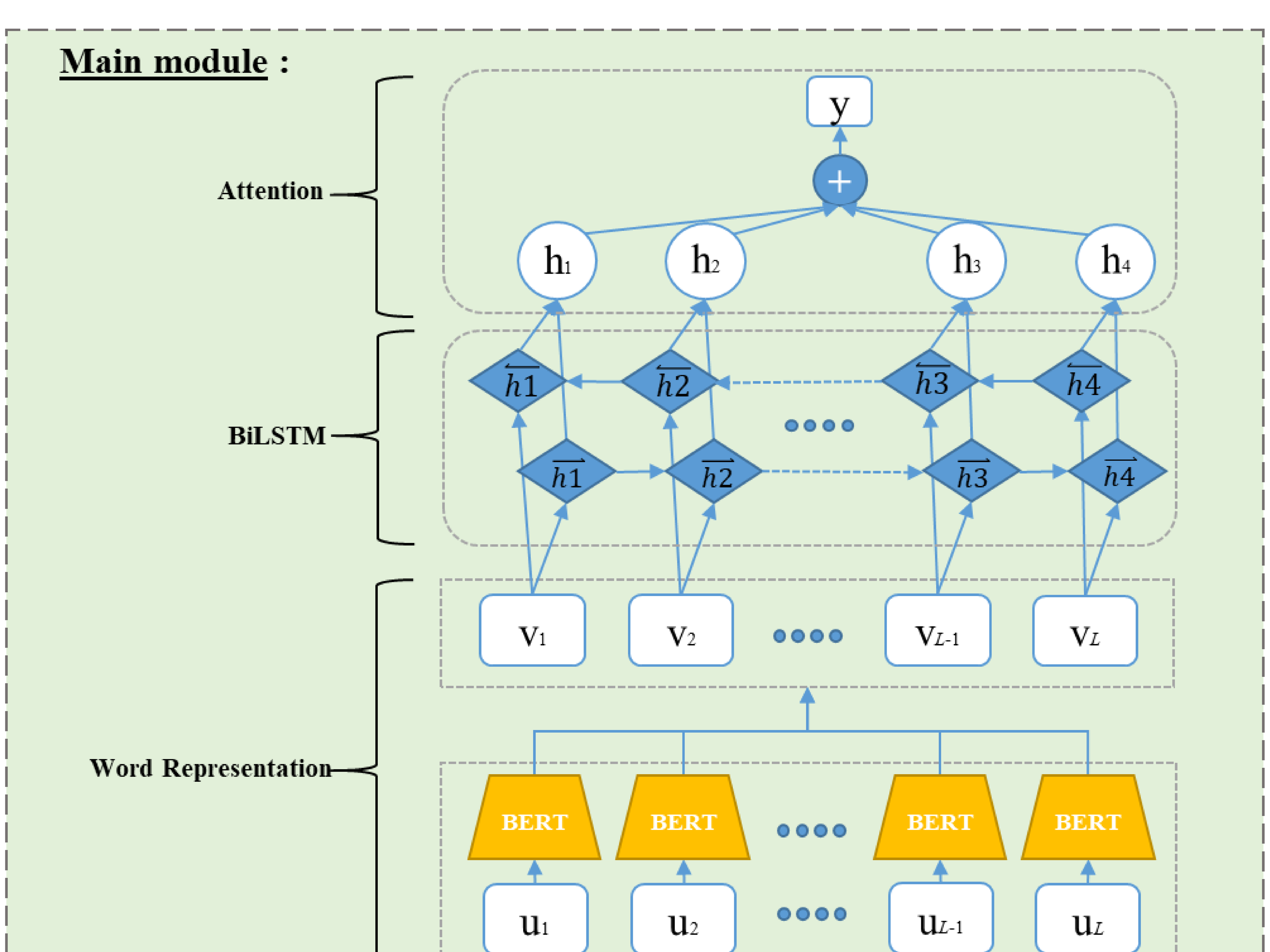
Methodology

First, we introduce Jieba, a Chinese text segmentation tool, to perform word segmentation. And then we extract the features out of the round order.

Secondly, we apply BERT to extract the word vector. Then, we feed them into the BiLSTM model and add the attention Layer. At the end, we concatenate ouputs with features and apply a softmax layer. Below graphs represent the overall architecture of the model :







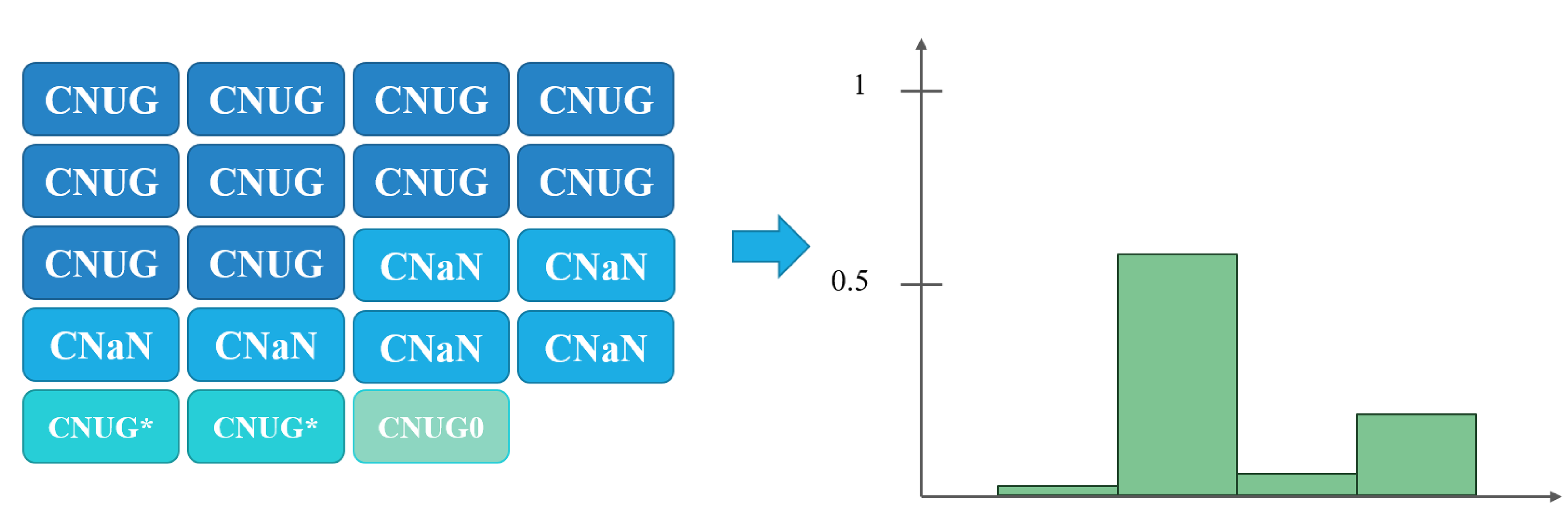
Experiment

The dataset contains Chinese dialogues, it has 3,700 dialogues of training data, 390 dialogues of development data, and 300 dialogues of test data. There are total 7 types of the nugget label in the dataset for customer and helpdesk utterances, and for each utterance it was annotaed by 19 annotators for training and development set (20 annotators for test set). In order to evaluate the multi-label experiment, we applied Jensen Shannon Divergence (JSD) and Root Normalizes Sum of Square (RNSS), in order to compare the difference between estimated distribution and golden standard distribution of each utterance.

The final results are based on three experiments, namely run-0, run-1 and run-2. The bottom table shows the structure of each run and we can see that adding a attention layer is much more effective.

Nugget Labels

CNUG0	Customer	Customer trigger
CNUG		Customer regular nugget
CNUG*		Customer goal
CNaN		Customer not-a-nugget
HNUG	Helpdesk	Helpdesk regular nugget
HNUG*		Helpdesk goal
HNaN		Helpdesk not-a-nugget



Run	JSD	RNSS	Structure
run-0	0.0906	0.1995	Bert emb + 2 BiLSTM
run-1	**0.0883**	0.1953	Bert emb + 2 BiLSTM + Att
run-2	0.0887	**0.1948**	Bert emb + 1 BiLSTM + Att

Conclusion

We introduced a novel model, extracting Bert embeddings to BiLSTM model with attention layer. The experimental results of the concatenation of the above three modules show that the model can make the results better. In future research, we plan to collect the corpus in different languages to test our model. *This research was supported by the Ministry of Science and Technology of Taiwan under grant MOST 107-2410-H-038-017-MY3, MOST 109-2410-H-038 -012 -MY2, and MOST 107-2634-F-001-005.*