

WUST at the NTCIR-14 STC-3 Dialogue Quality and Nugget Detection Subtask

Ming Yan, Maofu Liu, Junyi Xiang

School of Computer Science and Technology
Wuhan University of Science and Technology

liumaofu@wust.edu.cn



Contents

- Introduction
- Related work
- System Architecture
- Experiments
- Conclusions

Introduction

- ◆ Dialogue quality aims to build an evaluation system to evaluate a task-oriented, multi-round, text-based dialogue system.
- ◆ Nugget detection aims to extract the state of each sentence from multiple dialogues.
- ◆ The existing evaluation systems are limited to the specific domains, which are difficult to apply to the other domain universally.
- ◆ We adopt Bi-LSTM and attention mechanism to obtain the semantics of different parts of the conversation to achieve the dialogue quality and nugget detection for task-oriented dialogue.

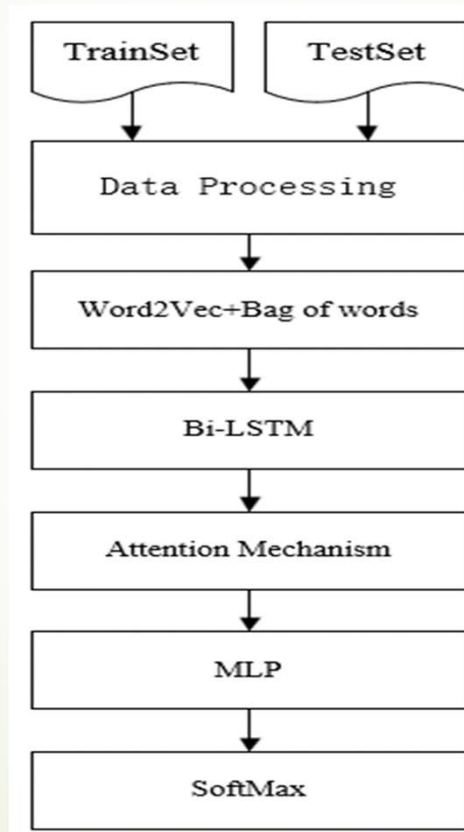


Related work

- ▶ Dialogue quality and nugget detection can help machines better understand natural languages, aiming to help them get important information from the text. Dialogue quality and nugget detection are just beginning to be studied
- ▶ Most of the nugget detection methods regard this problem as a classification task and construct a classifier based on some features of the text. However, the natural language processing tools and resources that extract features may make errors and pass these errors to the final classification. The neural network has a strong feature and semantic learning ability. It can automatically learn text representation from data. Nguyen proposed a convolutional neural network to analyze nugget in the text. This method constrains the context to a fixed window, resulting in the loss of word meaning representation in long sentences

System Architecture

- Our system consist of three main modules, i.e. data preprocessing, feature extraction, and SoftMax classifier.





Data Preprocessing

- ◆ First, we choose Jieba Chinese word segmenter to segment the Chinese word.
- ◆ Second, the stop words will cause noise interference to the effective information of the text, and the deactivation words have no special meaning. In order to reduce the noise interference caused by deactivation words to the sentences, first remove the stop words before extracting the features of the text.

Feature Extraction

- First, we add the word vectors after word segmentation to sentence vectors. Second, we adopt Bi-LSTM to decode sentence vectors, the information of “past” and “future” of the text can be retained at the same time. In this paper, we use three layers of Bi-LSTM to overlay, so that we can fully learn the semantic information and context dependency between dialogues. The attention mechanism is a model that simulates the attention of the human brain. In the field of natural language processing.

$$u_t = \tanh(W_w h_t + b_w) \quad (1)$$

$$\alpha_t = \frac{\exp(u_t^T u_w)}{\sum_t \exp(u_t^T u_w)} \quad (2)$$

$$V = \sum_t \alpha_t h_t \quad (3)$$

- The specific attention mechanism formula is shown in formula (1-3).

Experiments

- ▶ In this paper, we use the data sets provided by NTCIR-14. This dataset contains about 3,700 training instances and 390 instances for testing.
- ▶ Since the classes of DQ subtask are non-nominal, cross-bin measures is more suitable for bin-by-bin measures. As discussed by Sakai, bin-by-bin measures such as Jensen-Shannon Divergence are not adequate for this subtask as they do not consider the distance between classes. Thus, we utilize two cross-bin measures: Normalized Match Distance (NMD) and Root Symmetric Normalized Order-aware Divergence (RSNOD). In contrast to DQ subtask, the classes in ND subtask are nominal, so bin-by-bin measures should be more suitable. Specifically, two measures are used in ND subtask: Root Normalized Sum of Squares (RNSS) and Jensen-Shannon Divergence (JSD).

Experimental results

- We submitted the results of Chinese DQ and ND tasks for NTCIR-14 STC-3. The results given by the organizers are as follows Table 1-4.

Run	Mean RSNO	Run	Mean NMD
WUST-run0	0.1251	WUST-run0	0.0836
WUST-run1	0.1274	WUST-run1	0.0860
WUST-run2	0.1263	WUST-run2	0.0845

Table1. The results of A-score in dialogue quality

Run	Mean RSNO	Run	Mean NMD
WUST-run0	0.1226	WUST-run0	0.0779
WUST-run1	0.1270	WUST-run1	0.0808
WUST-run2	0.1248	WUST-run2	0.0779

Table2. The result of S-score in dialogue quality.

Run	Mean RSNO	Run	Mean NMD
WUST-run0	0.1200	WUST-run0	0.0780
WUST-run1	0.1236	WUST-run1	0.0828
WUST-run2	0.1167	WUST-run2	0.0774

Table3. The result of E-score in dialogue quality.

Run	Mean JSD	Run	Mean RNSS
WUST-run0	0.0223	WUST-run0	0.0909
WUST-run1	0.0233	WUST-run1	0.0931
WUST-run2	0.0250	WUST-run2	0.0980

Table4. The result of nugget detection.

Conclusions

- ▶ In our work, We propose a neural network method, which uses Bi-LSTM to extract the context dependency between dialogues, and adopts attention mechanism to learn the key sentences or phrases in the dialogue, and combines these two kinds of information to improve the dialogue quality and the identification of nugget detection ability.
- ▶ In the future, we can adopt multi-hop attention mechanism to learn the key sentences or phrases in the dialogue, and combines these two kinds of information to improve the dialogue quality and the identification of nugget detection ability.



END

Thanks for listening!