Dually Interactive Matching Network for Personalized Response Selection in Retrieval-Based Chatbots

Jia-Chen Gu1, Zhen-Hua Ling1, Xiaoan Zhu2 and Quan Liu1,3
1National Engineering Laboratory for Speech and Language Information Processing, University of Science and Technology of China
2ECE, Queen's University, 3State Key Laboratory of Cognitive Intelligence, iFLYTEK Research

Introduction

Personalized response selection is a task to select a correct response from a list of candidates considering:

(1) the context of the conversation
S1: Hello, how are you doing tonight?
S2: I am well an loving this interaction.
S1: I just got back from the club.
S2: This is my favorite season of a year.
... (2) the persona of the speaker
S1: I like to dance at the club.
S2: I love to meet new people.
S1: Autumn is my favorite season.
...

Problem 1: The context is treated at coarse-grained context-level while each utterance may play different roles.

Method: Compute similarities between each context utterance and each profile at fine-grained utterance-level.

Problem 2: The interactions between the persona and each response candidate are ignored.

Method: Add interactions between the persona and each response.

IMN-Based Persona Fusion

The representation vectors of context, context utterances and persona profiles are denoted as \(\mathbf{c}, \{\mathbf{u}_n\}_{m=1}^{h} \) and \(\{\mathbf{p}_n\}_{m=1}^{h}\) respectively.

Context-level persona fusion:

\[
\mathbf{c}^+ = \mathbf{c} + \sum_{n} \text{Softmax} (\mathbf{c} \cdot \mathbf{p}_n) \mathbf{p}_n
\]

Utterance-level persona fusion:

\[
\mathbf{u}_n^+ = \mathbf{u}_n + \sum_{n} \text{Softmax} (\mathbf{u}_n \cdot \mathbf{p}_n) \mathbf{p}_n
\]

\[
\mathbf{c}^+ = \text{Aggregation} (\{ \mathbf{u}_n^+ \}_{m=1}^{h})
\]

To add another interactive matching between the persona and each response, we formulate the task as a dual matching problem.

(1) Word Representation Layer: general, task-specific and character embeddings.
(2) Sentence Encoding Layer: a single-layer BiLSTM to encode each utterance, response and each profile separately.
(3) Matching Layer: a cross-attention module to derive the matching information by the interactive matching between the context and response, and that between the persona and response.
(4) Aggregation Layer: to convert matching matrices into the final matching vector.
(5) Prediction Layer: a MLP classifier to return a score denoting the matching degree.

Experiments

Dually Interactive Matching Network

Dataset: PERSONA-CHAT

Baseline IMN\_base: IMN-based persona fusion at the context-level
Baseline IMN\_un: IMN-based persona fusion at the utterance-level

Numbers in parentheses indicate gains or losses after adding the persona conditions.

DIM outperformed its baseline IMN\_base by a margin of 14.5% and previous models by margins larger than 27.7% in terms of hits@1 conditioned on original self personas.

Analysis

(1) Ablation tests of removing either persona-response matching or context-response matching in DIM show both contribute.

(2) Transfer tests were conducted by training and evaluating using mismatched types of personas.

Conclusion

In this paper, we formulate the task of personalized response selection as a dual matching problem. A new model named Dually Interactive Matching Network (DIM) is proposed, which achieves a new state-of-the-art performance on PERSONA-CHAT set.

Reference

Zhang S, Dinan E, Urbanek J, Szlam A, Kiela D, and Weston J. Personalizing dialogue agents: I have a dog, do you have pets too? ACL’18
Gu J, Ling Z and Liu Q. Interactive matching network for multi-turn response selection in retrieval-based chatbots. CIKM’19

Contact

Email: gujc@mail.ustc.edu.cn
Github: github.com/JasonForJoy/DIM
Homepage: home.ustc.edu.cn/~gujc/