

Context	我7月20日才到奥克兰呢,你机票订了吗 I will arrive at Auckland in July 20th, have you book the air ticket?	那说明你很熟悉俄罗斯这个国家了啊还开贴干嘛 That means you are familiar with the country Russia, why you open new post
Last Utterance	那太遗憾了我七月二十号就结束旅途了 What a shame, I will finish my trip in July 20th	对啊你都知道 Yes, you have known that
Rewritten-Utterance	7月20日才到奥克兰那太遗憾了我七月二十号就结束旅途了 Arrive at Auckland in July 20th? What a shame, I will finish my trip in July 20th	对啊你都知道俄罗斯这个国家了 Yes, you have known the country Russia
Rewritten-Utterance-RL	机票订了那太遗憾了,我七月二十号就结束旅途了 Air ticket is booked? What a shame, I will finish my trip in July 20th	对啊你都知道国家了 Yes, you have known the country
baseline	这样啊去上学 Go to school like this?	你东北哪噶的啊 Where is your home in northeast of China
Our Model	请问你行程规划了几天 How many days have you planned your trip	俄罗斯不了解 I don't know about Russia
Our Model+RL	可惜了那你还要去吗 What a shame, then are you going there?	说得容易啊 Easier said

Table 7: The examples of end-to-end response selection.

A Model Details

A.1 Context Rewriting Model

We employ the Adam algorithm to optimize the objective function with a batch size of 200. We set the initial learning rate as 0.0004 and reduce it by half if perplexity on validation begins to increase. Both of the encoder and decoder are 1-layer GRU, the word embedding size and hidden vector of encoder GRU are 512, the hidden size of decoder GRU is 1024. We use BPE to do Chinese word segmentation because it can solve the out-of-vocabulary word problem and reduce the size of the vocabulary. The vocabulary size of input and output are 34687. Drop out mechanism is used and equals to 0.3. We use beam search in generating rewritten last utterance and response, the beam size is always 5. For reinforcement learning, λ is equals to 0.1.

A.2 Single-turn Response Generation Model

The model is pre-trained by an encoder-decoder framework with an attention mechanism, where the word embedding size is 512, the encoder is 1-layer bidirectional GRU with 1024 units in both

directions, and the decoder is 1-layer unidirectional GRU with 2048 units.

B Case Study

B.1 End-to-end Retrieval Chatbots

Table 7 presents the selected examples of our models and baselines, our model can extract the keywords from context to construct an informative utterance, which is helpful to retrieve response relative candidates from the index and select the correct response. But the baseline model retrieves candidates by concatenating extracted keywords from the context with the last utterance, the extracted keywords may be useless and even noisy, so some candidates with low correlation to utterances will be retrieved, such error will propagate to next step.

As shown in Table 7, after reinforcement learning, our model tends to generate shorter rewritten utterance than our pre-trained model. And the extracted keywords may be useless or lose important information. It will reduce the quality of retrieved candidates and the quality of the final selected response.