Multi-Turn Dialogue System in Open-Domain

Overview

Multi-turn dialogue system in open-domain is a research hotspot in the field of natural language processing. However, given a user-issued query, it is still challenging for computers to give a feasible reply according to dialogue context. To addresses this issue, this evaluation task mainly targets at generating or retrieving proper replies through better context understanding.

The datasets employed in this task are collected from Sina Weibo, which contain training set and testing set. There are 5,000,000 conversation sessions in the training set and extra 40,000 conversation sessions in the testing set. Examples of the datasets are given below, where the left side is two sessions of the training set and the right side presents two sessions of the testing set.

Training Set

Context: 谢谢你所做的一切
Context: 你开心就好
Context: 开心
Context: 嗯因为你的心里只有学习
Query: 某某某，还有你
Reply: 这个某某某用的好
\n
Context: 你们宿舍都是这么厉害的人吗
Query: 眼睛特别搞笑这土也不好捏但就是觉得挺可爱
Reply: 特别可爱啊

Testing Set

Context: 米芝莲在福州哪儿哈？
Context: 金山万达
Context: 好吃咩？
Query: 不好吃还不如一般菠萝包跟香港的比真的差多了
\n
Context: 第八张是给人家拍的街拍么
Query: 你好烦

Data Format:

- The last sentence of each session is viewed as the reply.
- The second-to-last sentence of each session is defined as the query.
- Other sentences except for reply and query are the context.
- These sessions are separated by blank lines.

- There is no reply in the testing set.

Task Description

This evaluation task consists of two sub-task.

Task 1: Generative Dialogue System

All participants can conduct experiments on the training set. During testing, we will evaluate the performance of varying algorithms on the testing set.

Metric: BLEU 1-4

Task 2: Retrieval Dialogue System

It is necessary for participants to prepare the format of reply themselves. In the process of testing, we will prepare 10 candidates for each dialogue session within the testing set. Among candidates for each session, only one reply is the ground truth while other candidates are randomly sampled from the datasets.

Metric: Precision of retrieved results.