**Evaluation Rules for Dialog Summary Dataset**

This task is to evaluate the quality of dialog summary. First, let's introduce the given task content:

We will give you multiple summary texts to be tested for the same dialog, as well as a standard summary answer. The labeler needs to evaluate the summary texts to be tested by referring to the standard summary answer. The format of the table is as follows:

手机屏幕截图

描述已自动生成

The first column is the dialog number, the second column is the standard summary answer, and the third column is the summary text to be tested. Each standard answer corresponds to 4 summary texts to be tested. Then the next two columns are the indicators to be evaluated, which are informativeness and succinctness.

The detailed rules for each indicator are as follows:

Informativeness: judge whether the information in the standard answer is covered in the summary text to be tested, and the score is 1-5. 1 represents that the main information of the dialogue is basically not in the summary, and 5 represents that the main information of the dialogue is basically all in the summary.

Succinctness: judge whether the information in the summary text to be tested is concise and non-redundant (redundant is defined as: the content is not in the standard answer or there is content duplication), and the score is 1-5.1 represents basically all redundant, and 5 is basically non-redundant. If there is duplicate information in the summary text, the score will be reduced according to the number of duplicate information.