

闲聊系统评价方法

张程

任务型对话的评价

“明天天气怎么样”

precision、recall

客服型对话的评价

“信用卡被盗刷了怎么办”

recall@N

闲聊型对话的评价

“我今天心情不太好”

？ ？ ？

闲聊评价：我太难了

今天心情不太好

开心点啊

我带你去吃夜宵吧

没有标准回答

线上指标

平均对话轮数

留存率：平均session数

信任：话题深度、广度

线上指标

用户直接评价 bot

没想到你挺智能

没想到你挺智障

你可真智能

客观评价

流畅度

今天心情不太好

我带你去撸夜宵吧

我带你去吃夜宵吧

语言模型：
$$p(x_1, x_2 \dots x_T) = \sum_{t=1}^T p(x_t | x_1, x_2 \dots x_{t-1})$$

客观评价

与参考答案的相似度

今天心情不太好

那去跑跑步吧

去跑步吧

词重叠：BLEU、ROUGE、METEOR

客观评价

与参考答案的相似度

今天心情不太好

那去跑跑步吧

去锻炼一会吧

语义相似：词向量距离，相似度模型

主观评价

上线之前，客观之后

开放测试

相关性

主题相关

今天心情不太好

我吃过了

俺也一样

相关性

任务完成度

心情不好，怎么排解？

可怜一番。

可怜一番，跑跑步吧

流畅度

流畅度

今天心情不太好

我带你去撸夜宵吧

我带你去吃夜宵吧

流畅度

逻辑性

今天心情不太好

心情是遭遇的内在表达，提升能力则可以改善遭遇

要是再努力一点，你也许就不会碰到不高兴的事情了

流畅度

精练

今天心情不太好

我一直认为，那个，昨天的心情不好，啊，今天的心情不好，那么呢，明天的心情可能也不会好

昨天和今天就算了，明天一定要高兴起来呀

上下文一致性

逻辑一致性

今天心情不太好

那去跑跑步吧

我不喜欢跑步

嗯，跑步是一个无聊的运动

那挑一个喜欢的运动

上下文一致性

主题一致性

今天心情不太好

那去跑跑步吧

我脚崴了

赶紧去医院看看

是因为脚崴了才心情不好吗

上下文一致性

人格一致性

今天心情不太好

去旅游吧，我猜你和我一样旅游，推荐你去浪漫的土耳其

我不喜欢旅游，喜欢宅着

我也喜欢宅着！来开黑呀

参与度

冒犯

今天心情不太好

呵呵

参与度

多样性

今天心情不太好

那去跑跑步吧

今天心情不太好

跑跑步去吧

参与度

富含信息

今天心情不太好

开心起来吧

心疼一番，我给你唱一首
五环之歌吧

参与度

富含情绪

今天心情不太好

哦，怎么回事

哦？！怎么回事？！

幽默

您为什么没有微信？

没有微信，但人比较有威信

指标冲突：幽默的奥义是反常

实用方案

只为上线

实用方案

合适性

1分或2分：完全答非所问，以及含有不友好内容；

3分：基本可用，问答逻辑正确；

4分：能解决用户问题且足够精炼；

5分：在4分基础上，能让人感受到情感及人设。

实用方案

版本间 good/same/bad 对比

今天心情不太好

哦，怎么回事

哦？！怎么回事？！

总结

- 线上指标

- 平均对话轮数
- 留存率
- 信任
- 用户直接评价 bot

- 客观评价

- 流畅度
- 相似度
 - 词重叠
 - 语义相似

- 主观评价

- 相关性
 - 任务完成度
 - 主题
- 流畅度
 - 流畅度
 - 逻辑性
 - 精练
- 上下文一致性
 - 逻辑一致性
 - 主题一致性
 - 人格一致性

- 参与度

- 冒犯
- 多样性
- 富含信息
- 富含情绪

- 有趣

- 实用方案

- 合适性
- good/same/bad

谢谢!