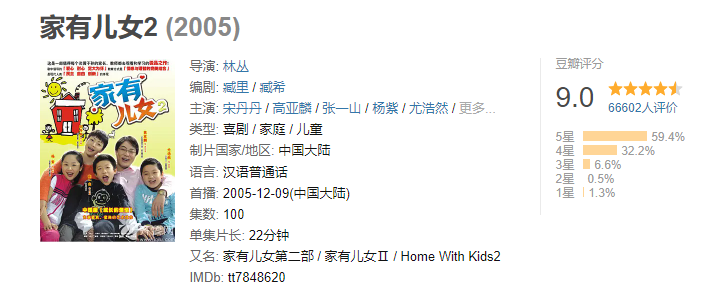
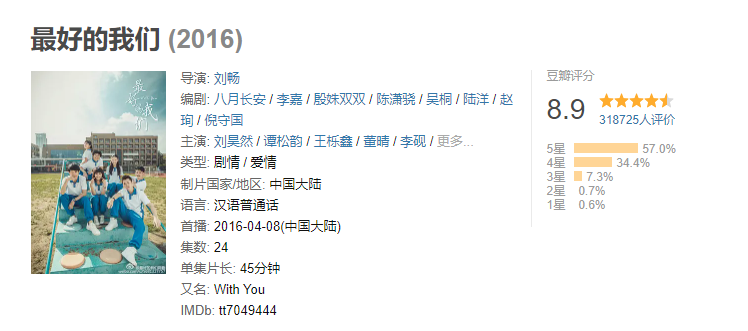
1. +数据集来源

找取豆瓣评分8.0以上的电视剧，电视剧为现代，涵盖了爱情、家庭、悬疑、创业等主题，从互联网上找出相关电视剧每集的台词、剧本，筛选出其中只有两个人对话的对话内容。一共选取八部电视剧，共443集，时长为16055分钟。

* 《家有儿女》2005 共2部/100集 喜剧/家庭/儿童



* 《最好的我们》2016 共24集 爱情



* 《大江大河》2018共两部 86集 剧情/历史



* 《在一起》2020 共20集 剧情（防疫）



* 《我在他乡挺好的》2021 共12集 剧情/爱情



* 《人民的名义》2017 共52集 剧情/悬疑/犯罪



* 《小欢喜》2019 共49集 剧情/家庭



* 《鸡毛飞上天》2017 共55集 剧情



1. 标注内容及规则

对话中有AB两个人，A为说话者，B为倾听者，要根据历史对话生成B（机器人说的话）。情绪、对话行为、原因以及原因来源AB均需要标注，情感需求则标注A。

1. 情绪详解

采用Plutchik的情绪论，8种基本情绪+中性

期待（anticipation）：对未来的未知的某个时刻或者事物产生一种憧憬、向往

快乐（joy）：人的一种开心、高兴的状态

信任（trust）：信任被认为是一种依赖关系，是由个人价值观、态度、心情及情绪、个人魅力交互作用的结果，是一组心理活动的产物

惊讶（surprise）：觉得很意外，很奇怪，很惊异

生气（anger）：因不合心意而不愉快从而非常生气

厌恶（disgust）：是一种反感的情绪

悲伤（sadness）：表示伤心难过

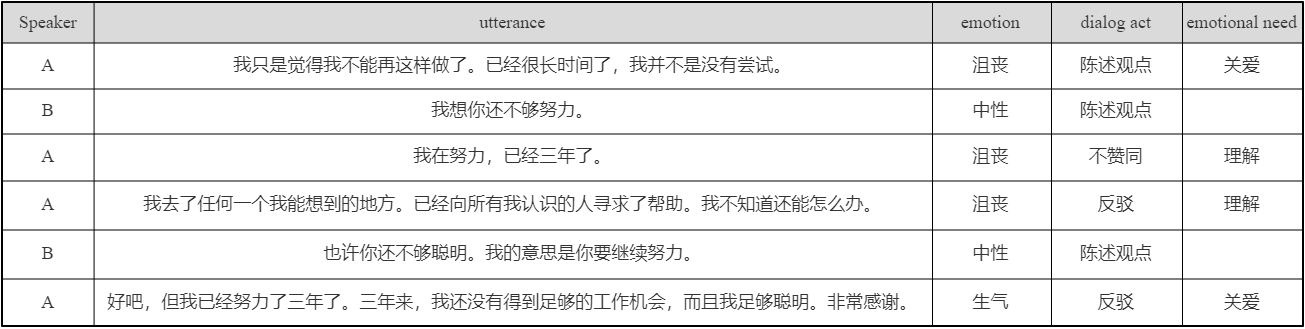
恐惧（fear）：惊慌害怕，惶惶不安

中性（neutral）：无明显情感

1. 对话行为

对话行为为说话者说某句话的动机，我们依据Pareti和Welivita提出的对话行为分类进行整合，得到十五分类的对话行为，具体示意如下：

表1 对话行为标注示例



* 问候（greeting）：问候，问安
* 道歉（apology）：道歉,谢罪,辩白
* 命令（command）：居高临下地掌控、控制
* 回答（answer）：对提出的问题给予解答或对要求表示意见
* 提问（questioning）：提出问题
* 承认（acknowledging）：认可
* 赞同（agreement）：对某种主张表示相同的意见，赞成
* 不赞同（disagreement）：对某种主张表示相反的意见，不赞成
* 陈述（statement-opinion/statement-non-opinion）：有条有理地表达事物、事实或者说话人的看法。
* 安慰（consoling）：使人心情安静舒适
* 鼓励（encouraging）：激发；勉励
* 同情（sympathizing）：对于别人遭遇在感情上产生共鸣
* 希望（wishing）：心想达到某种目的或出现某种情况
* 建议（suggesting）：提出主张

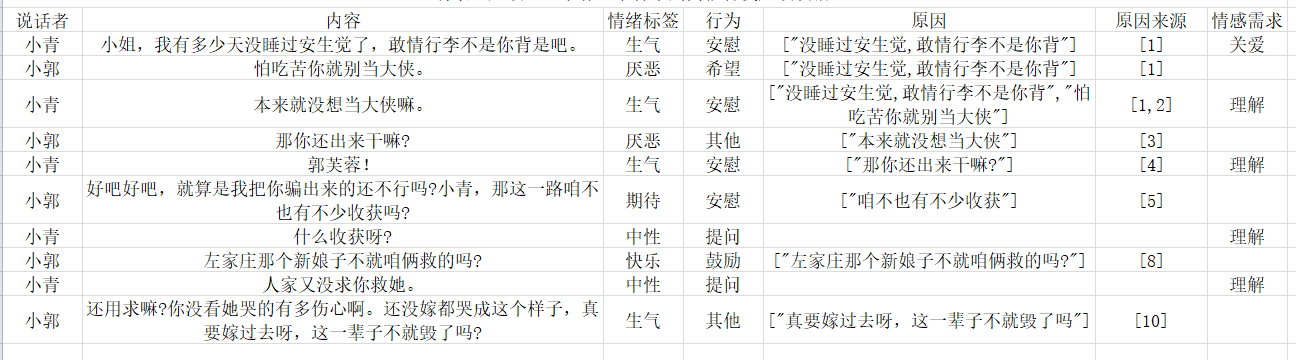
1. 个性

个性即为用户画像，需包括用户性别、用户兴趣爱好等内容，来自于角色介绍以及日常对话中。个性句需以我、我的为开头，至少包含一个名词、动词/形容词、内容词。例子：

我喜欢跑步。

1. 原因跨度以及原因来源

原因来自于话语，要看整个对话，大部分来自于历史话语，要找出原因话语中（例如A说的一句话）原因跨度（例如A说的话语中的连续的最大原因片段）。除去标注原因跨度之外，还要标注原因来源（来自于那个话语）。示例如下：



1. 情感需求

情感需求是指人们对某相应特定情感的缺乏、等待、期盼状态和获取该特定情感的对象物以实现情感满足的强烈愿望。情感需求只需标注A即可。

我们依据Gray撰写的《男人来自金星,女人来自火星》一书选取男女的12种需求，女性需求（关爱、理解、尊重、忠诚、体贴、安全感），男性需求（信任、接纳、欣赏、崇拜、认可、鼓励）

* 关爱(caring)：希望得到关心爱护
* 理解(understand)：希望被懂得，了解
* 尊重(respect)：希望得到尊敬或重视
* 忠诚(loyal)：希望得到真心诚意，无二心
* 体贴(considerate)：希望别人为自己着想，给予关怀、照顾
* 安全感(sense of security)：渴望稳定、安全的心理需求
* 信任(trust)：对方相信自己是诚实、可信赖、正直的
* 接纳(accept)：被接受、采纳
* 欣赏(appreciate)：被认为是好和喜欢
* 崇拜(worship)：希望得到尊敬、钦佩
* 认可(approval)：被许可
* 鼓励(encourage)：被激发、勉励