

Aspect情感分析旅游领域项目

需求

目前想要分两种的情感：积极和消极
比如消费者说：
我准备去日本旅行，可是现在日本的旅行政策太严厉了，想去根本去不了。但是现在三亚的关口开了，我感觉可以去三亚顺便去免税店逛一逛，里面的东西还是挺便宜的。
这里需要判断两个词的情感：日本，和三亚

去某个地区游玩的情感分析，某个城市或国家等

ChatGPT标注数据

因为没有相关领域的标注数据，我们使用chatgpt进行标注

给定的句子是"我准备去日本旅行，可是现在日本的旅行政策太严厉了，想去根本去不了。但是现在三亚的关口开了，我感觉可以去三亚顺便去免税店逛一逛，里面的东西还是挺便宜的。"
给出的情感是：["积极","消极"]
给出句子中涉及的实体词语是：["日本","三亚"]
需要进行的任务是Aspect sentiment analysis，判断给定的实体词语在句子中的情感，回答使用json格式，例如：{"实体1":"积极", "实体2":"消极"}

prompt格式

期待返回格式

{'日本': '消极', '三亚': '积极'}

{'日本': "消极", "三亚": "积极"}

解释：
"日本"这个实体在句子中的情感是消极，因为虽然作者准备去日本旅行，但是现在日本的旅行政策太严厉了，想去根本去不了。这表明作者对于去日本旅行的期望和计划受到了挫折和阻碍，感到失望和沮丧，因此情感是消极的。
"三亚"这个实体在句子中的情感是积极，因为虽然作者无法去日本旅行，但是现在三亚的关口开了，可以去那里旅行，作者感觉可以去三亚顺便去免税店逛一逛，里面的东西还是挺便宜的，这表明作者对于去三亚旅行的想法和计划是积极的，并且对于免税店里的物品也有积极的评价。

chatgptweb

api接口

\n\n{'日本': '消极', '三亚': '积极'}

实际返回格式

有些还是返回中性，因为确实是中性的

例如：{'content': {'雅加达': '中立'}, 'entity': ['雅加达'], 'sentence': '【玩车趣闻】在印尼雅加达，一名男子在一家购物中心外观看一辆由57000罐废弃饮料制成的“卡车”。\u200b'}

对chatgpt的返回格式进行格式化，尝试json处理后，即可得到正确结果

观察1000条数据返回的结果，业务人员进行评估，效果挺好

ChatGLM标注数据

搭建一个自己的ChatGLM模型，设计Prompt，测试返回结果，发现效果不如ChatGPT，暂时不用于Apsect的情感分析

数据预处理

原始数据统计

3个excel，大概242万条数据，列数是21，字段包括

'da_index', 'channel', 'url', 'datetime', 'keyword', 'content', 'title', 'forward_content', 'nickname', 'userid', 'reposts', 'comments', 'likes', 'u_url', 'tag_Sheet1_keyword', 'tag_Sheet1_id', 'tag_Sheet1_aspect', 'gender', 'city', 'introduce', 'collections'

keyword字段整理去重后剩余

627条数据

过滤后还剩余: ['奥克兰', '欧洲', '英国', '胡志明市', '海口', '悉尼', '日上', '美国', '葡萄牙', '马尔代夫', '阿根廷', '丹麦', '清迈', '俄罗斯', '旧金山', '斯里兰卡', '吉隆坡', '三亚', '济州岛', '意大利', '墨尔本', '札幌', '阿姆斯特丹', '韩国', '柬埔寨', '伦敦', '北海道', '澳大利亚', '海棠湾', '东京', '澳洲', '巴厘岛', '普吉岛', '新加坡', '出国', '曼谷', '日本', '芭提雅', '米兰', '洛杉矶', '纽约', '苏梅岛', '首尔', '斐济', '土耳其', '名古屋', '里斯本', '巴黎', '荷兰', '加拿大', '希思罗', '印度尼西亚', '海南', '大阪', '挪威', '毛里求斯', '香港', '东南亚', '莫斯科', '匈牙利', '澳门', '奥地利', '法国', '雅加达', '迪拜', '罗马', '马尼拉', '西班牙', '泰国', '古巴', '京都']