# 大模型引入

功能：“大模型分析”模块是对其他分析模块（热点话题识别、情感分析、趋势分析）输出的数据进行整合，生成最终的舆情分析报告。通过引入大模型（本报告使用chatgpt大模型），可以实现对多源数据的综合分析，提供更深层次的洞察和预测。

* 1. 数据准备与预处理

1.1.1数据收集

从热点话题识别、情感分析和趋势分析模块收集输出的数据。确保每个模块输出的数据准确有代表性，并且格式统一，便于后续整合。

热点话题识别模块输出：每条记录包含具体话题、关键词、热度评分等信息。

情感分析模块输出：每条记录包含文本内容、情感分类（正面、负面、中性）和情感评分。

趋势分析模块输出：每条记录包含文本内容、趋势指标（增长率、未来发展方向）、趋势评分等信息。

1.1.2 数据整合

使用Pandas库将不同模块的输出数据整合为一个综合数据集。包括加载模块数据，数据清洗，数据合并，数据转换（转换成向量，便于大模型输入）等步骤。

* 1. 训练 ChatGPT 的 Prompt 设计

1.2.1 热点话题识别模块

目的：让 ChatGPT 进行整体热度分析，生成一个综合的热度分析报告，包含每个话题的热度对比、总体趋势和建议。

【输入数据格式】

{"话题": "具体话题", "关键词": ["关键词1", "关键词2", "关键词3"], "热度评分": 具体评分}

【Prompt示例】

请基于以上数据进行整体热度分析，并生成一份报告。

输出格式：

1. 每个话题的综合热度分析，包括关键词和热度评分。

2. 各话题的热度对比，说明哪些话题最热门，哪些话题相对较冷。

3. 总体趋势分析，总结当前舆情的主要方向和变化趋势。

4. 提出建议，说明如何利用当前的热点话题来制定策略。

为了确保模型输出符合预期，可以进行如下调试：

- 确保每个话题的热度分析准确反映关键词和评分。

- 对比各话题热度，突出最热门和最冷门的话题。

- 总体趋势应总结当前数据中的主要方向和变化趋势。

- 提出的建议应结合热度分析结果，具有可操作性。

【关键字】热度对比、最热门话题、相对较冷话题、总体趋势、策略建议

1.2.2 情感分析模块

目的：让 ChatGPT 进行整体情感分析，生成一个综合的情感分析报告，包含情感分布、总体情感趋势和建议。

【输入数据格式】

{"评论": "具体评论内容", "情感倾向": "正面/中性/负面", "情感评分": 具体评分}

【Prompt示例】

请基于以上数据进行整体情感分析，并生成一份综合报告。

输出格式：

1. 情感分布分析，包括正面、负面和中性情感的比例。

2. 总体情感趋势，总结当前评论的主要情感方向。

3. 提出建议，说明如何利用当前的情感分析结果来制定策略。

为了确保模型输出符合预期，可以进行如下调试：

- 确保情感分布分析准确反映数据中的情感倾向和评分。

- 总体情感趋势应总结当前数据中的主要情感方向。

- 提出的建议应结合情感分析结果，具有可操作性。

【关键字】情感分布、正面情感、负面情感、中性情感、总体趋势、策略建议

1.2.3 趋势分析模块

目的：让 ChatGPT 进行整体趋势分析，生成一个综合的趋势分析报告，包含趋势分布、总体趋势和建议。

【输入数据格式】

{"评论": "具体评论", "趋势预测": "具体趋势预测", "未来发展方向": "具体未来发展方向", "趋势评分": 8}

【Prompt示例】

请基于以上数据进行整体舆情趋势分析，并生成一份综合报告。

输出格式：

1. 趋势分布分析，包括各趋势的比例。

2. 总体趋势，总结当前评论的主要趋势方向。

3. 提出建议，说明如何利用当前的趋势分析结果来制定策略。

为了确保模型输出符合预期，可以进行如下调试：

- 确保趋势分布分析准确反映数据中的趋势预测和评分。

- 总体趋势应总结当前数据中的主要趋势方向。

- 提出的建议应结合趋势分析结果，具有可操作性。

【关键字】趋势分布、趋势明显、趋势较弱、总体趋势、策略建议

1.2.4 综合分析

为了能使大模型更好的切合公司现状以及发展方向，可以补充背景信息、发展方向等其他Prompt，如：

背景信息：

- 公司在最近一个季度发布了一款新产品，市场反响热烈，获得了很多好评。

- 最新市场调查显示，消费者对公司的服务满意度有所下降，需要关注服务质量。

- 许多公司在科技创新方面投入了大量资源，本季度的热门话题是科技研发和创新。

* 1. 综合分析生成报告

生成自然语言报告需要从模型推理结果中提取关键信息，并将这些信息整合成易于理解的自然语言形式。

1.3.1 情感分析部分

使用大模型对数据进行综合分析阐述，对于结果加以概括。情感分析部分需要描述整体情感分布，例如：“在分析的文本中，正面情感占比为X%，负面情感占比为Y%，中性情感占比为Z%。整体来看，xx情感站主导，可见大部分客户对于本产品持xx态度。”同时给出相应具体实例，说明各类情感的典型文本。

1.3.2 热点话题部分

使用热点话题识别结果，描述每个热点话题的关键词和热度。列出识别出的热点话题及其热度，例如：“本次分析共识别出5个热点话题，分别为A、B、C、D和E，其中话题A的热度最高。”同时提供每个话题的关键词和相关示例。

1.3.3 趋势分析部分

对大模型不断投喂与训练，针对现有数据尝试预测未来舆情变化与发展。对于所得结果描述舆情变化的趋势，例如：“在过去一个月中，关于公司的正面情感呈上升趋势，负面情感有所下降。”描述预测未来的发展方向，例如：“预计未来两周内，正面情感将继续上升，负面情感将保持稳定。”

1.3.4 综合分析与结论部分

总结各部分的分析结果，提出整体的舆情结论。对于以上结果，进行综合分析，例如：“综合以上分析结果，可以得出结论，公司在近期的公众舆情中表现良好，正面情感占多数。”同时提出建议或行动方案，例如：“建议公司继续保持良好的市场表现，针对负面情感较集中的话题进行重点关注和处理。”

最终将上述部分整合，确保上述部分有有效逻辑连接，整体连贯一致。之后选择合适的格式进行输出。

# 可视化展示

功能：针对分析层各模块所得数据进行可视化展示，辅助对于舆情分析的综合判断。其中工具所需工具有：

1. Matplotlib：功能强大，适合绘制各种类型的图表。适用于趋势分析曲线图、词频分布直方图等。

2.Seaborn：基于Matplotlib，提供更加美观和高级的绘图接口。适用于热力图、分布图、回归图等。

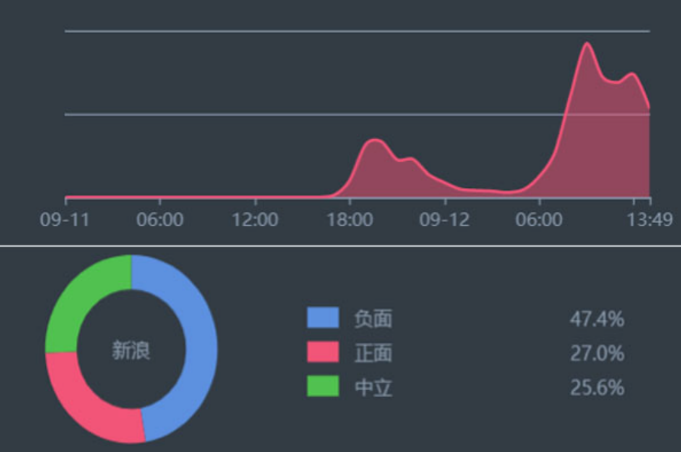
3.Plotly：交互性强，适合制作动态和交互式图表。适用于交互式趋势图、词云等。

4. D3.js：高度可定制化，适合生成复杂的动态和交互式图表。适用于需要高度定制化和交互性的图表。

2.1舆情分析曲线图

从分析层获取舆情趋势分析的数据，包括时间和对应的情感指数。并对数据进行清洗和整理，确保时间序列数据完整且有序。使用Matplotlib绘制舆情趋势分析曲线图

功能：直观具体的看出舆情走势以及未来发展趋向。

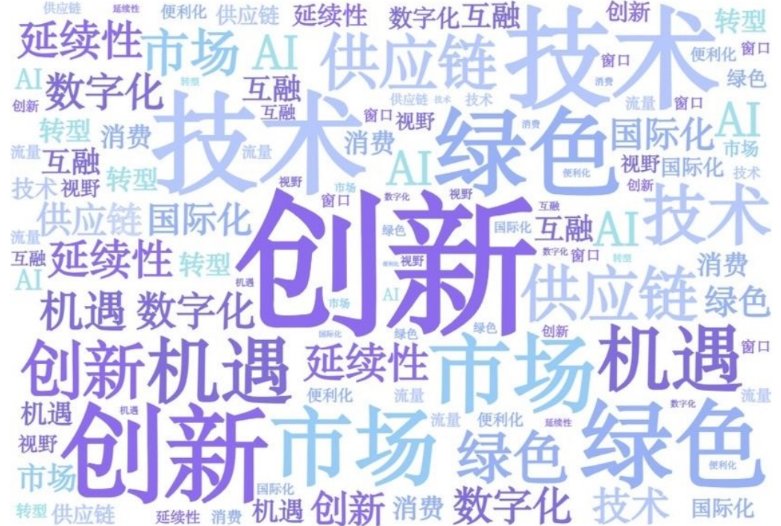


热点趋势分析曲线图示例

2.2热点词云图

从热点话题识别模块获取关键词及其频数并对关键词进行计数，整理成频数数据。使用Plotly绘制热点词云图。

功能：明确了解公众对于产品的评价，辅助公司对于产品的公众定位。

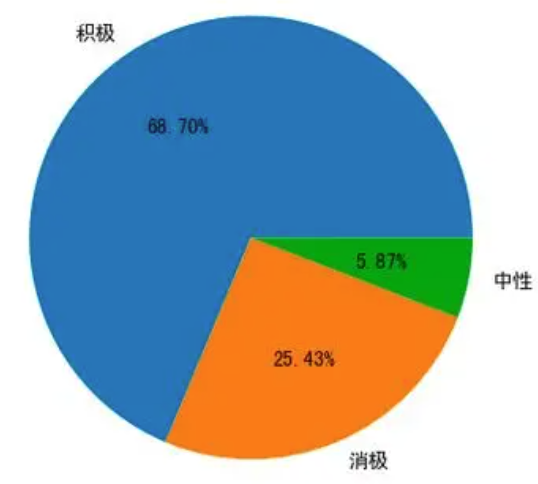


热点词云图示例

2.3情感分布饼图

从情感分析模块获取各情感类型的比例数据，整理成比例形式使用Matplotlib绘制情感分布饼图。

功能：直接了然地看出公众对于产品的情感态度，辅助公司对产品的进一步调整。

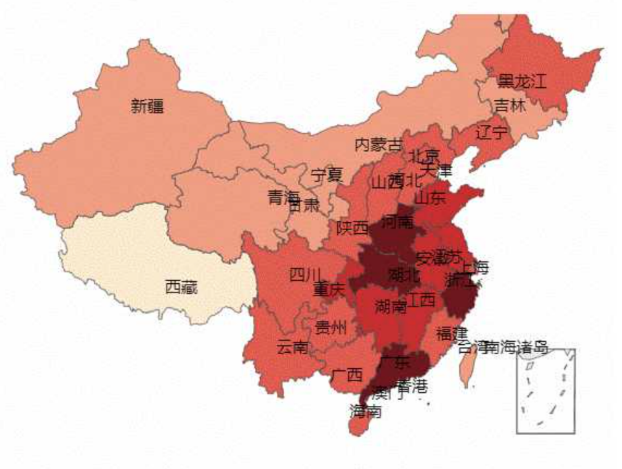


情感分析饼图示例

2.4热力图

从分析层获取热点话题与情感分布的交叉数据。将数据整理成矩阵形式，以便绘制热力图并使用Seaborn绘制热力图。

功能：整体高效地看出公众对于产品的态度并通过关键字分析出客户需求，利于公司对产品的进一步优化。



热力图示意