

测试报告-实验文档

组长：欧立言 231250088

组员：王钰荣 231250094 岑若琛 231250089

项目仓库：<https://github.com/Wang-Y-R/DataScience-Group>

1. 数据预处理

1.1 图文融合与信息提取

1. 图片信息的提取：

对图片信息的提取方法：

1. OCR光学识别图片文本内容
2. 利用小组件识别模型识别图标中存在的组件（按钮等）
3. 使用AI大语言模型分析图片内容

图文融合的方法：

1. 将图片的信息提取出来，与报告文本用某种方式拼接，共同计算
2. 采用特征拼接的方法

而我们最终选择了**AI大语言模型分析**的方法。同时为AI模型提供问题描述的文本以及对应的图片，让AI在软件测试的情境下综合分析图片和文本，并结合二者给出新的问题描述。将其给出的描述作为图文融合的结果。

选择此方法而不选择其它方法的理由：

1. OCR光学识别图片文本再与描述拼接的方法不适合所有的图片。大部分的图片上有大量的冗余信息，亦或是只有图形按键而没有文本。用OCR光学识别图片文本的方法并不妥当。而AI大语言模型则可以分析图片，识别出那些内容是重要的信息。并且AI模型还能分析图片内的非文本的内容，效果更好。
2. 不使用小组件模型识别的方法是因为AI模型分析图片的过程就已经包括了对图片上各种小组件的识别分析。

2. AI模型图文融合：

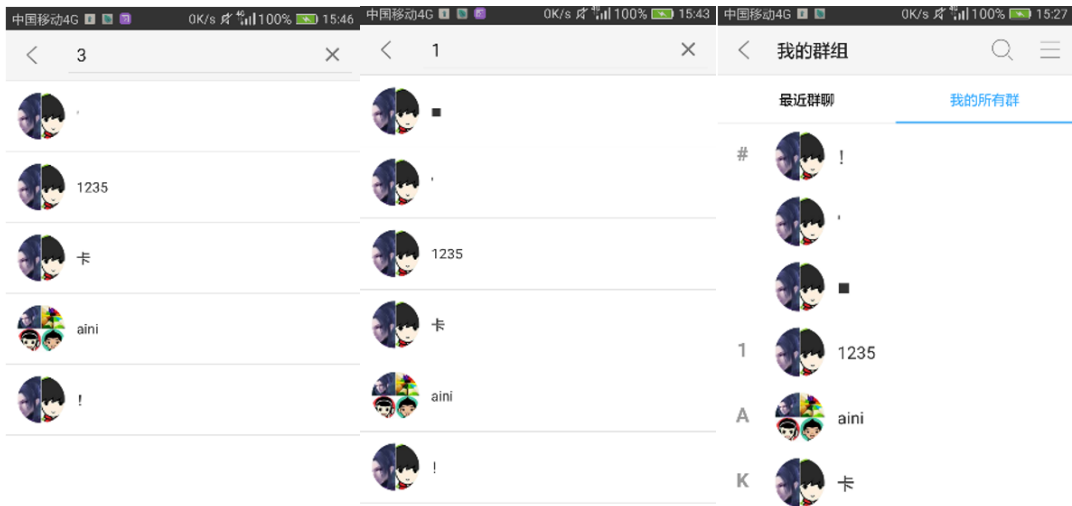
选择的是讯飞语言大模型。选择的理由如下：

1. 讯飞语言大模型提供了相当数量的免费tokens供我们使用
2. 讯飞语言大模型中提供了图片分析的模型，符合我们的要求。
3. 代码文件路径 "1.GraphAndTextFusion\fusion.py"

3. 图文融合的效果示例

1. ID: 10010000009320

问题描述：在我的群组中通过搜索框进行群组搜索，输入数字，出现与数字不匹配的群名称。
相应的图片：



AI的分析结果：在“我的群组”页面中，当用户尝试通过搜索框输入数字进行群组搜索时，系统返回了与输入数字不匹配的群组名称。例如，用户可能会输入“1235”，但系统返回的群组名称却与“1235”无关。这可能导致用户难以找到他们想要的群组，从而降低了用户体验。

2. Id: 10010000009233

问题描述：卸载应用后，最近群聊和消息记录全部清空了，只可以在我的群组里查看到我的所有群。

相应的图片：



AI分析结果：在执行应用的卸载操作后，用户的最近群聊和消息记录被意外地清空了。此外，用户只能在我的群组里查看到他们之前加入的所有群聊，而无法访问任何具体的聊天内容或历史消息。这可能表明在应用的卸载过程中，相关的数据存储和检索功能出现了问题。

可以看出：AI模型融合了图片和文字，重新对问题做出了描述，实现了图文融合。

图文融合后的数据路径："1.GraphAndTextFusion\afterFusion.xlsx"

1.2 去除停用词和清除垃圾数据

我们结合数据大体特征以及通用的停用词列表制定了一份适合的停用词列表用来清除潜在的干扰词/无用词，并使用jieba库实现分词，对于分词后词组数少于5的报告认定为垃圾数据并清除。在第一次分词后我们进行了词频统计，将出现次数过多的词组认定为报告共有的词组并将其加入停用词列表，进行第二次预处理。

预处理数据路径: "2.Preprocessing\afterPreprocessed.xlsx"

词频统计数据路径: "2.Preprocessing\WF_afterPreprocessed.xlsx"

2. 特征提取

2.a Sentence-Bert特征提取法

1. **思路**: 利用huggingface提供的接口, 使用sentence-bert的方法将中文的语句提取特征, 转化为向量, 用于后续的聚类分析。
2. **提取向量**: 使用**bert-base-chinese模型**, 利用huggingface的接口对图文融合的问题描述提取语句向量。
 1. 遇到的问题: 由于样本集数据量过大(1400+条)的数据, 且模型计算花费时间较长, 资源开销大。
 2. 解决方法: 我们将图文融合的数据分五次分别进行特征提取, 将向量依次存储。在后续使用时再读取。解决了性能问题, 同时避免重复计算, 浪费资源。
 3. 向量文件路径: "3.a.SentenceBert\embeddings*.npy"
3. **向量的降维**: 由于模型计算出来的向量过大(shape 为 $1400 * 128 * 768$)的三维向量, 此处先对其降维。考虑PCA降维方法与T-SNE方法。由于向量太大, T-SNE方法效率低。故采用**PCA降维**方法, 将向量降为2维, 用于后续聚类分析以及散点图绘制。

2.b TF-IDF矩阵

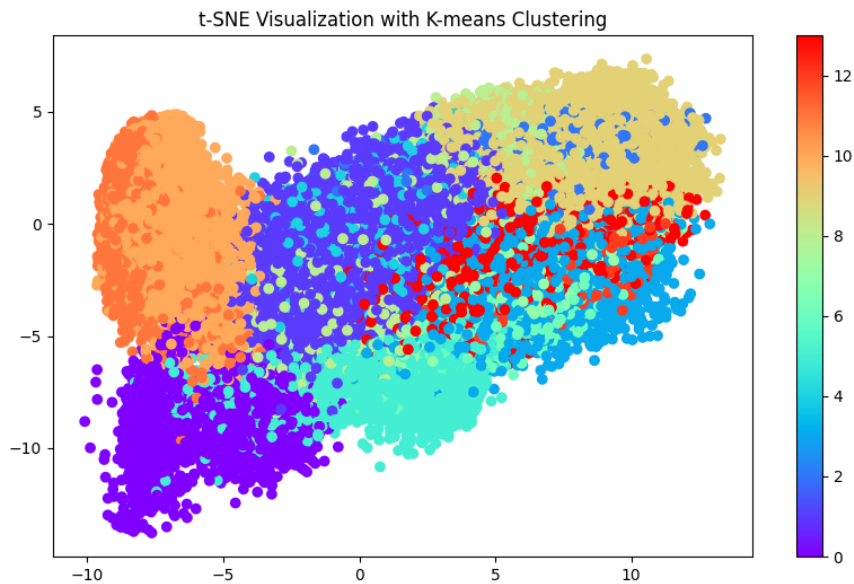
1. **思路**: 利用去除停用词和jieba分词的数据生成TF-IDF向量并构建TF-IDF矩阵, 用于后续的聚类分析。TF-IDF向量是基于词频构建的, 能有效反映数据的内容和特征。
2. **生成TF-IDF向量**:
 1. 遇到的问题: ①大模型的回答模板中会重复提及回答模板用词(例如: “用户”, “功能”), 会提高向量的相似度, 模糊了聚类。②数据量过大, 词语次数过多, 导致TF-IDF向量维度溢出, 不利于聚类分析。
 2. 解决方法: ①二次处理: 预处理阶段的词频统计对词语进行分析与筛选, 重新确定停用词, 再次清洗数据。②筛选高频词语: 根据词频统计结果, 筛选出词频大于统计标准的词语, 将其作为TF-IDF向量的维度。
3. **构建TF-IDF矩阵**: 对于筛选出的高频词语, 生成TF-IDF向量, 将所有文档的TF-IDF向量组合起来, 构建一个TF-IDF矩阵
文件路径: "3.b.TF-IDF\TFIDFMatrix10.xlsx"

3. 聚类分析及结果

3.a Sentence-Bert特征提取法的聚类及其结果分析

1. **聚类分析**: 采用K-means聚类方法: 使用语句向量作为输入, 通过多次尝试确定适合的聚类数, 执行K-means聚类算法。并将聚类结果分类存储在excel之中。

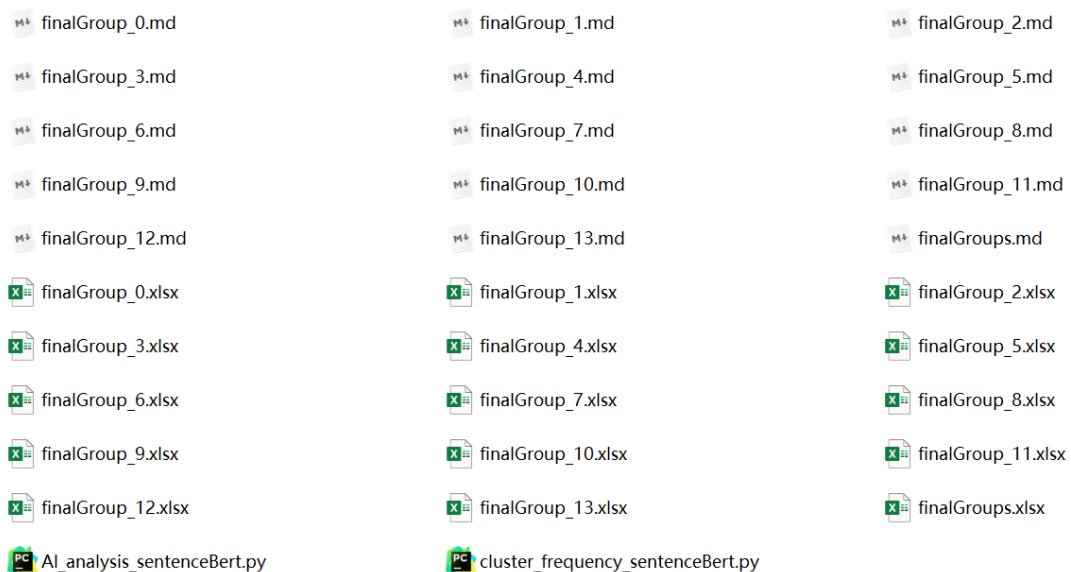
2. **聚类散点图绘制**：以14聚类数为例，不同的点代表不同样本，颜色相同代表同一个类别



3. 采取两种方法对聚类结果进行考察分析

1. **主题词提取**：对不同聚类进行词频统计，将频数最高的定义为该聚类的主题词，之后可生成词云进行问题分析。
2. **大模型分析**：通过把聚类后的报告使用智谱清言GLM4大模型分析，找出关键问题，为报告撰写提供材料。

4. **结果展示实例**：以14聚类数的结果为例，Excel表格中存储该类别对应的信息以及词频结果，Markdown文本中存储大模型分析的结果



1. 从**词频结果**来看，每一个类别都有着许多相同的大量出现的词语，对直接识别该类的特征造成一定的障碍。以14聚类数为例，选取三个类的词频统计情况。可以看出，类似“群组”，“群”，“系统”之类的词语大量的出现。说明这些词语并不是这些类区分的关键要素。但他们可能最终影响到了聚类的结果。

1	Word	Frequency
2	群组	85
3	群	82
4	系统	55
5	没有	53
6	搜索	51
7	加入	45
8	无法	45
9	界面	41
10	群聊	41
11	名称	39
12	提示	36
13	信息	33
14	好友	33
15	点击	32
16	成员	32
17	添加	27
18	消息	27
19	输入	26
20	设置	26
21	聊天记录	24
22	软件	23
23	错误	23
24	出现	23
25	创建	22
26	进入	21
27	限制	19
28	超过	19
29	预期	19
30	情况	18
31	并未	18
32	查找	18
33	进行	18
34	邀请	17
35	选择	16
36	提供	16

图一

1	Word	Frequency
2	群	286
3	群组	285
4	搜索	254
5	没有	206
6	群聊	195
7	无法	191
8	界面	179
9	系统	169
10	输入	162
11	添加	156
12	信息	139
13	点击	135
14	创建	133
15	消息	125
16	成员	123
17	好友	118
18	名称	116
19	提示	115
20	选择	112
21	加入	111
22	软件	106
23	进行	98
24	联系人	94
25	出现	94
26	预期	85
27	设置	84
28	错误	79
29	进入	79
30	限制	78
31	聊天	77
32	页面	77
33	提供	67
34	查找	65
35	按钮	63
36	正确	62

图二

1	Word	Frequency
2	群组	22
3	输入	18
4	无法	18
5	系统	17
6	发送	15
7	没有	14
8	搜索	13
9	群	12
10	群聊	12
11	添加	10
12	昵称	10
13	创建	10
14	成员	9
15	界面	9
16	名称	9
17	内容	9
18	提示	8
19	消息	8
20	信息	8
21	点击	8
22	出现	8
23	表情	8
24	手机	8
25	过程	7
26	进行	7
27	加入	7
28	图片	7
29	查找	6
30	聊天	6
31	请求	6
32	设置	6
33	错误	5
34	头像	5
35	允许	5
36	限制	5

图三

改进思路：可以在数据预处理/图文融合后对结果进一步处理，去除如“群组”，“群”，“系统”之类的词语，使得不同问题的区别更加突出。要提取每个类别的特征需要从排名15或20+的词语中提取。如图一、二、三的核心词分别为：聊天记录、联系人、表情/图片。这些核心词可能才是对应类别的特征。虽然最终得到的结果并不十分突出显著，但从这些核心词来看，该方法在一定程度上可以对大量的图文信息做出一定的分类。

2. 从**AI分析**结果来看，以14聚类数为例，选取三个类的分析情况，可以看出AI模型分析在一定水平上可以体现出类的不同之处。但效果不佳。原因如下：①同属一类的问题文本量较大，AI处理概括的能力也有限。②如词频分析所属，许多问题在描述方面存在许多公用的词语，AI概括中也会大量出现此类词语，并不能体现出核心词。难以反映类的特征。

finalGroup_7

从提供的软件测试报告中，可以分析出以下一到两个主要问题：

1. **用户界面和体验问题：**
 - **界面显示错误：**软件界面存在显示错误，如登录进度条未完成、图片中书籍堆叠不整齐等问题，这影响了用户的视觉体验，造成混乱和不稳定感。
 - **搜索和加入群聊的体验问题：**搜索框在输入特定序列或数字时没有给出有效的提示或指导，搜索结果与用户的查询意图不符，群聊人数信息未及时更新，这都会导致用户在搜索和加入群聊时感到困惑，降低用户体验。
 - **操作流程不符合用户预期：**例如在创建群聊时，系统没有按照规定的群名称长度进行限制和提示，群设置中的选项与用户初始选择不符，操作人员非群主却能直接添加新成员等，这些都不符合用户的预期行为，增加了用户的操作难度。
2. **功能性和交互设计问题：**
 - **功能缺失或不完善：**如缺少群名称长度提示、群聊加入方式表述不清、缺乏网络断开提示、缺少确认跳转按钮等，这些都是功能设计上的缺陷，导致用户在使用过程中遇到不便。
 - **用户交互设计不足：**软件在搜索群聊时必须点击放大镜才能输入信息，缺少对特殊字符输入的反馈，以及没有利用定位功能为用户提供附加价值，如显示活动定位、帖子定位等，这些都是交互设计上的不足，未能充分考虑用户的使用习惯和需求。
 - **推荐系统缺失：**报告中提到缺少推荐群的功能，这限制了用户之间的互动和社交体验，减少了用户粘性。

finalGroup_8

从提供的软件测试报告中，可以分析出以下一到两个主要问题：

1. **群组管理和邀请机制问题**
 - 在群组设置中，当群主设置了拒绝加入群请求并由管理员审批时，成员却可以随意邀请他人加入，且群主和管理员未收到任何加入请求的信息。这表明群组邀请和管理机制存在严重缺陷，可能导致权限控制和群组安全管理失效。
2. **搜索功能准确性问题**
 - 软件的搜索功能似乎存在多个准确性问题。例如，“ceshi”群组名称被错误地重复显示，搜索结果与输入关键词不匹配，以及使用阿拉伯数字搜索时出现无关群组。这些问题表明搜索算法或数据处理逻辑可能存在错误，导致用户无法获得精确的搜索结果。

以下是其他一些次要问题：

- **用户界面和体验问题：**如注册页面缺少微博直接登录选项、重复加群请求、无法预览发送内容、昵称更新不一致、界面设计过于简单等，这些问题共同影响了用户体验。
- **功能性和性能问题：**如闪退现象、无法收藏发送的表情、搜索功能模糊匹配问题、公告推送逻辑错误、图片发送不支持裁剪、聊天记录保存数量有限等，这些问题涉及软件的功能完整性和性能稳定性。
- **数据输入和显示问题：**如搜索字段无限字符输入、创建群组时换行符自动替换为空格、输入限制过于严格等，这些问题影响了用户的数据输入和显示效果。

finalGroup_10

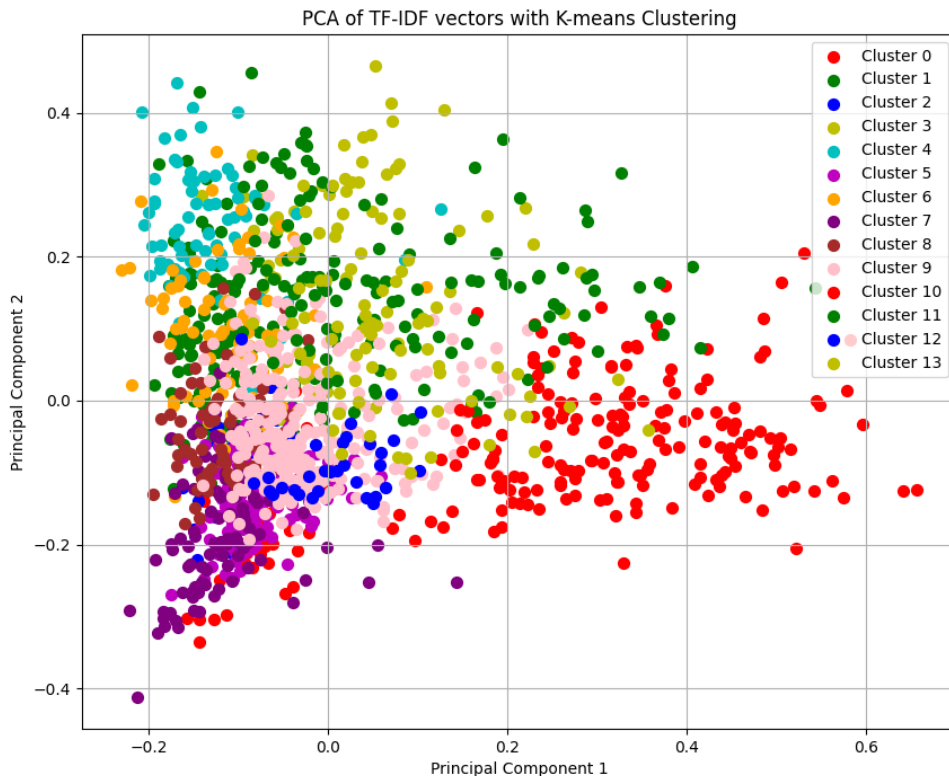
根据测试报告，软件存在以下主要问题：

1. **群组功能问题：**在创建群组时，可以添加超过20个成员，但最后创建时才提示超出人数限制，导致填写信息作废。建议在添加超过20个成员时就给出提示。
2. **搜索功能问题：**搜索群组时，搜索结果与输入关键字相关性不强，导致难以找到目标群组。建议优化搜索算法，提高搜索结果的相关性。
3. **用户隐私问题：**非好友用户可以直接被拉入群聊，缺乏隐私保护。建议增加拉入群时需要用户确认的功能。
4. **用户体验问题：**部分功能的操作流程繁琐，如添加群成员时需要逐个选择。建议优化操作流程，提高用户体验。
5. **界面显示问题：**部分页面布局不合理，信息展示不直观，需要优化界面设计。
6. **功能缺失问题：**部分功能如修改好友备注、设置聊天背景等缺失，需要完善相关功能。
7. **交互逻辑问题：**部分功能的交互逻辑存在问题，如删除群组聊天记录后仍可再次删除，不符合用户预期。

综上所述，软件在群组管理、搜索、用户隐私保护、用户体验设计等方面存在问题，需要进行优化和改进。

3.b TF-IDF矩阵的聚类及其结果分析

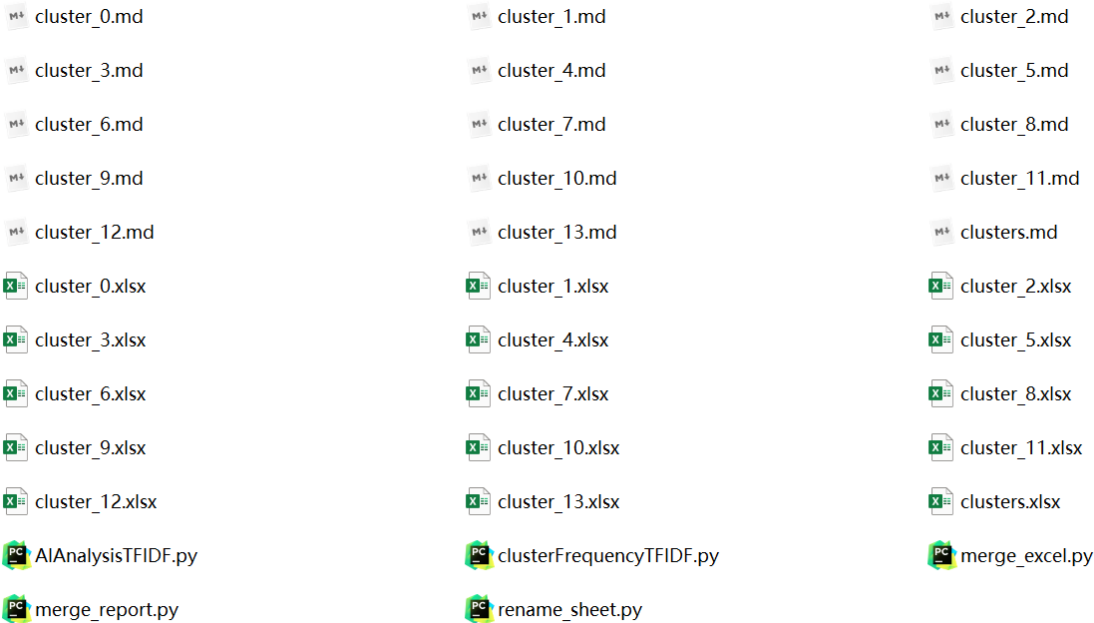
1. **执行K-means聚类**：使用TF-IDF矩阵作为输入，执行K-means聚类算法。
2. **确定最佳聚类数**：选择具有最高轮廓系数的聚类数，重新执行K-means聚类将报告分为若干个聚类。（最佳聚类数为14）
3. **PCA降维**：使用PCA降维到2维，绘制出所有聚类的散点图，以展示它们在特征空间中的分布。



4. 采取两种方法对聚类结果进行考察分析

1. **主题词提取**：对不同聚类进行词频统计，将频数最高的定义为该聚类的主题词，之后可生成词云进行问题分析。
2. **大模型分析**：通过把聚类后的报告使用智谱清言GLM4大模型分析，找出关键问题，为报告撰写提供材料。

5. **结果展示实例：** Excel表格中存储该类别对应的信息以及词频结果， Markdown文本中存储大模型分析的结果。



1. 从**词频统计**结果来看，每个聚类的词频统计中都有1~2个频数较为突出的词，我们可以将其视作该聚类的主题词并推测该聚类集中体现的软件问题。比如图中第一个统计反映了聊天记录删除和清空功能可能存在缺陷，第二个统计反映了软件在添加好友模块可能存在缺陷，第三个统计反映了软件在搜索功能上可能存在缺陷。但词频统计也会记录不同聚类的共有词（例如“群聊”），可能会影响聚类结果和问题判断，这说明需要在二次处理阶段对共有词进行去除。总的来说，该方法可以在一定程度上对大量的图文信息进行主题词标注并分类。

Word	Frequency	Word	Frequency	Word	Frequency
聊天记录	86	好友	187	搜索	460
删除	22	添加	135	群	126
查看	20	联系人	128	输入	117
消息	19	群聊	108	进行	80
清空	18	选择	63	无法	75
聊天	14	列表	52	群组	72
仍然	12	界面	50	信息	72
界面	12	无法	50	没有	71
历史	11	创建	45	查找	62
点击	11	点击	44	系统	57
无法	11	没有	41	点击	55
预期	10	最近	39	界面	46
软件	10	搜索	37	匹配	41
时间	9	软件	36	群聊	41
记录	9	群组	34	出现	36
没有	9	出现	33	找到	35
页面	9	系统	32	预期	33
查询	8	成员	32	软件	33
搜索	8	发起	29	图标	32
群聊	8	页面	28	加入	29
查找	7	群	28	提供	27
设置	7	头像	27	并未	27
群	7	错误	21	名称	26
信息	7	信息	21	正确	26
.	7	选中	20	添加	25

2. 从**AI分析**结果来看，AI在一方面对特定的聚类报告总结提炼出软件集中体现的问题，展示聚类的特征，另一方面面对文本量较大的报告或受限于自身概括能力时，AI的分析便显得过于分散

无法突出重点。但从总体上看AI对各个聚类的分析依然能在一定程度上反映了聚类的特征与软件的问题，说明该方法可以实现对图文信息的分类与信息提炼。

cluster_0

从这份软件测试报告中，可以分析出以下一到两个主要问题：

1. **聊天记录清空功能缺陷：**

报告中多次提到，用户在执行清空聊天记录的操作后，聊天记录并没有被完全删除，无论是在群聊还是私聊中。这个缺陷表现在多个方面：

- 即使清空了聊天记录，用户仍然可以在历史消息中查看到被删除的记录。
- 清空操作后，部分历史消息仍然可见，导致用户的隐私可能受到威胁。
- 在不同的设备或软件重新安装后，聊天记录的状态更新不正确，用户无法获取最新的聊天信息。

这个问题的主要影响是用户的隐私安全和数据完整性。用户期望清空操作能够彻底删除聊天记录，但实际上软件并没有按照用户的预期工作。

2. **聊天记录搜索和显示功能问题：**

- 用户在进行关键字搜索时，系统没有返回正确的结果，即使本地存在相关聊天内容。
- 在查看历史消息时，聊天记录显示不全，或者时间顺序混乱，或者页面布局不友好，导致用户难以进行有效查找和阅读。
- 搜索功能的界面设计存在错误，如搜索按钮重叠、显示混乱等，影响了用户的使用体验。

这些搜索和显示方面的问题影响了用户查找信息的效率和体验，导致用户可能无法快速准确地找到所需的聊天记录。

综上所述，这份测试报告中的主要问题集中在聊天记录的清空功能以及搜索和显示功能的缺陷上，这些问题需要开发团队重视并尽快解决，以确保用户的隐私安全和使用体验。

cluster_5

根据软件测试报告的内容，我们可以总结出以下几个问题：

- 表情显示问题：**在聊天界面中，用户输入的表情符号有时无法正确显示，甚至出现系统默认表情，或者输入的表情与实际发送的表情不一致。
- 语音播放问题：**用户在群聊中发送的语音消息，接收方听不到声音，或者声音太小，无法调整音量。
- 图片发送问题：**用户在发送图片时，出现图片无法预览、发送顺序错乱、无法裁剪图片等问题。
- 文件发送问题：**用户发送的文件无法预览，需要下载后才能查看；同时存在文件发送失败、无法取消上传等问题。
- 公告功能问题：**公告编辑功能不完善，无法保存历史公告，公告无法正常显示表情等。
- 评论显示问题：**用户发表评论后，评论数未立即更新，需要手动刷新才能看到。
- 语音消息问题：**语音消息发送失败，无法双向收听，听到的语音与实际声音不同等问题。
- 界面交互问题：**表情框遮挡文字，文字与图片重叠，无法设置聊天背景等问题。
- 操作逻辑问题：**用户界面缺乏提示，操作不符合预期，如上传图片后无法选择，表情显示错误等。
- 性能问题：**软件在处理大量表情时出现卡顿或崩溃。

4.结论

两种方法的对比

得出哪一种的分类结果更清晰

oly：我感觉tfidf矩阵中主题词更突出，所以聚类效果更好？