



200233

上海桂平路 435 号 上海专利商标事务所有限公司  
施浩(021-34183200)

发文日:

2024 年 09 月 12 日



申请号: 202210804315.6

发文序号: 2024091201483880

申请人: 上海金融期货信息技术有限公司

发明创造名称: 融合专家推荐与文本聚类的智能事件分析方法和系统

## 第一次审查意见通知书

1. ☒ 应申请人提出的实质审查请求, 根据专利法第 35 条第 1 款的规定, 国家知识产权局对上述发明专利申请进行实质审查。

☐ 根据专利法第 35 条第 2 款的规定, 国家知识产权局决定自行对上述发明专利申请进行审查。

2. ☐ 申请人要求以其在:

☐ 申请人已经提交了经原受理机构证明的第一次提出的在先申请文件的副本。

☐ 申请人尚未提交经原受理机构证明的第一次提出的在先申请文件的副本, 根据专利法第 30 条的规定视为未要求优先权要求。

3. ☐ 经审查, 申请人于\_\_\_\_\_提交的修改文件, 不符合专利法实施细则第 57 条第 1 款的规定, 不予接受。

4. 审查针对的申请文件:

☒ 原始申请文件。 ☐ 分案申请递交日提交的文件。 ☐ 下列申请文件:

5. ☐ 本通知书是在未进行检索的情况下作出的。

☒ 本通知书是在进行了检索的情况下作出的。

☒ 本通知书引用下列对比文件(其编号在今后的审查过程中继续沿用):

编号	文件号或名称	公开日期 (或抵触申请的申请日)
1	CN114238638A	2022-03-25
2	CN113515624A	2021-10-19

6. 审查的结论性意见:

关于说明书:

☐ 申请的内容属于专利法第 5 条规定的不授予专利权的范围。

☐ 说明书不符合专利法第 26 条第 3 款的规定。

☐ 说明书不符合专利法第 33 条的规定。

☐ 说明书的撰写不符合专利法实施细则第 20 条的规定。



# 国家知识产权局

☐ \_\_\_\_\_

关于权利要求书：

- ☐ 权利要求\_\_\_\_\_不符合专利法第 2 条第 2 款的规定。
- ☐ 权利要求\_\_\_\_\_不符合专利法第 9 条第 1 款的规定。
- ☐ 权利要求\_\_\_\_\_不具备专利法第 22 条第 2 款规定的新颖性。
- ☒ 权利要求 1-12 不具备专利法第 22 条第 3 款规定的创造性。
- ☐ 权利要求\_\_\_\_\_不具备专利法第 22 条第 4 款规定的实用性。
- ☐ 权利要求\_\_\_\_\_属于专利法第 25 条规定的不授予专利权的范围。
- ☐ 权利要求\_\_\_\_\_不符合专利法第 26 条第 4 款的规定。
- ☐ 权利要求\_\_\_\_\_不符合专利法第 31 条第 1 款的规定。
- ☐ 权利要求\_\_\_\_\_不符合专利法第 33 条的规定。
- ☐ 权利要求\_\_\_\_\_不符合专利法实施细则第 22 条的规定。
- ☐ 权利要求\_\_\_\_\_不符合专利法实施细则第 23 条的规定。
- ☐ 权利要求\_\_\_\_\_不符合专利法实施细则第 24 条的规定。
- ☐ 权利要求\_\_\_\_\_不符合专利法实施细则第 25 条的规定。
- ☐ \_\_\_\_\_

- ☐ 申请不符合专利法第 26 条第 5 款或者实施细则第 29 条的规定。
- ☐ 申请不符合专利法第 19 条第 1 款的规定。
- ☐ 申请不符合专利法实施细则第 11 条的规定。
- ☐ 分案申请不符合专利法实施细则第 49 条第 1 款的规定。

上述结论性意见的具体分析见本通知书的正文部分。

7. 基于上述结论性意见，审查员认为：

- ☐ 申请人应当按照通知书正文部分提出的要求，对申请文件进行修改。
- ☐ 申请人应当在意见陈述书中论述其专利申请可以被授予专利权的理由，并对通知书正文部分中指出的不符合规定之处进行修改，否则将不能授予专利权。
- ☒ 专利申请中没有可以被授予专利权的实质性内容，如果申请人没有陈述理由或者陈述理由不充分，其申请将被驳回。

☐ \_\_\_\_\_

8. 申请人应注意下列事项：

- (1) 根据专利法第 37 条的规定，申请人应在收到本通知书之日起的 4 个月内陈述意见，如果申请人无正当理由逾期不答复，其申请被视为撤回。
- (2) 申请人对其申请的修改应当符合专利法第 33 条的规定，不得超出原说明书和权利要求书记载的范围，同时申请人对专利申请文件进行的修改应当符合专利法实施细则第 57 条第 3 款的规定，按照本通知书的要求进行修改。
- (3) 申请人的意见陈述书和/或修改文本应邮寄或递交国家知识产权局专利局受理处，凡未邮寄或递交给受理处的文件不具备法律效力。
- (4) 未经预约，申请人和/或代理师不得前来国家知识产权局专利局与审查员举行会晤。
- (5) 对进入实质审查阶段的发明专利申请，在第一次审查意见通知书答复期限届满前（已提交答复意见的除外），主动申请撤回的，可以请求退还 50% 的专利申请实质审查费。

9. 本通知书正文部分共有 3 页，并附有下列附件：

- ☐ 引用的对比文件的复印件共\_\_\_\_\_份\_\_\_\_\_页。
- ☐ \_\_\_\_\_

审查员：樊静雯

联系电话：028-62968103

审查部门：专利审查协作四川中心



210401  
2023.03

纸件申请，回函请寄：100088 北京市海淀区蓟门桥西土城路 6 号 国家知识产权局专利局受理处收  
电子申请，应当通过电子专利申请系统以电子文件形式提交相关文件。除另有规定外，以纸件等其他形式提交的文件视为未提交。



## 第一次审查意见通知书

申请号:2022108043156

本申请涉及一种融合专家推荐与文本聚类的智能事件分析方法和系统,经审查,现提出如下审查意见:

1. 权利要求 1 请求保护一种融合专家推荐与文本聚类的智能事件分析方法,对比文件 1 (CN114238638 A) 为最接近的现有技术,其公开了一种基于成对约束和集群引导的深度半监督文本聚类方法,并具体公开了如下技术特征(参见说明书第[0006]段-[0134]段): S101, 对目标文本数据进行预处理和向量化,得到多维向量;例如,当需要对某个特定领域内例如注塑领域内的文本数据进行聚类处理时,可以搜集注塑成型过程的机理经验文本描  $H(k)$ ,  $k=1,2,\dots,N3$ , 其中  $H(k)$  为不定长的句子,  $k$  为句子编号,  $N3$  为句子总数, 然后进行繁简转换、中文分词和去停用词等预处理操作, 采用词嵌入技术获取每条注塑文本对应的多维向量表示, 再进行向量标准化处理, 利用同义词替换等 EDA (Easy data augmentation) 技术得到文本增强后的注塑向量(相当于**数据预处理, 完成事件数据库中的已知事件类和已知事件的预处理, 并剔除事件中的冗余信息, 提取事件的关键字**)。S102, 根据所述多维向量学习所述目标文本数据的隐层特征, 并将所述隐层特征输入预设算法进行聚类, 得到初始聚类中心(相当于**为建立事件分类模型, 进行预聚类**); S105, 根据数据标签将全部所述有标签样本分为多个不同类型的有标签集群, 根据实时聚类结果将所述目标文本数据中的无标签样本分为多个无标签集群, 并计算全部所述有标签集群和全部所述无标签集群的集群分配损失; 所述集群分配损失用来反映两两集群间的平均相似度与其集群分配结果的差异程度(相当于**计算事件类别分类相似度阈值, 从而建立事件分类模型; 处理未分类的监控事件数据**)。S106, 根据重构误差、所述聚类损失、所述成对约束损失和所述集群分配损失计算联合损失函数, 并根据所述联合损失函数迭代达到预设条件时, 得到聚类结果。具体实施时, 在得到所述聚类损失、所述成对约束损失和所述集群分配损失后, 可以设置各个损失函数的权重, 联合上述目标函数  $L=L1+\alpha L2+\beta L3+\gamma L4$  来优化网络参数, 迭代更新直至满足一定的停止条件, 得到所述聚类结果(相当于**更新事件分类模型**)。由此可知, 对比文件 1 公开了一种融合专家推荐与文本聚类的智能事件分析方法。

该权利要求所要求保护的技术方案与对比文件 1 所公开的技术内容相比, 区别特征在于: 通过分析事件类, 获取关键词库。基于上述区别特征, 可以确定本申请相对于对比文件 1 实际解决的技术问题是: 如何优化分类效果。

然而, 对比文件 2 (CN113515624A) 公开了一种针对突发事件新闻的文本分类方法, 并具体公开了如下技术特征(参见说明书第[0007]段-[0083]段): S3、对新闻数据集 D 中的每篇新闻 d 进行事件抽取, 所对应的事件采用主谓宾结构的三元组原子事件表示, 并构造事件词典(相当于**通过分析事件类, 获取关键词库**), S4、采用无参聚类方法对事件词典 ED 中的所有事件进行聚类。上述技术特征在对比文件 2 中的作用与在该权利要求中的作用相同, 均是为了提高分类效率, 即对比文件 2 给出了将上述技术特征应用于对比文件 1 以解决其技术问题的启示。

由此可见, 在对比文件 1 的基础上, 结合对比文件 2 和本领域惯用技术手段得到该权利要求请求保护的技术方案对本领域技术人员来说是显而易见的。因此, 该权利要求不具备突出的实质性特点和显著的进步, 不具有专利法第 22 条第 3 款规定的创造性。



2. 权利要求 2 为权利要求 1 的从属权利要求，对比文件 1 公开了如下技术特征（参见说明书第[0006]段-[0134]段）：S101，对目标文本数据进行预处理和向量化，得到多维向量；例如，当需要对某个特定领域内例如注塑领域内的文本数据进行聚类处理时，可以搜集注塑成型过程的机理经验文本描  $H(k)$  , $k=1,2,...,N3$ ，其中  $H(k)$  为不定长的句子， $k$  为句子编号， $N3$  为句子总数，然后进行繁简转换、中文分词和去停用词等预处理操作，采用词嵌入技术获取每条注塑文本对应的多维向量表示，再进行向量标准化处理，利用同义词替换等 EDA (Easy data augmentation) 技术得到文本增强后的注塑向量（相当于**训练数据预处理；使用正则表达式按照固定格式识别和删除冗余信息；使用训练数据的分词结果；生成训练数据的代表性向量**）。此外，对比文件 2 公开了如下技术特征（参见说明书第[0007]段-[0083]段）：S3、对新闻数据集  $D$  中的每篇新闻  $d$  进行事件抽取，所对应的事件采用主谓宾结构的三元组原子事件表示，并构造事件词典（相当于**事件类别数据预处理；建立起关键字字典**），S4、采用无参聚类方法对事件词典  $ED$  中的所有事件进行聚类。即对比文件 2 给出了构造关键字字典并基于此进行事件聚类的启示，结合对比文件 1 的发明构思，本领域技术人员有动机构建自定义字典，向自定义字典中添加某些固定短语以构造关键字字典，用 BERT 预训练模型对关键字字典中的所有关键字进行词向量转化，获取关键字字典中的关键字对应的词向量列表；获取各事件类别的文本描述中的关键词，根据关键词词典获取各事件类别的文本描述中的关键词所对应的词向量，建立三维词向量矩阵；通过动态更新和加载自定义字典和停止词列表，根据关键字字典，将分词结果中权重前列的分词作为监控事件报警信息的关键字；代表性向量是在关键字字典中找到对应于训练数据中所有关键字的单词向量和单词权重，并使用加权平均方法生成的，这属于本领域惯用技术手段。并且，基于所建立的关键字字典，提取未分类监控事件的关键字也属于本领域惯用技术手段。因此，当其引用的权利要求不具备创造性的前提下，上述权利要求也不具有专利法第 22 条第 3 款规定的创造性。

3. 权利要求 3 为权利要求 1 的从属权利要求，对比文件 1 公开了如下技术特征（参见说明书第[0006]段-[0134]段）：S102，根据所述多维向量学习所述目标文本数据的隐层特征，并将所述隐层特征输入预设算法进行聚类，得到初始聚类中心（相当于**训练数据聚类**）。首先从  $N2$  个数据样本中随机抽取  $K$  个聚类中心，然后采用欧氏距离遍历度量所有数据点到聚类中心的距离，并将其划分到最近的类别  $\lambda_i = \arg \min_{j \in \{1,2,...,K\}} r_{ij}$  中，使得每个数据点到其所属簇心的距离总和最小，再计算每个类簇的均值作为新的聚类中心，最后 Kmeans 算法朝着  $\min E = \sum_{j=1}^K \sum_{x_i \in C_j} \|x_i - \mu_j\|^2$  的方向不断优化迭代上述步骤，直至簇心不再改变或达到足够的迭代次数，并根据最终的聚类结果得到深度嵌入聚类模型的初始聚类中心（相当于**聚类结果处理；为保证每个聚类中的所有监控事件只对应于同一个类，将聚类结果进行二次划分，将初始聚类按照打上的类标签顺次生成新的聚类；根据新的聚类结果，对每个聚类使用 K-means 算法获得聚类中心向量来表示聚类**）。此外，使用 T-SNE 方法对训练数据进行可视化，以及使用 DBSCAN 聚类算法的轮廓系数和噪声率来评估聚类效果均为本领域惯用技术手段。因此，当其引用的权利要求不具备创造性的前提下，上述权利要求也不具有专利法第 22 条第 3 款规定的创造性。

4. 权利要求 4 为权利要求 1 的从属权利要求，对比文件 1 已经公开了遍历度量所有数据点到聚类中心的距离最后得到初始聚类中心的发明构思，在此基础上，本领域技术人员有动机获取各事件类别与其对应的聚类中心和噪点数据间的相似度；采用注意力机制，计算每个聚类中心向量或噪声数据向量对应的事件类别代



表向量；计算聚类中心向量或噪声数据向量与事件类别代表向量之间的余弦相似度，这属于本领域惯用技术手段。此外，获取各事件类别的分类相似度阈值，以每个事件类别对应的最小相似度值作为阈值，生成所有事件类别的分类相似度阈值表是本领域技术人员根据实际需求进行的灵活设置，属于本领域惯用技术手段。因此，当其引用的权利要求不具备创造性的前提下，上述权利要求也不具有专利法第 22 条第 3 款规定的创造性。

5. 权利要求 5 为权利要求 1 的从属权利要求，对比文件 1 公开了如下技术特征（参见说明书第[0006]段-[0134]段）：S101，对目标文本数据进行预处理和向量化，得到多维向量（相当于**数据预处理；根据训练数据预处理的方法，对未分类的新监测事件的报警信息进行分词，提取关键字，并生成其代表向量**）；S105，根据数据标签将全部所述有标签样本分为多个不同类型的有标签集群，根据实时聚类结果将所述目标文本数据中的无标签样本分为多个无标签集群，并计算全部所述有标签集群和全部所述无标签集群的集群分配损失；所述集群分配损失用来反映两两集群间的平均相似度与其集群分配结果的差异程度（相当于**获取分类结果；计算未分类的监控事件数据的代表向量，并计算其和各个聚类中心向量之间的相似度**）。在此基础上，为了提高分类效果，本领域技术人员有动机查询已形成的事件知识库中事件关键字-类的映射关系。此外，如果新事件未被分配到任何聚类，则将其标记为噪声数据，计算噪声数据的代表向量与每个事件类别之间的余弦相似度为本领域技术人员根据实际需求进行的灵活设置，属于本领域惯用技术手段。因此，当其引用的权利要求不具备创造性的前提下，上述权利要求也不具有专利法第 22 条第 3 款规定的创造性。

6. 权利要求 6 为权利要求 1 的从属权利要求，对比文件 1 给出了对特征进行更新学习以得到聚类结果的启示，在此基础上，本领域技术人员有动机更新关键字字典，包括更新事件类别、关键字字典；更新监控事件数据库，包括更新监视事件数据、词向量矩阵、事件分类模型，这属于本领域惯用技术手段。因此，当其引用的权利要求不具备创造性的前提下，上述权利要求也不具有专利法第 22 条第 3 款规定的创造性。

7. 权利要求 7-12 请求保护一种融合专家推荐与文本聚类的智能事件分析系统，其功能模块与方法权利要求 1-6 中的步骤一一对应，参照本通知书中对权利要求 1-6 的评述可知，在权利要求 1-6 不具备创造性的情况下，上述权利要求也不具备专利法第 22 条第 3 款规定的创造性。

基于上述理由，本申请的独立权利要求以及从属权利要求都不具备创造性，同时说明书中也没有记载其他任何可以授予专利权的实质性内容，因而本申请不具备被授予专利权的前景。如果申请人不能在本通知书规定的答复期限内提出表明本申请具有创造性的充分理由，本申请将被驳回。

审查员姓名:樊静雯  
审查员代码:30141841



# 国家知识产权局

## 检索报告

申请号：2022108043156		申请日：2022 年 07 月 07 日		首次检索	
申请人：上海金融期货信息技术有限公司		最早的优先权日：			
权利要求项数：12		说明书段数：208			
审查员确定的 IPC 分类号：G06F 16/35,G06F 16/36,G06N 5/02,G06N 5/04,G06K 9/62,G06F 40/289,G06F 40/216,G06F 40/242					
检索记录信息：CN114238638A: 160 CNTXT, (词 or 字 or 文本) and (k means or DBSCAN or 聚类中心簇心 or 预聚类) and 相似度 and 分类 and 更新 and 半监督; 语义排序,语义基准:2022108043156 CN113515624A: 251 CNTXT, (文本 s 分类)and (k means or DBSCAN or 聚类中心 or 簇心 or 预聚类)(相似度 or 距离) and (字典 or 词典) and pd<20220707; 语义排序,语义基准:2022108043156 CN113378090A: 63 CNTXT, (((已知 or 已分) 4w 类) and 关键词) and (未 3W (分类 or 处理)) and 聚类分类; 语义排序,语义基准:2022108043156					
相 关 专 利 文 献					
类型	国别以及代码[11] 给出的文献号	代码[43]或[45] 给出的日期	IPC 分类号	相关的段落 和 / 或图号	涉及的权 利要求
Y	CN114238638A	2022-03-25	G06F16/35	说明书第 [0006]段 -[0134]段	1-12
Y	CN113515624A	2021-10-19	G06F16/35	说明书第 [0007]段 -[0083]段	1-12
A	CN113378090A	2021-09-10	G06F16/958	全文	1-12

相 关 非 专 利 文 献					
类型	书名（包括版本号和卷号）	出版日期	作者姓名和出版者名称	相关页数	涉及的权利要求



## 国家知识产权局

类型	期刊或文摘名称 (包括卷号和期号)	发行日期	作者姓名和文章标题	相关页数	涉及的权利要求
类型	网址	网络发布日 或公开日	作者姓名和网页标题	相关部分	涉及的权利要求

### 表格填写说明事项:

1. 审查员实际检索领域的 IPC 分类号应当填写到大组和 / 或小组所在的分类位置。
2. 期刊或其它定期出版物的名称可以使用符合一般公认的国际惯例的缩写名称。
3. 相关文件的类型说明:
  - X: 单独影响权利要求的新颖性或创造性的文件;
  - Y: 与本检索报告中其他 Y 类文件组合后影响权利要求的创造性的文件;
  - A: 背景技术文件, 即反映权利要求的部分技术特征或者有关的现有技术的文件;
  - R: 任何单位或个人在申请日向专利局提交的、属于同样的发明创造的专利或专利申请文件。
  - P: 中间文件, 其公开日在申请的申请日与所要求的优先权日之间的文件, 或者会导致需要核实该申请优先权的文件;
  - E: 单独影响权利要求新颖性的抵触申请文件;
  - T: 申请日或优先权日当天或之后公布的, 可以对所要求保护发明的理论或原理提供清楚解释的文件, 或者可显示出所要求保护发明的推理或事实不成立的文件;
  - L: 除 X、Y、A、R、P、E 和 T 类文件之外的文件。

审 查 员: 樊静雯  
2024 年 09 月 11 日

审查部门: 专利审查协作四川中心