



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:

2025 年 02 月 28 日



申请号: 202210804315.6

发文序号: 2025022800061360

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

## 驳 回 决 定

1. 根据专利法第 38 条及其实施细则第 59 条的规定, 决定驳回上述专利申请, 驳回的依据是:

- ☐ 申请不符合专利法第 2 条第 2 款的规定。
- ☐ 申请属于专利法第 5 条或者第 25 条规定的不授予专利权的范围。
- ☐ 申请不符合专利法第 9 条第 1 款的规定。
- ☐ 申请不符合专利法第 19 条第 1 款的规定。
- ☐ 申请不符合专利法第 22 条第 2 款的规定。
- ☒ 申请不符合专利法第 22 条第 3 款的规定。
- ☐ 申请不符合专利法第 22 条第 4 款的规定。
- ☐ 申请不符合专利法第 26 条第 3 款或者第 4 款的规定。
- ☐ 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 31 条第 1 款的规定。
- ☐ 申请的修改不符合专利法第 33 条的规定。
- ☐ 申请不符合专利法实施细则第 11 条的规定。
- ☐ 申请不符合专利法实施细则第 23 条第 2 款的规定。
- ☐ 分案申请不符合专利法实施细则第 49 条第 1 款的规定。
- ☐ \_\_\_\_\_

详细的驳回理由见驳回决定正文部分(共 5 页)。

2. 本驳回决定是针对下列申请文件作出的:

- ☐ 原始申请文件。 ☐ 分案申请递交日提交的文件。 ☒ 下列申请文件:

申请日提交的摘要附图、说明书摘要、说明书第 1-208 段、说明书附图; 2025 年 1 月 2 日提交的权利要求第 1-8 项。

3. 根据专利法第 41 条及实施细则第 65 条的规定, 申请人对本驳回决定不服的, 可以在收到本决定之日起 3 个月内向专利局复审和无效审理部请求复审。根据专利法实施细则第 113 条的规定, 复审费应在上述期限内缴纳, 期满未缴纳或者未缴足的, 视为未提出请求。

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





## 驳回决定

本决定涉及的是申请号为 2022108043156 的名称为“融合专家推荐与文本聚类的智能事件分析方法和系统”的发明专利申请（下称“本申请”），申请人为上海金融期货信息技术有限公司，申请日为 2022 年 07 月 07 日。

### 一、案由

本申请原申请文件权利要求书包括 2 项独立权利要求 1、7 以及 10 项从属权利要求 2-6,8-12。

应申请人于 2022 年 07 月 07 日提出的实质审查请求，审查员对本申请进行了实质审查，并于 2024 年 09 月 12 日发出了第一次审查意见通知书，指出权利要求 1-12 不具备专利法第 22 条第 3 款规定的创造性。通知书中引用了如下对比文件：

对比文件 1：CN114238638A，公开日为 2022 年 03 月 25 日；

对比文件 2：CN113515624A，公开日为 2021 年 10 月 19 日。

针对该审查意见通知书，申请人于 2025 年 01 月 02 日提交了意见陈述书和修改后的权利要求书，将从属权利要求 2-3 的附加技术特征并入原权利要求 1 中，形成新的独立权利要求 1，将从属权利要求 8-9 的附加技术特征并入原权利要求 7 中，形成新的独立权利要求 5，保留其他从属权利要求的附加技术特征，适应性修改权利要求编号，并在意见陈述书中论述了修改后的权利要求 1-8 具备创造性的理由。

审查员认为，本案事实已经清楚，因此针对申请日提交的说明书摘要、说明书第 1-208 段、摘要附图、说明书附图、2025 年 1 月 2 日提交的权利要求 1-8 项作出本驳回决定。

### 二、驳回理由

#### （一）权利要求 1-8 不具备专利法第 22 条第 3 款规定的创造性。

1. 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  $H(k)$ ， $k=1,2,...,N3$ ，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；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**）。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；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  $\min E = \sum_{i=1}^K \sum_{x \in C_i} \|x_i - \mu_i\|^2$  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**）。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；事件类别数据预处理；构建自定义字典，向自定义字典中添加某些固定短语，建立起关键字字典；基于所建立的关键字字典，提取未分类监控事件的关键字；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；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；使用 T-SNE 方法对训练数据进行可视化；使用 DBSCAN 聚类算法的轮廓系数和噪声率来评估聚类效果。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；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。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；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。



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 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。

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）：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。

4. 权利要求 4 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。

5. 权利要求 5-8 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 1-4 中的步骤一一对应，参照本通知书中对权利要求 1-4 的评述可知，在权利要求 1-4 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。

## （二）针对申请人的意见陈述进行答复

申请人认为修改后的独立权利要求 1 的技术方案与对比文件 1-2 的技术方案相比，区别如下：步骤 1：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，通过分析事件类，获取关键词库，并剔除事件中的冗余信息，提取事件的关键字。步骤 1 进一步包括：步骤 1-1：事件类别数据预处理；步骤 1-2：



训练数据预处理；其中步骤 1-1 进一步包括：步骤 1-1-1：构建自定义字典，向自定义字典中添加某些固定短语，建立起关键字字典；步骤 1-1-2：基于所建立的关键字字典，提取未分类监控事件的关键字；步骤 1-1-3：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；步骤 1-1-4：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；其中步骤 1-2 进一步包括：步骤 1-2-1：使用正则表达式按照固定格式识别和删除冗余信息；步骤 1-2-2：通过动态更新和加载自定义字典和停止词列表，使用训练数据的分词结果；步骤 1-2-3：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；步骤 1-2-4：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训练数据的代表性向量。步骤 2 进一步包括：步骤 2-1：训练数据聚类；步骤 2-2：聚类结果处理；其中，步骤 2-1 进一步包括：步骤 2-1-1：使用 T-SNE 方法对训练数据进行可视化；步骤 2-1-2：使用 DBSCAN 聚类算法的轮廓系数和噪声率来评估聚类效果；其中，步骤 2-2 进一步包括：步骤 2-2-1：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；步骤 2-2-2：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。

**申请人认为：**对比文件 1-2 并没有完全揭示上述的关键技术特征。且，上述区别技术特征并不是本领域的惯用技术手段。上述区别技术特征可以为本发明带来提高事件类别识别的准确性、提升智能事件分析方法的智能化、能够更好地适应复杂多变的实际应用场景的技术效果。

**对此，审查员认为：**

对于上述关键技术特征，参照上述评述部分，对比文件 1 以及对比文件 2 已经公开了部分技术特征。在对比文件 1 公开了数据预处理、聚类、计算有标签集群与无标签集群的集群分配损失以构建联合损失函数，最终得到聚类结果并更新网络参数的发明构思，对比文件 2 给出的构建事件词典，之后对事件词典中所有事件聚类启示，参照上述评述，对于本领域技术人员而言，其余区别技术特征为本领域惯用技术手段。本申请声称可以达到的“提高事件类别识别的准确性、提升智能事件分析方法的智能化、能够更好地适应复杂多变的实际应用场景”的技术效果也是本领域技术人员可以预期的。

综上所述，申请人的意见陈述不具备说服力。

### 三、决定

综上所述，发明专利申请不符合专利法第二十二条第三款的规定，属于专利法实施细则第五十九条第二项的情况，因此根据专利法第三十八条予以驳回。

根据专利法第四十一条第一款的规定，申请人如果对本驳回决定不服，可以在收到本驳回决定之日起三个月内，向专利局复审和无效审理部请求复审。

审查员姓名:樊静雯  
审查员代码:30141841