

200127

上海市浦东新区杨高南路288号19-21层
上海金融期货信息技术有限公司
李悦萌 女士

关于：申请号为**202210804315.6**的专利申请案
申 请 人：上海金融期货信息技术有限公司
发明名称：融合专家推荐与文本聚类的智能事件分析方法和系统
我方编号：CNJRH-0154. 224538

李悦萌 女士
您好！

关于本案的第1次审查意见通知书，已按贵方指令于期限内答复国知局，附件是答复文本的副本，请查收。

如有问题请随时联系我们。谢谢！

顺颂 业祺



专利代理师 施浩
2025年1月2日

Encls.

hdc.doc

意见陈述书

① 专 利 或 申 请 利	申请号或专利号 2022108043156
	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
	申请人或专利权人（第一署名人）上海金融期货信息技术有限公司
② 陈述事项：关于费用的意见陈述请使用意见陈述书(关于费用) 以下选项只能选择一项 <input checked="" type="checkbox"/> 针对国家知识产权局于 2024 年 09 月 12 日发出的 第 1 次审查意见通知书（发文序号 2024091201483880）陈述意见。 <input type="checkbox"/> 针对国家知识产权局于__年__月__日发出的__（发文序号__）补充陈述意见。 <input type="checkbox"/> 针对国家知识产权局于__年__月__日发出的药品专利权期限补偿审查意见通知书（发文序号__）陈述意见。 <input type="checkbox"/> 主动提出修改（根据专利法实施细则第 57 条第 1 款、第 2 款的规定） <input type="checkbox"/> 公布公告事项 <input type="checkbox"/> 其他事宜	
③ 陈述的意见： 意见陈述正文请见附件	
④ 附件清单 【附件名称】权利要求书 【附件名称】修改对照页 【附件名称】其他证明文件 已备案的证明文件备案编号：__	

意 见 陈 述 书

⑤ 当事人或专利代理机构

上海专利商标事务所有限公司

意见陈述书正文

尊敬的审查员，您好！

针对本案的审查意见，申请人的意见陈述如下。

一、针对权利要求1-12不具备创造性的审查意见

权利要求的修改：

将“其中，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

其中，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”的技术特征加入到原独立权利要求1中；

将“其中，数据预处理模块进一步配置为：

事件类别数据预处理，包括：

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；

训练数据预处理，包括：

使用正则表达式按照固定格式识别和删除冗余信息；

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；

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生成训练数据的代表性向量；

其中，预聚类模块进一步配置为：

训练数据聚类，包括：

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；

聚类结果处理，包括：

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”的技术特征加入到原独立权利要求7中。

上述修改没有超出申请文件原始记载的范围，符合专利法第33条的规定。

意见陈述理由：

申请人认为，修改后的独立权利要求1是具备创造性的，理由如下。

本案所要解决的技术问题是：现有的事件分析算法一般采用现场运维、事后处置的方法，即利用自动化工具对系统进行监控、发布部署，依靠专业人士对于事件告警信息进行分析与处理的技术来实现，其具有如下的缺陷：

1. 事件分析效率较低，占用大量计算资源，无法实时处理并发事件。
2. 事件匹配准确率低，无法快速推荐准确处置方案。
3. 运维事件专业性、针对性强，现有市面上的事件分析算法无法提供针对性的服务。

因此，如何研发一种新的事件分析技术，来解决现有事件分析算法的上述缺点，是目前业界亟待解决的问题。

为解决这一技术问题，本案的关键技术特征是：本发明的方案具有如下的创新点。

1：在读取监控事件报警信息时，使用正则表达式，识别日志中冗余的格式化信息并删除；动态识别带有连字符或下划线组成的英文词组，并将其添加至自定义字典中；在分词时使用自定义字典，能够有效防止在分词的步骤时将其分开，并且之后再之前出现过的英文词组时也会保证其不会被分开。

2：在设定关键字权重时使用IDF（逆文档频率）而没有使用常规的TF*IDF（词频*逆文档频率），突出了强区分度的词语的重要性，获得了较好的分类结果。

3：该算法中通过聚类算法进行预分类的目的是将更为相似的事件划分为一个聚类，以便在分类新监控事件时能找到与其最为相似的事件形成的聚类，根据

训练数据学习到的聚类与事件类别的对应关系，能够更准确的为其推荐分类结果。

4: 在聚类或噪点数据与类group进行相似度衡量过程中，通过使用注意力机制（Attention）机制，能够使得类名group_name的所有分词中与聚类或噪点数据有关的分词的权重更高，从而去减少那些与聚类或噪点数据无关的分词在形成group代表向量时的影响，这样在判断聚类或者噪点所属 group 的过程中，生成出来的group代表向量能够更好的与聚类或噪点数据进行相似度衡量。

5: 对所有的噪点数据进行聚类，可以让较为相似的噪点数据形成新的聚类，通过不断增加新的聚类和group的对应关系，也能够让未处理过的监控事件的分类结果更加准确。并且只对噪点数据进行聚类而不是对所有数据重新进行聚类，保持了原有聚类和事件类别对应关系的稳定性。

6: 在算法运行阶段，业务人员通过不断对新事件的分类、监控事件库更新、事件类别及其描述的更新，让模型对不断增多的带有分类标签的监控事件数据重新按照既定步骤进行学习，能够不断提升对于未处理过的监控事件的聚类效果。

7: 该算法的优势在于，根据少量的带有标签的训练数据，对于与训练数据同类型的监控事件能够迅速地建立起有效的分类模型；并且在新增少量新类型的监控事件时，能够通过聚类和相似度衡量的手段快速形成新类型监控事件的聚类与事件类别的对应关系，从而灵活地动态调整模型分类效果，避免了传统事件分类方法在面对新类型监控事件出现时，因为需要积累大量的同类型数据来优化模型分类效果，而导致重复的预测分类错误。

对比文件1的技术摘要为：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，属于数据处理技术领域，具体包括：对目标文本数据进行预处理和向量化，得到多维向量；根据多维向量学习目标文本数据的隐层特征，并将隐层特征输入预设算法进行聚类，得到初始聚类中心；计算聚类损失；利用交叉熵生成成对约束损失；并计算全部有标签集群和全部无标签集群的集群分配损失；根据重构误差、聚类损失、成对约束损失和集群分配损失计算联合损失函数，并根据联合损失函数迭代达到预设条件时，得到聚类结果。通过本公开的方案，充分挖掘标签中的监督信息，加强了其与无标签集群之间的引导学习，提高了深度半监督聚类模型的鲁棒性，同时提升了文本聚类的精度。

对比文件2的技术摘要为：一种针对突发事件新闻的文本分类方法，属于自然语言处理领域，包括：收集新闻文档，完成数据清洗，及文档的分词、依存分析和指代消解等预处理操作，得到新闻数据集D；将新闻数据集D加入到背景语料，利用Word2Vec训练后学习词的分布式表示；对新闻数据集D中的每篇新闻d进行事件抽取并构造事件词典；采用无参聚类的中文口哨方法对事件词典中的所有事件进行聚类以得到事件簇；对聚类后得到的每个事件簇，计算其出现频率和倒排文档频率，以提取特征事件；根据特征事件为每篇新闻文档构建其特征向量；采用支持向量机的分类算法完成对新闻文档的分类。该方法语义表征能力和类别区分度强。

将本发明的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。

一、本发明中，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。

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。若审查意见认为是惯用手段，应不难通过检索提出对比文件，仅在未提出任何书面证据的情况下就简单断言区别技术特征是惯用手段，并不符合审查指南第二部份第八章第4.10.2.2节的规定：「审查员在审查意见通知书中引用的本领域的公知常识应当是确凿的，如果申请人对审查员引用的公知常识提出异议，审查员应当能够说明理由或提供相应的证据予以证明」。

且，上述区别技术特征可以为本发明带来如下的技术效果：

1、提高事件类别识别的准确性：

通过步骤1-1-1构建自定义字典，并添加固定短语，可以确保在事件类别识别过程中，能够更准确地捕捉到与特定事件类别相关的关键字，减少误判和漏判。

步骤1-1-2和1-1-3利用BERT预训练模型对关键字进行词向量转化，进一步提升了关键字表示的准确性和丰富性，有助于后续的分类和聚类操作。

2、增强文本数据的预处理能力：

步骤1-2-1使用正则表达式识别和删除冗余信息，有效减少了数据噪声，提高了后续处理的效率和准确性。

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，使得分词结果更加符合实际业务场景，提高了分词的质量。

3、优化关键字提取和代表性向量生成：

步骤1-1-3和1-2-3通过关键字字典和分词结果的结合，更加精准地提取了监控事件的关键字，为事件分类提供了有力支持。

步骤1-1-4和1-2-4通过词向量和加权平均方法生成了事件类别和训练数据的代表性向量，这些向量不仅包含了丰富的语义信息，还便于后续的分类和聚类

算法进行处理。

4、提升事件分类模型的泛化能力：

通过上述预处理和特征提取步骤，使得事件分类模型在训练过程中能够学习到更加准确和丰富的特征表示，从而提高了模型的泛化能力，使其能够更好地处理未知或新出现的事件类别。

5、促进智能化和自动化处理：

整个预处理流程高度自动化，减少了人工干预的需求，提高了处理效率和准确性。同时，通过引入BERT等先进的自然语言处理技术，使得事件分析过程更加智能化。

二、本发明中，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。

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。若审查意见认为是惯用手段，应不难通过检索提出对比文件，仅在未提出任何书面证据的情况下就简单断言区别技术特征是惯用手段，并不符合审查指南第二部份第八章第4.10.2.2节的规定：「审查员在审查意见通知书中引用的本领域的公知常识应当是确凿的，如果申请人对审查员引用的公知常识提出异议，审查员应当能够说明理由或提供相应的证据予以证明」。

且，上述区别技术特征可以为本发明带来如下的技术效果：

1、提升聚类分析的准确性和可视化：

通过步骤2-1-1使用T-SNE方法对训练数据进行可视化，使得数据的分布和潜在的聚类结构更加直观。这有助于分析人员更好地理解数据，并初步判断聚类的效果。

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，提供了一种量化手段来评价聚类结果的好坏。这有助于选择最佳的聚类参数，提高聚类的准确性。

2、优化聚类结果的实用性和准确性：

步骤2-2-1的二次划分确保了每个聚类中的所有监控事件都对应于同一个类，避免了因聚类算法本身的局限性而导致的聚类混淆。这有助于提高后续事件分类的准确性和实用性。

步骤2-2-2通过K-means算法为每个聚类获得聚类中心向量，这些向量可以作为聚类的代表性特征，用于后续的事件分类和模型更新。使用聚类中心向量表示聚类不仅简化了聚类结果，还提高了聚类的可解释性和可操作性。

3、增强事件分类模型的稳定性和适应性：

通过在预聚类阶段引入更加精细和科学的聚类分析和处理，所得到的聚类结果更加准确和可靠。这有助于建立更加稳定和准确的事件分类模型，提高模型对未知事件的分类能力。

同时，由于聚类结果的优化，使得事件分类模型在更新时能够更加准确地反映数据的变化，增强了模型的适应性和鲁棒性。

4、促进智能事件分析方法的自动化和智能化：

所增加的技术特征进一步推动了智能事件分析方法的自动化进程。通过引入T-SNE可视化、DBSCAN聚类评估、K-means聚类中心提取等自动化工具和技术，减少了人工干预的需求，提高了处理效率和准确性。

此外，这些技术特征的引入也使得智能事件分析方法更加智能化，能够更好地适应复杂多变的实际应用场景。

综上，修改后的独立权利要求 1 具有突出的实质性特点和显著的进步，具备创造性。依附其的全部从属权利要求均依法具备创造性。

同理，修改后的独立权利要求 5 具有突出的实质性特点和显著的进步，具备创造性。依附其的全部从属权利要求均依法具备创造性。

以上陈述，如有不妥，请指正并给与再次修改的机会，也可通过 021-34183200-2409 和代理人联系。申请人愿意积极配合以使本申请能够早日授权。谢谢！

权 利 要 求 书

1、一种融合专家推荐与文本聚类的智能事件分析方法，其特征在于，方法包括：

5 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 2：为建立事件分类模型，进行预聚类；

步骤 3：计算事件类别分类相似度阈值，从而建立事件分类模型；

步骤 4：处理未分类的监控事件数据；

10 步骤 5：更新事件分类模型；

其中，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：

15 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；

20 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

25 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

其中, 步骤 2 进一步包括:

步骤 2-1: 训练数据聚类;

5 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:

10 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。

15 2、根据权利要求 1 所述的融合专家推荐与文本聚类的智能事件分析方法, 其特征在于, 步骤 3 进一步包括:

步骤 3-1: 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;

步骤 3-2: 获取各事件类别的分类相似度阈值;

其中步骤 3-1 进一步包括:

20 步骤 3-1-1: 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;

步骤 3-1-2: 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;

其中步骤 3-2 进一步包括:

25 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。

3、根据权利要求 1 所述的融合专家推荐与文本聚类的智能事件分析方法, 其特征在于, 步骤 4 进一步包括:

步骤 4-1: 数据预处理;

步骤 4-2: 获取分类结果;

其中步骤 4-1 中, 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;

5 其中步骤 4-2 进一步包括:

步骤 4-2-1: 查询已形成的事件知识库中事件关键字-类的映射关系;

步骤 4-2-2: 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;

步骤 4-2-3: 如果新事件未被分配到任何聚类, 则将其标记为噪声数据;

10 步骤 4-2-4: 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。

4、根据权利要求 1 所述的融合专家推荐与文本聚类的智能事件分析方法, 其特征在于, 步骤 5 进一步包括:

步骤 5-1: 更新关键字字典;

15 步骤 5-2: 更新监控事件数据库;

其中, 步骤 5-1 进一步包括:

步骤 5-1-1: 更新事件类别;

步骤 5-1-2: 更新关键字字典;

其中, 步骤 5-2 进一步包括:

20 步骤 5-2-1: 更新监视事件数据;

步骤 5-2-2: 更新词向量矩阵;

步骤 5-2-3: 更新事件分类模型。

25 5、一种融合专家推荐与文本聚类的智能事件分析系统, 其特征在于, 系统包括:

数据预处理模块, 用于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;

预聚类模块, 用于为建立事件分类模型, 进行预聚类;

事件类别分类相似度阈值计算模块, 用于计算事件类别分类相似度阈值;

未分类监控事件数据处理模块，用于处理未分类的监控事件数据；

模型更新模块，更新事件分类模型；

其中，数据预处理模块进一步配置为：

事件类别数据预处理，包括：

5 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；

10 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；

 训练数据预处理，包括：

 使用正则表达式按照固定格式识别和删除冗余信息；

15 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；

 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生成训练数据的代表性向量；

20 其中，预聚类模块进一步配置为：

 训练数据聚类，包括：

 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；

 聚类结果处理，包括：

25 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。

6、根据权利要求5所述的融合专家推荐与文本聚类的智能事件分析系统，其特征在于，事件类别分类相似度阈值计算模块进一步配置为：

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，包括：

5 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；

 获取各事件类别的分类相似度阈值，包括以每个事件类别对应的最小相似度
10 值作为阈值，生成所有事件类别的分类相似度阈值表。

7、根据权利要求5所述的融合专家推荐与文本聚类的智能事件分析系统，其特征在于，未分类监控事件数据处理模块进一步配置为：

 数据预处理，包括根据训练数据预处理的方法，对未分类的新监测事件的报
15 警信息进行分词，提取关键字，并生成其代表向量；

 获取分类结果，包括：

 查询已形成的事件知识库中事件关键字-类的映射关系；

 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；

20 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。

8、根据权利要求5所述的融合专家推荐与文本聚类的智能事件分析系统，其特征在于，模型更新模块进一步配置为：

25 更新关键字字典，包括：

 更新事件类别；

 更新关键字字典；

 更新监控事件数据库，包括：

 更新监视事件数据；

更新词向量矩阵；
更新事件分类模型。

权 利 要 求 书

1、一种融合专家推荐与文本聚类的智能事件分析方法，其特征在于，方法包括：

5 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 2：为建立事件分类模型，进行预聚类；

步骤 3：计算事件类别分类相似度阈值，从而建立事件分类模型；

步骤 4：处理未分类的监控事件数据；

10 步骤 5：更新事件分类模型；

其中，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：

15 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；

20 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；

25 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

其中, 步骤 2 进一步包括:

步骤 2-1: 训练数据聚类;

5 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:

10 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。

15 ~~2、根据权利要求 1 所述的融合专家推荐与文本聚类的智能事件分析方法, 其特征在于, 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:~~

20 ~~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;~~

25 ~~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;~~

5 ~~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。~~

~~3、根据权利要求 1 所述的融合专家推荐与文本聚类的智能事件分析方法, 其特征在于, 步骤 2 进一步包括:~~

10 ~~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;~~

15 ~~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。~~

20

~~42、根据权利要求 1 所述的融合专家推荐与文本聚类的智能事件分析方法, 其特征在于, 步骤 3 进一步包括:~~

~~步骤 3-1: 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;~~

~~步骤 3-2: 获取各事件类别的分类相似度阈值;~~

25 ~~其中步骤 3-1 进一步包括:~~

~~步骤 3-1-1: 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;~~

~~步骤 3-1-2: 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;~~

其中步骤 3-2 进一步包括：

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。

5 **53**、根据权利要求 1 所述的融合专家推荐与文本聚类的智能事件分析方法，其特征在于，步骤 4 进一步包括：

步骤 4-1：数据预处理；

步骤 4-2：获取分类结果；

10 其中步骤 4-1 中，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；

其中步骤 4-2 进一步包括：

步骤 4-2-1：查询已形成的事件知识库中事件关键字-类的映射关系；

步骤 4-2-2：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；

15 步骤 4-2-3：如果新事件未被分配到任何聚类，则将其标记为噪声数据；

步骤 4-2-4：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。

64、根据权利要求 1 所述的融合专家推荐与文本聚类的智能事件分析方法，其特征在于，步骤 5 进一步包括：

20 步骤 5-1：更新关键字字典；

步骤 5-2：更新监控事件数据库；

其中，步骤 5-1 进一步包括：

步骤 5-1-1：更新事件类别；

步骤 5-1-2：更新关键字字典；

25 其中，步骤 5-2 进一步包括：

步骤 5-2-1：更新监视事件数据；

步骤 5-2-2：更新词向量矩阵；

步骤 5-2-3：更新事件分类模型。

75、一种融合专家推荐与文本聚类的智能事件分析系统，其特征在于，系统包括：

数据预处理模块，用于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；

5 预聚类模块，用于为建立事件分类模型，进行预聚类；

事件类别分类相似度阈值计算模块，用于计算事件类别分类相似度阈值；

未分类监控事件数据处理模块，用于处理未分类的监控事件数据；

模型更新模块，更新事件分类模型；

其中，数据预处理模块进一步配置为：

10 事件类别数据预处理，包括：

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；

15 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；

训练数据预处理，包括：

使用正则表达式按照固定格式识别和删除冗余信息；

20 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；

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；

25 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；

其中，预聚类模块进一步配置为：

训练数据聚类，包括：

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；

聚类结果处理，包括：

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。

~~8、根据权利要求 7 所述的融合专家推荐与文本聚类的智能事件分析系统，其特征在于，数据预处理模块进一步配置为：—~~

~~事件类别数据预处理，包括：—~~

~~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；—~~

~~训练数据预处理，包括：—~~

~~使用正则表达式按照固定格式识别和删除冗余信息；—~~

~~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；—~~

~~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生成训练数据的代表性向量。—~~

~~9、根据权利要求 7 所述的融合专家推荐与文本聚类的智能事件分析系统，其特征在于，预聚类模块进一步配置为：—~~

~~训练数据聚类，包括：—~~

~~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；~~

~~聚类结果处理，包括：~~

~~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；~~

5 ~~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。~~

406、根据权利要求 75 所述的融合专家推荐与文本聚类的智能事件分析系统，其特征在于，事件类别分类相似度阈值计算模块进一步配置为：

10 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，包括：

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；

15 获取各事件类别的分类相似度阈值，包括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。

447、根据权利要求 75 所述的融合专家推荐与文本聚类的智能事件分析系统，其特征在于，未分类监控事件数据处理模块进一步配置为：

20 数据预处理，包括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；

获取分类结果，包括：

查询已形成的事件知识库中事件关键字-类的映射关系；

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

25 之间的相似度；

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。

128、根据权利要求 75 所述的融合专家推荐与文本聚类的智能事件分析系统，其特征在于，模型更新模块进一步配置为：

更新关键字字典，包括：

更新事件类别；

5 更新关键字字典；

更新监控事件数据库，包括：

更新监视事件数据；

更新词向量矩阵；

更新事件分类模型。

10