

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 4 月 30 日 (30.04.2020)



(10) 国际公布号
WO 2020/082613 A1

(51) 国际专利分类号:
G06F 17/27 (2006.01)

(21) 国际申请号: PCT/CN2019/070292

(22) 国际申请日: 2019 年 1 月 3 日 (03.01.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201811228490.5 2018 年 10 月 22 日 (22.10.2018) CN

(71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];
中国广东省深圳市福田区福田街道福

安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(72) 发明人: 叶曙峰(YE, Shufeng); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。 梁桂贤(LIANG, Guixian); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。 吴冶成(WU, Yecheng); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。 黄鸿顺(NG, Jon Tzen); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(54) Title: METHOD AND DEVICE FOR EXTRACTION OF CORE VIEWPOINT FROM SECURITIES RESEARCH REPORT USING DEEP LEARNING MODEL

(54) 发明名称: 利用深度学习模型的证券研报核心观点提取方法及装置

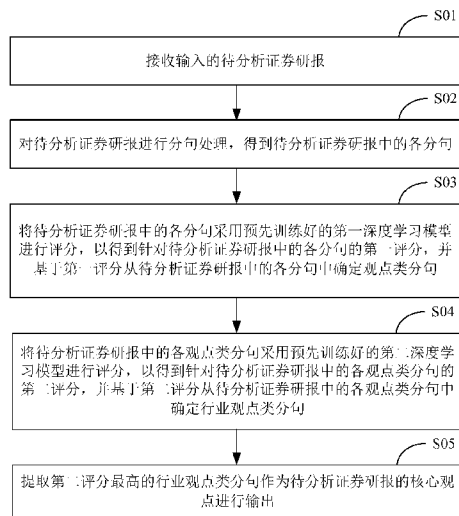


图 1

S01 Receive input securities research report to be analyzed
S02 Perform sentence segmentation processing on securities research report to be analyzed to obtain sentence segments in securities research report to be analyzed
S03 Score each sentence segment in securities research report to be analyzed using pre-trained first deep learning model to obtain first score for each sentence segment in securities research report to be analyzed, and on basis of first scores, determine viewpoint-type sentence segments from among sentence segments in securities research report to be analyzed
S04 Score each viewpoint-type sentence segment in securities research report to be analyzed using pre-trained second deep learning model to obtain second score for each viewpoint-type sentence segment in securities research report to be analyzed, and on basis of second scores, determine industry viewpoint-type sentence segments from among viewpoint-type sentence segments in securities research report to be analyzed
S05 Extract industry viewpoint-type sentence segment having highest second score as core viewpoint of securities research report to be analyzed, and output same

(57) Abstract: A method for extraction of a core viewpoint from a securities research report using a deep learning model, comprising: receiving a securities research report to be analyzed; performing sentence segmentation on said securities research report to obtain sentence segments; scoring each sentence segment using a first deep learning model to obtain a first score for each sentence segment, and determining viewpoint-type sentence segments on the basis of the first scores; scoring each viewpoint-type sentence segment using a second deep learning model to obtain a second score for each viewpoint-type sentence segment, and determining industry viewpoint-type sentence segments on the basis of the second scores; and extracting the industry viewpoint-type sentence segment having the highest second score as a core viewpoint of the securities research report to be analyzed, and outputting same. The use of deep learning models for smart scoring allows for extraction of core viewpoints from securities research reports, solving the problem of low efficiency and accuracy in securities research report analysis schemes, and efficiently and accurately extracting core viewpoints from securities research reports.

(57) 摘要: 一种利用深度学习模型的证券研报核心观点提取方法, 包括: 接收待分析证券研报; 对该证券研报进行分句得到各分句; 将各分句采用第一深度学习模型进行评分得到各分句的第一评分并基于第一评分确定观点类分句; 将各观点类分句采用第二深度学习模型进行评分得到各观点类分句的第二评分并基于第二评分确定行业观点类分句; 提取第二评分最高的行业观点类分句作为待分析证券研报的核心观点进行输出。利用深度学习模型进行智能化评分的方式实现对证券研报中核心观点的提取, 可解决针对证券研报分析方案的效率及准确率较低的问题, 可高效且准确的提取出证券研报中的核心观点。

(74) 代理人: 北京英特普罗知识产权代理有限公司 (INTELLECPRO CHINA LIMITED); 中国北京市西城区车公庄大街9号五栋大楼C座11层, Beijing 100044 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

利用深度学习模型的证券研报核心观点提取方法及装置

相关申请的交叉引用

5 本申请申明享有 2018 年 10 月 22 日递交的申请号为 CN2018112284905、名称为“利用深度学习模型的证券研报核心观点提取方法及装置”的中国专利申请的优先权，该中国专利申请的整体内容以参考的方式结合在本申请中。

技术领域

10 本申请涉及计算机技术领域，具体涉及一种利用深度学习模型的证券研报核心观点提取方法及装置。

背景技术

15 证券研究报告，也可简称为证券研报，是指相关研究人员（比如证券公司内的研究人员等）对证券及相关产品的价值、或者影响其市场价格的因素进行分析，所作出的研究报告。

对证券研报进行分析，可以及时了解证券研报中关于行业、政策、投资可行性等方面情况，但目前主要还是通过人工方式对证券研报进行阅读分析，以获得作者的核心观点等有用信息，此种方式需要耗费大量人力，效率及准确率都较低。

20 相关技术中针对证券研报分析方案的效率及准确率较低的问题，目前尚未提出有效的解决方案。

发明内容

25 本申请的目的在于提供一种利用深度学习模型的证券研报核心观点提取方法、装置、计算机设备及可读存储介质，进而在一定程度上克服上述现有技术中存在的问题，可高效且准确的提取出证券研报中的核心观点。

本申请是通过下述技术方案来解决上述技术问题：

根据本申请的一个方面，提供了一种利用深度学习模型的证券研报核心观点提取方法，包括如下步骤：

30 S01，接收输入的待分析证券研报；

S02，对待分析证券研报进行分句处理，得到待分析证券研报中的各分句；

S03，将待分析证券研报中的各分句采用预先训练好的第一深度学习模型进行评分，以得到针对待分析证券研报中的各分句的第一评分，并基于第一评分从待分析证券研报中的各分句中确定观点类分句；

35 S04，将待分析证券研报中的各观点类分句采用预先训练好的第二深度学习模型进行评分，以得到针对待分析证券研报中的各观点类分句的第二评分，并基于第二评分从待分析证券研报中的各观点类分句中确定行业观点类分句；

S05，提取第二评分最高的行业观点类分句作为待分析证券研报的核心观点进行输出。

为了实现上述目的，本申请还提供一种利用深度学习模型的证券研报核心观点提取装置，包括：

40 接收模块，用于接收输入的待分析证券研报；

分句模块，用于对待分析证券研报进行分句处理，得到待分析证券研报中的各分句；

第一评分模块，用于将待分析证券研报中的各分句采用预先训练好的第一深度学习模

型进行评分，以得到针对待分析证券研报中的各分句的第一评分，并基于第一评分从待分析证券研报中的各分句中确定观点类分句；

第二评分模块，用于将待分析证券研报中的各观点类分句采用预先训练好的第二深度学习模型进行评分，以得到针对待分析证券研报中的各观点类分句的第二评分，并基于第二评分从待分析证券研报中的各观点类分句中确定行业观点类分句；

提取模块，用于提取第二评分最高的行业观点类分句作为待分析证券研报的核心观点进行输出。

为了实现上述目的，本申请还提供一种计算机设备，包括存储器、处理器以及存储在存储器上并可在处理器上运行的计算机程序，所述处理器执行所述程序时实现利用深度学习模型的证券研报核心观点提取方法的以下步骤：

S01，接收输入的待分析证券研报；

S02，对待分析证券研报进行分句处理，得到待分析证券研报中的各分句；

S03，将待分析证券研报中的各分句采用预先训练好的第一深度学习模型进行评分，以得到针对待分析证券研报中的各分句的第一评分，并基于第一评分从待分析证券研报中的各分句中确定观点类分句；

S04，将待分析证券研报中的各观点类分句采用预先训练好的第二深度学习模型进行评分，以得到针对待分析证券研报中的各观点类分句的第二评分，并基于第二评分从待分析证券研报中的各观点类分句中确定行业观点类分句；

S05，提取第二评分最高的行业观点类分句作为待分析证券研报的核心观点进行输出。

为了实现上述目的，本申请还提供一种计算机可读存储介质，其上存储有计算机程序，所述程序被处理器执行时实现利用深度学习模型的证券研报核心观点提取方法的以下步骤：

S01，接收输入的待分析证券研报；

S02，对待分析证券研报进行分句处理，得到待分析证券研报中的各分句；

S03，将待分析证券研报中的各分句采用预先训练好的第一深度学习模型进行评分，以得到针对待分析证券研报中的各分句的第一评分，并基于第一评分从待分析证券研报中的各分句中确定观点类分句；

S04，将待分析证券研报中的各观点类分句采用预先训练好的第二深度学习模型进行评分，以得到针对待分析证券研报中的各观点类分句的第二评分，并基于第二评分从待分析证券研报中的各观点类分句中确定行业观点类分句；

S05，提取第二评分最高的行业观点类分句作为待分析证券研报的核心观点进行输出。

本申请提供的利用深度学习模型的证券研报核心观点提取方法、装置、计算机设备及可读存储介质，可先将待分析证券研报进行分句处理得到各分句，再将各分句采用预先训练好的第一深度学习模型进行评分以得到各分句的第一评分，并根据第一评分判断分句是否为观点类分句，然后将判断出的观点类分句采用预先训练好的第二深度学习模型进行评分以得到各观点类分句的第二评分，并根据第二评分判断各观点类分句是否为行业观点类分句，在判断出行业观点类分句后，可提取第二评分最高的行业观点类分句作为该待分析证券研报中的核心观点进行输出。通过上述方案，可利用预先训练好的深度学习模型通过智能化评分的方式挑选出观点类分句及行业观点类分句，并最终通过客观的评分从行业观点类分句中提取出能够代表整篇证券研报的核心观点的分句，由此，上述利用深度学习模型智能评分并分析的过程，不但可大大节省人力，而且可提高操作效率及提取结果的准确率。

附图说明

图 1 是根据本申请实施例的利用深度学习模型的证券研报核心观点提取方法的一种可

选的流程示意图；

图 2 是根据本申请实施例的利用深度学习模型的证券研报核心观点提取装置的一种可选的程序模块示意图；

图 3 是根据本申请实施例的利用深度学习模型的证券研报核心观点提取装置的另一种可选的程序模块示意图；

图 4 是根据本申请实施例的计算机设备的一种可选的硬件架构示意图。

具体实施方式

为了使本申请的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处所描述的具体实施例仅用以解释本申请，并不用于限定本申请。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

实施例 1

下面结合附图对本申请提供的利用深度学习模型的证券研报核心观点提取方法进行说明。

图 1 为本申请利用深度学习模型的证券研报核心观点提取方法的一种可选的流程示意图，如图 1 所示，该方法可以包括以下步骤：

S01，接收输入的待分析证券研报。

在本实施例中，比如可以一篇证券研报为单位进行其核心观点的提取，当接收到一篇待分析的证券研报后，可先对该篇证券研报的格式进行判断。具体的，可判断该篇证券研报的格式是否为文本类型，比如“.txt”格式、“.doc”格式等。若判断结果是否为，也就是说该证券研报的格式不是文本类型，则可将该证券研报的格式转换为文本类型，比如该证券研报的格式为便携文件格式（PDF 格式），则可使用现有的“PDFParser”等工具将 PDF 格式的证券研报转换 txt 格式的证券研报。以此，可统一接收到的待分析证券研报的格式，以更方便且更高效的对待分析证券研报的内容进行读取。

S02，对待分析证券研报进行分句处理，得到待分析证券研报中的各分句。

在本实施例中，可根据预置类型的符号对待分析证券研报进行分句处理，比如可根据逗号“，”、句号“。”、点号“.”、破折号“—”、中括号“[]、【】”、分号“；”等符号，对待分析证券研报进行分句处理，以得到待分析证券研报中的各分句。

在得到待分析证券研报中的各分句后，可进一步采用 jieba 分词模块对各分句进行分词处理，以得到待分析证券研报中的各分词。

在得到待分析证券研报中的各分词后，可基于在预置的包括分词与数值的对应转换关系的字典，确定待分析证券研报中各分词所对应的数值（在本实施例中为整数型数值）。

然后，再根据各分词与其对应数值的确定结果，将待分析证券研报中每个由多个分词组成的分句转换成数值向量格式的分句。

以此，可在后续步骤中，将数值向量格式的分句输入深度学习模型中，以便深度学习模型对各分句进行评分并可根据评分对分句的类型进行判断。

S03，将待分析证券研报中的各分句采用预先训练好的第一深度学习模型进行评分，以得到针对待分析证券研报中的各分句的第一评分，并基于第一评分从待分析证券研报中的各分句中确定观点类分句。

首先，先对第一深度学习模型的训练过程进行说明，该训练过程可包括如下步骤：

步骤 110，确定第一数据集。

比如，我们可预先精心挑选预置篇数（比如 500 篇）证券研报，比如可挑选一定比例的观点比较明显、尤其是行业观点比较明显的证券研报，还可挑选一定比例的观点比较模糊、尤其是行业观点比较模糊的证券研报等，然后可对这些证券研报的摘要部分进行分句

处理以得到多个分句，然后以人工方式对每个分句进行第一类型标签的标注，该第一类型标签可包括观点类和非观点类。也就是说，经第一类型标签标注后，在第一数据集中的分句可包括观点类分句和非观点类分句。

步骤 120，将标注有第一类型标签的各分句进行分词处理，得到标注有第一类型标签的各分词。

比如，可使用现有的 jieba 分词模块对第一数据集中的标注有第一类型标签的各分句进行分词处理，以得到标注有第一类型标签的各分词。

步骤 130，根据预置规则将标注有第一类型标签的各分词转换为对应的第一数值，并将标注有第一类型标签的各分词与其对应的第一数值存储于预置的包括分词与数值的对应转换关系的字典中。

具体的，可根据预置规则（比如随机转换，或者可根据实际需求预先设定词语到数值的转换规则等）将标注有第一类型标签的各分词一一转换为数值（在本实施例中，可称为第一数值），并可将标注有第一类型标签的各分词与第一数值存储于预置的包括分词与数值的对应转换关系的字典中。

在本实施例中，可将标注有第一类型标签的各分词与第一数值的对应转换关系以文件或者数据表的形式保存于字典中，内容可如表 1 所示。

表 1 分词与数值的对应转换关系

分词	数值
分词 1	2
分词 2	175
分词 3	50
...	...

步骤 140，根据字典，将标注有第一类型标签的各分句转换成数值向量格式的分句，形成第一数值向量分句集合。

也即，可根据上述包括分词与数值的对应转换关系的字典，将每个标注有第一类型标签的由多个分词组成的分句，转换成数值向量格式的分句，并形成数值向量分句集合，在本实施例中，可称为第一数值向量分句集合。

步骤 150，从第一数值向量分句集合中选取第一预置数量的分句作为第一训练数据。

通常在进行深度学习模型训练之前，可先确定用于训练的训练集和用于测试的测试集，在本实施例中，可从上述第一数值向量分句集合中选取预置数量（比如占比 90%）的分句作为第一训练数据，剩余（比如占比 10%）的分句则可作为第一测试数据。

步骤 160，将第一训练数据经深度学习模型进行训练，以获得第一深度学习模型。

在本实施例中，该深度学习模型可为长短期记忆网络（LSTM，Long Short-Term Memory）机器学习模型，在训练之前，可对一些关键参数进行设置，比如嵌入层大小（embedding size）、隐藏层大小（hidden layer size）、批次训练大小（batch size）、全样本循环次数（num epochs）、遗忘率（dropout）、激活函数（activation，分类型变量使用 sigmoid）、损失函数（loss，分类型问题使用 binary-crossentropy）等参数。然后，可将第一训练数据输入 LSTM 机器学习模型进行训练，以在训练完毕后获得第一深度学习模型。

此外，在训练完毕获得第一深度学习模型后，还可将第一测试数据输入第一深度学习模型进行测试，以得到该第一深度学习模型的准确率，经多次试验，我们的第一深度学习模型在第一测试数据上的准确率可达到 85% 左右。在准确率低于预定标准（比如 70%）时，可通过调整参数、对训练集数据进行调整等方式重新进行训练，以得到符合预定标准的准确率的第一深度学习模型，从而可保证利用训练好的第一深度模型可更为准确的对分句进行评分并判断出句子类型。

由此，可将待分析证券研报中的各分句采用上述第一深度学习模型进行评分以得到针对待分析证券研报中的各分句的第一评分，并可基于第一评分从待分析证券研报中的各分

句中确定观点类分句。

在本实施例中，该第一评分可为该模型认为的评分，该评分可视为概率（在 0 至 1 之间），比如模型认为比较偏向于观点类的分句，则评分更接近于 1（比如 0.75），模型认为比较偏向于非观点类的分句，则评分更接近于 0（比如 0.22）。

在得到第一评分后，可根据该第一评分与预先设置的第一分数阈值进行比较，并根据比较结果判断出观点类分句，比如可将第一评分大于第一分数阈值的分句判定为观点类分句，将第一评分不大于第一分数阈值的分句判定为非观点分句。在本实施例中，可将该第一分数阈值设置为 0.5，也就是说，第一评分不大于 0.5 的分句均可确定为非观点类分句，不再参与后续步骤；第一评分大于 0.5 的分句均可确定为观点类分句，可以继续参与后续步骤。

以此，采用预先训练好的第一深度学习模型，以智能化评分的方式来判断分句的类型（观点类分句或非观点类分句），一方面，可大大节省人力，提高判断效率；另一方面，可提高判断结果的准确率。

S04，将待分析证券研报中的各观点类分句采用预先训练好的第二深度学习模型进行评分，以得到针对待分析证券研报中的各观点类分句的第二评分，并基于第二评分从待分析证券研报中的各观点类分句中确定行业观点类分句。

首先，先对第二深度学习模型的训练过程进行说明，该训练过程可包括如下步骤：

步骤 210，确定第二数据集。

比如，预先挑选预置篇数（比如 500 篇）证券研报，比如可挑选一定比例的观点比较明显、尤其是行业观点比较明显的证券研报，还可挑选一定比例的观点比较模糊、尤其是行业观点比较模糊的证券研报等，然后可对证券研报的摘要部分进行分句处理以得到多个分句，然后以人工方式对每个分句进行第一类型标签的标注，该第一类型标签可包括观点类和非观点类。也就是说，经第一类型标签标注后，第二数据集中的分句可包括观点类分句和非观点类分句。然后，再以人工方式对每个观点类分句进行第二类型标签的标注，该第二类型标签可包括行业观点类和非行业观点类。也就是说，经第二类型标签标注后，在第二数据集中的分句可包括观点类分句和非观点类分句，且观点类分句中还包括行业观点类分句和非行业观点类分句。

步骤 220，将标注有第二类型标签的各分句进行分词处理，得到标注有第二类型标签的各分词。

也就是说，将第二数据集中标准有第二类型标签的各分句（也即第二数据集中的行业观点类分句和非行业观点类分句）进行分词处理，比如，可使用现有的 `jieba` 分词模块对上述各分句进行分词处理，以得到标注有第二类型标签的各分词。

步骤 230，根据预置规则将标注有第二类型标签的各分词转换为对应的第二数值，并将标注有第二类型标签的各分词与其对应的第二数值存储于预置的包括分词与数值的对应转换关系的字典中。

具体的，可根据预置规则（比如随机转换，或者可根据实际需求预先设定词语到数值的转换规则等）将标注有第二类型标签的各分词一一转换为数值（在本实施例中，可称为第二数值），并可将标注有第二类型标签的各分词与第二数值的转换结果存储于预置的包括分词与数值的对应转换关系的字典中。

在本实施例中，可将标注有第二类型标签的各分词与第二数值的对应转换关系以文件或者数据表的形式保存于字典中，内容可如上述表 1 所示。

步骤 240，根据字典，将标注有第二类型标签的各分句转换成数值向量格式的分句，形成第二数值向量分句集合。

也即，可根据上述包括分词与数值的对应转换关系的字典，将每个标注有第二类型标签的由多个分词组成的分句，转换成数值向量格式的分句，并形成数值向量分句集合，在本实施例中，可称为第二数值向量分句集合。

步骤 250, 从第二数值向量分句集合中选取第二预置数量的分句作为第二训练数据。

通常在进行深度学习模型训练之前, 可先确定用于训练的训练集和用于测试的测试集, 在本实施例中, 可从上述第二数值向量分句集合中选取预置数量 (比如占比 90%) 的分句作为第二训练数据, 剩余 (比如占比 10%) 的分句则可作为第二测试数据。

5 步骤 260, 将第二训练数据经深度学习模型进行训练, 以获得第二深度学习模型。

在本实施例中, 该深度学习模型可为长短期记忆网络 (LSTM, Long Short-Term Memory) 机器学习模型, 在训练之前, 可对一些关键参数进行设置, 比如嵌入层大小 (embedding size)、隐藏层大小 (hidden layer size)、批次训练大小 (batch size)、全样本循环次数 (num epochs)、遗忘率 (dropout)、激活函数 (activation, 分类型变量使用 sigmoid)、
10 损失函数 (loss, 分类型问题使用 binary-crossentropy) 等参数。然后, 可将第二训练数据输入 LSTM 机器学习模型进行训练, 以在训练完毕后获得第二深度学习模型。

此外, 在训练完毕获得第二深度学习模型后, 还可将第二测试数据输入第二深度学习模型进行测试, 以得到该第二深度学习模型的准确率。经多次试验, 我们的第二深度学习在第二测试数据上的准确率可达到 85% 左右。在准确率低于预定标准 (比如 70%) 时, 可
15 通过调整参数、对训练集数据进行调整等方式重新进行训练, 以得到符合预定标准的准确率的第二深度学习模型, 从而可保证利用训练好的第二深度模型可更为准确的对分句进行评分并判断出句子类型。

由此, 可将由 S03 中判定的待分析证券研报中的各观点类分句采用上述第二深度学习模型进行评分以得到针对待分析证券研报中的各观点类分句的第二评分, 并可基于第二评分从观点类分句中确定待分析证券研报中的行业观点类分句。
20

在本实施例中, 该第二评分可为该模型认为的评分, 该评分可视为概率 (在 0 至 1 之间), 比如模型认为比较偏向于行业观点类的分句, 则评分更接近于 1 (比如 0.85), 模型认为比较偏向于非行业观点类的分句, 则评分更接近于 0 (比如 0.3)。

在得到第二评分后, 可根据该第二评分与预先设置的第二分数阈值进行比较, 并根据
25 比较结果判断出观点类分句, 比如可将第二评分大于二分数阈值的分句判定为行业观点类分句, 第二评分不大于第二分数阈值的分句判定为非行业观点分句。在本实施例中, 可将该第二分数阈值设置为 0.5, 也就是说, 第二评分不大于 0.5 的分句均可确定为非行业观点类分句, 不再参与后续步骤; 第二评分大于 0.5 的分句均可确定为行业观点类分句, 可以继续参与后续步骤。

30 以此, 采用预先训练好的第二深度学习模型, 以智能化评分的方式来判断观点类分句的具体类型 (行业观点类分句或非行业观点类分句), 一方面, 可大大节省人力, 提高判断效率; 另一方面, 可提高判断结果的准确率。

S05, 提取第二评分最高的行业观点类分句作为待分析证券研报的核心观点进行输出。

在本实施例中, 一方面, 由于第二评分是由第二深度学习模型根据分句是偏向于行业
35 观点类还是偏向于非行业观点类而给出的分数, 且越偏向于行业观点类的分句的第二评分越高 (即越接近 1), 另一方面, 由于在通常情况下, 整篇证券研报的核心观点必然包含于行业观点中, 因此, 可将第二评分最高的行业观点类分句看作是最为重要且核心的行业观点。

在具体实现时, 可在得到第二评分后, 将第二评分进行排序 (比如以正序排序), 并将
40 第二评分最高 (比如排序第一) 的行业观点类分句作为待分析证券研报的核心观点进行输出。

根据本实施例的各个实施方式, 可先将待分析证券研报进行分句处理得到各分句, 再将各分句采用预先训练好的第一深度学习模型进行评分以得到各分句的第一评分, 并根据
45 第一评分判断分句是否为观点类分句, 然后将判断出的观点类分句采用预先训练好的第二深度学习模型进行评分以得到各观点类分句的第二评分, 并根据第二评分判断各观点类分句是否为行业观点类分句, 在判断出行业观点类分句后, 可提取第二评分最高的行业观点

类分句作为该待分析证券研报中的核心观点进行输出。通过上述方案，可利用预先训练好的深度学习模型通过智能化评分的方式挑选出观点类分句及行业观点类分句，并最终通过客观的评分从行业观点类分句中提取出能够代表整篇证券研报的核心观点的分句，由此，上述利用深度学习模型智能评分并分析的过程，不但可大大节省人力，而且可提高操作效率及提取结果的准确率。

通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到上述实施例方法可借助软件加必需的通用硬件平台的方式来实现，当然也可以通过硬件，但很多情况下前者是更佳的实施方式。

实施例 2

基于上述实施例 1 中提供的利用深度学习模型的证券研报核心观点提取方法，本实施例中提供一种利用深度学习模型的证券研报核心观点提取装置，具体地，图 2 至 3 示出了该利用深度学习模型的证券研报核心观点提取装置的可选的结构框图，该利用深度学习模型的证券研报核心观点提取装置被分割成一个或多个程序模块，一个或者多个程序模块被存储于存储介质中，并由一个或多个处理器所执行，以完成本申请。本申请所称的程序模块是指能够完成特定功能的一系列计算机程序指令段，比程序本身更适合描述利用深度学习模型的证券研报核心观点提取装置在存储介质中的执行过程，以下描述将具体介绍本实施例各程序模块的功能。

如图 2 所示，该利用深度学习模型的证券研报核心观点提取装置 20 可包括：

接收模块 21，可用于接收输入的待分析证券研报；

分句模块 22，可用于对待分析证券研报进行分句处理，得到待分析证券研报中的各分句；

第一评分模块 23，可用于将待分析证券研报中的各分句采用预先训练好的第一深度学习模型进行评分，以得到针对待分析证券研报中的各分句的第一评分，并基于第一评分从待分析证券研报中的各分句中确定观点类分句；

第二评分模块 24，可用于将待分析证券研报中的各观点类分句采用预先训练好的第二深度学习模型进行评分，以得到针对待分析证券研报中的各观点类分句的第二评分，并基于第二评分从待分析证券研报中的各观点类分句中确定行业观点类分句；

提取模块 25，可用于提取第二评分最高的行业观点类分句作为待分析证券研报的核心观点进行输出。

在具体实现时，接收模块 21，可具体用于：

判断接收到的待分析证券研报的格式是否为文本类型；

若否，则将所述待分析证券研报的格式转换为文本类型。

进一步的，参看图 3 所示，分句模块 22，可具体包括：

分句单元 221，用于根据预置类型的符号对待分析证券研报进行分句处理，得到待分析证券研报中的各分句；

分词单元 222，用于对待分析证券研报中的各分句进行分词处理，得到待分析证券研报中的各分词；

数值确定单元 223，用于基于预置的包括分词与数值的对应转换关系的字典，确定待分析证券研报中的各分词对应的数值；

转换单元 224，用于根据确定结果，将待分析证券研报中的各分句转换成数值向量格式的分句。

在本实施例中，第一深度学习模型的训练过程，包括如下步骤：

步骤 110，确定第一数据集，其中，第一数据集包括预先对预置篇数的证券研报的摘要部分进行分句处理后得到的多个分句，其中每个分句具有预先标注的第一类型标签，第一类型标签包括观点类和非观点类；

步骤 120, 将标注有第一类型标签的各分句进行分词处理, 得到标注有第一类型标签的各分词;

步骤 130, 根据预置规则将标注有第一类型标签的各分词转换为对应的第一数值, 并将标注有第一类型标签的各分词与其对应的第一数值存储于预置的包括分词与数值的对应转换关系的字典中;

步骤 140, 根据词典, 将标注有第一类型标签的各分句转换成数值向量格式的分句, 形成第一数值向量分句集合;

步骤 150, 从第一数值向量分句集合中选取第一预置数量的分句作为第一训练数据;

步骤 160, 将第一训练数据经深度学习模型进行训练, 以获得第一深度学习模型。

在本实施例中, 第二深度学习模型的训练过程, 包括如下步骤:

步骤 210, 确定第二数据集, 其中, 第二数据集包括预先对预置篇数的证券研报的摘要部分进行分句处理后得到的多个分句, 其中每个分句具有预先标注的第一类型标签, 第一类型标签包括观点类和非观点类, 且标注有观点类标签的分句同时具有预先标注的第二类型标签, 第二类型标签包括行业观点类和非行业观点类;

步骤 220, 将标注有第二类型标签的各分句进行分词处理, 得到标注有第二类型标签的各分词;

步骤 230, 根据预置规则将标注有第二类型标签的各分词转换为对应的第二数值, 并将标注有第二类型标签的各分词与其对应的第二数值存储于预置的包括分词与数值的对应转换关系的字典中;

步骤 240, 根据字典, 将标注有第二类型标签的各分句转换成数值向量格式的分句, 形成第二数值向量分句集合;

步骤 250, 从第二数值向量分句集合中选取第二预置数量的分句作为第二训练数据;

步骤 260, 将第二训练数据经深度学习模型进行训练, 以获得第二深度学习模型。

进一步的, 深度学习模型为长短期记忆网络机器学习模型。

进一步的, 采用 jieba 分词模块对各分句进行分词处理。

关于上述实施例中的装置, 其中各个单元、模块执行操作的具体方式已经在有关该方法的实施例中进行了详细描述, 此处将不做详细阐述说明。

通过本实施例的各个实施方式, 可先将待分析证券研报进行分句处理得到各分句, 再将各分句采用预先训练好的第一深度学习模型进行评分以得到各分句的第一评分, 并根据第一评分判断分句是否为观点类分句, 然后将判断出的观点类分句采用预先训练好的第二深度学习模型进行评分以得到各观点类分句的第二评分, 并根据第二评分判断各观点类分句是否为行业观点类分句, 在判断出行业观点类分句后, 可提取第二评分最高的行业观点类分句作为该待分析证券研报中的核心观点进行输出。通过上述方案, 可利用预先训练好的深度学习模型通过智能化评分的方式挑选出观点类分句及行业观点类分句, 并最终通过客观的评分从行业观点类分句中提取出能够代表整篇证券研报的核心观点的分句, 由此, 上述利用深度学习模型智能评分并分析的过程, 不但可大大节省人力, 而且可提高操作效率及提取结果的准确率。

实施例 3

本实施例还提供一种计算机设备, 如可以执行程序的智能手机、平板电脑、笔记本电脑、台式计算机、机架式服务器、刀片式服务器、塔式服务器或机柜式服务器(包括独立的服务器, 或者多个服务器所组成的服务器集群)等。如图 4 所示, 本实施例的计算机设备 40 至少包括但不限于: 可通过系统总线相互通信连接的存储器 41、处理器 42, 如图 4 所示。需要指出的是, 图 4 仅示出了具有组件 41-42 的计算机设备 40, 但是应理解的是, 并不要求实施所有示出的组件, 可以替代的实施更多或者更少的组件。

本实施例中, 存储器 41(即可读存储介质)包括闪存、硬盘、多媒体卡、卡型存储器

(例如, SD 或 DX 存储器等)、随机访问存储器 (RAM)、静态随机访问存储器 (SRAM)、只读存储器 (ROM)、电可擦除可编程只读存储器 (EEPROM)、可编程只读存储器 (PROM)、磁性存储器、磁盘、光盘等。在一些实施例中, 存储器 41 可以是计算机设备 40 的内部存储单元, 例如该计算机设备 40 的硬盘或内存。在另一些实施例中, 存储器 41 也可以是计算机设备 40 的外部存储设备, 例如该计算机设备 40 上配备的插接式硬盘, 智能存储卡 (Smart Media Card, SMC), 安全数字 (Secure Digital, SD) 卡, 闪存卡 (Flash Card) 等。当然, 存储器 41 还可以既包括计算机设备 40 的内部存储单元也包括其外部存储设备。本实施例中, 存储器 41 通常用于存储安装于计算机设备 40 的操作系统和各类应用软件, 例如实施例 2 的利用深度学习模型的证券研报核心观点提取装置的程序代码等。此外, 存储器 41 还可以用于暂时地存储已经输出或者将要输出的各类数据。

处理器 42 在一些实施例中可以是中央处理器 (Central Processing Unit, CPU)、控制器、微控制器、微处理器、或其他数据处理芯片。该处理器 42 通常用于控制计算机设备 40 的总体操作。本实施例中, 处理器 42 用于运行存储器 41 中存储的程序代码或者处理数据, 例如利用深度学习模型的证券研报核心观点提取装置等。

实施例 4

本实施例还提供一种计算机可读存储介质, 如闪存、硬盘、多媒体卡、卡型存储器 (例如, SD 或 DX 存储器等)、随机访问存储器 (RAM)、静态随机访问存储器 (SRAM)、只读存储器 (ROM)、电可擦除可编程只读存储器 (EEPROM)、可编程只读存储器 (PROM)、磁性存储器、磁盘、光盘、服务器、App 应用商城等等, 其上存储有计算机程序, 程序被处理器执行时实现相应功能。本实施例的计算机可读存储介质用于利用深度学习模型的证券研报核心观点提取装置, 被处理器执行时实现实施例 1 的利用深度学习模型的证券研报核心观点提取方法。

上述本申请实施例序号仅仅为了描述, 不代表实施例的优劣。

通过以上的实施方式的描述, 本领域的技术人员可以清楚地了解到上述实施例方法可借助软件加必需的通用硬件平台的方式来实现, 当然也可以通过硬件, 但很多情况下前者是更佳的实施方式。

以上仅为本申请的优选实施例, 并非因此限制本申请的专利范围, 凡是利用本申请说明书及附图内容所作的等效结构或等效流程变换, 或直接或间接运用在其他相关的技术领域, 均同理包括在本申请的专利保护范围内。

权利要求书

1、一种利用深度学习模型的证券研报核心观点提取方法，其特征在于，包括如下步骤：

S01，接收输入的待分析证券研报；

5 S02，对待分析证券研报进行分句处理，得到待分析证券研报中的各分句；

S03，将待分析证券研报中的各分句采用预先训练好的第一深度学习模型进行评分，以得到针对待分析证券研报中的各分句的第一评分，并基于第一评分从待分析证券研报中的各分句中确定观点类分句；

10 S04，将待分析证券研报中的各观点类分句采用预先训练好的第二深度学习模型进行评分，以得到针对待分析证券研报中的各观点类分句的第二评分，并基于第二评分从待分析证券研报中的各观点类分句中确定行业观点类分句；

S05，提取第二评分最高的行业观点类分句作为待分析证券研报的核心观点进行输出。

2、根据权利要求1所述的利用深度学习模型的证券研报核心观点提取方法，其特征在于，S01接收输入的待分析证券研报，包括：

15 判断接收到的待分析证券研报的格式是否为文本类型；

若否，则将所述待分析证券研报的格式转换为文本类型。

3、根据权利要求1所述的利用深度学习模型的证券研报核心观点提取方法，其特征在于，S02对待分析证券研报进行分句处理，得到待分析证券研报中的各分句，包括：

20 根据预置类型的符号对待分析证券研报进行分句处理，得到待分析证券研报中的各分句；

对待分析证券研报中的各分句进行分词处理，得到待分析证券研报中的各分词；

基于预置的包括分词与数值的对应转换关系的字典，确定待分析证券研报中的各分词对应的数值；

根据确定结果，将待分析证券研报中的各分句转换成数值向量格式的分句。

25 4、根据权利要求1所述的利用深度学习模型的证券研报核心观点提取方法，其特征在于，第一深度学习模型的训练过程，包括如下步骤：

步骤 110，确定第一数据集，所述第一数据集包括预先对预置篇数的证券研报的摘要部分进行分句处理后得到的多个分句，其中每个分句具有预先标注的第一类型标签，第一类型标签包括观点类和非观点类；

30 步骤 120，将标注有第一类型标签的各分句进行分词处理，得到标注有第一类型标签的各分词；

步骤 130，根据预置规则将标注有第一类型标签的各分词转换为对应的第一数值，并将标注有第一类型标签的各分词与其对应的第一数值存储于预置的包括分词与数值的对应转换关系的字典中；

35 步骤 140，根据所述词典，将标注有第一类型标签的各分句转换成数值向量格式的分句，形成第一数值向量分句集合；

步骤 150，从第一数值向量分句集合中选取第一预置数量的分句作为第一训练数据；

步骤 160，将第一训练数据经深度学习模型进行训练，以获得第一深度学习模型。

40 5、根据权利要求1所述的利用深度学习模型的证券研报核心观点提取方法，其特征在于，第二深度学习模型的训练过程，包括如下步骤：

步骤 210，确定第二数据集，所述第二数据集包括预先对预置篇数的证券研报的摘要部分进行分句处理后得到的多个分句，其中每个分句具有预先标注的第一类型标签，第一类型标签包括观点类和非观点类，且标注有观点类标签的分句同时具有预先标注的第二类型标签，第二类型标签包括行业观点类和非行业观点类；

45 步骤 220，将标注有第二类型标签的各分句进行分词处理，得到标注有第二类型标签

的各分词；

步骤 230，根据预置规则将标注有第二类型标签的各分词转换为对应的第二数值，并将标注有第二类型标签的各分词与其对应的第二数值存储于预置的包括分词与数值的对应转换关系的字典中；

5 步骤 240，根据所述字典，将标注有第二类型标签的各分句转换成数值向量格式的分句，形成第二数值向量分句集合；

步骤 250，从第二数值向量分句集合中选取第二预置数量的分句作为第二训练数据；

步骤 260，将第二训练数据经深度学习模型进行训练，以获得第二深度学习模型。

10 6、根据权利要求 4 或 5 所述的利用深度学习模型的证券研报核心观点提取方法，其特征在于，所述深度学习模型为长短期记忆网络机器学习模型。

7、根据权利要求 3 至 5 任一项所述的利用深度学习模型的证券研报核心观点提取方法，其特征在于，采用 jieba 分词模块对各分句进行分词处理。

8、一种利用深度学习模型的证券研报核心观点提取装置，其特征在于，包括：

接收模块，用于接收输入的待分析证券研报；

15 分句模块，用于对待分析证券研报进行分句处理，得到待分析证券研报中的各分句；

第一评分模块，用于将待分析证券研报中的各分句采用预先训练好的第一深度学习模型进行评分，以得到针对待分析证券研报中的各分句的第一评分，并基于第一评分从待分析证券研报中的各分句中确定观点类分句；

20 第二评分模块，用于将待分析证券研报中的各观点类分句采用预先训练好的第二深度学习模型进行评分，以得到针对待分析证券研报中的各观点类分句的第二评分，并基于第二评分从待分析证券研报中的各观点类分句中确定行业观点类分句；

提取模块，用于提取第二评分最高的行业观点类分句作为待分析证券研报的核心观点进行输出。

25 9、一种计算机设备，所述计算机设备包括存储器、处理器以及存储在存储器上并可在处理器上运行的计算机程序，其特征在于，所述处理器执行所述程序时实现利用深度学习模型的证券研报核心观点提取方法的以下步骤：

S01，接收输入的待分析证券研报；

S02，对待分析证券研报进行分句处理，得到待分析证券研报中的各分句；

30 S03，将待分析证券研报中的各分句采用预先训练好的第一深度学习模型进行评分，以得到针对待分析证券研报中的各分句的第一评分，并基于第一评分从待分析证券研报中的各分句中确定观点类分句；

S04，将待分析证券研报中的各观点类分句采用预先训练好的第二深度学习模型进行评分，以得到针对待分析证券研报中的各观点类分句的第二评分，并基于第二评分从待分析证券研报中的各观点类分句中确定行业观点类分句；

35 S05，提取第二评分最高的行业观点类分句作为待分析证券研报的核心观点进行输出。

10、根据权利要求 9 所述的计算机设备，其特征在于，S01 接收输入的待分析证券研报，包括：

判断接收到的待分析证券研报的格式是否为文本类型；

若否，则将所述待分析证券研报的格式转换为文本类型。

40 11、根据权利要求 9 所述的计算机设备，其特征在于，S02 对待分析证券研报进行分句处理，得到待分析证券研报中的各分句，包括：

根据预置类型的符号对待分析证券研报进行分句处理，得到待分析证券研报中的各分句；

对待分析证券研报中的各分句进行分词处理，得到待分析证券研报中的各分词；

45 基于预置的包括分词与数值的对应转换关系的字典，确定待分析证券研报中的各分词对应的数值；

根据确定结果，将待分析证券研报中的各分句转换成数值向量格式的分句。

12、根据权利要求 9 所述的计算机设备，其特征在于，第一深度学习模型的训练过程，包括如下步骤：

步骤 110，确定第一数据集，所述第一数据集包括预先对预置篇数的证券研报的摘要部分进行分句处理后得到的多个分句，其中每个分句具有预先标注的第一类型标签，第一类型标签包括观点类和非观点类；

步骤 120，将标注有第一类型标签的各分句进行分词处理，得到标注有第一类型标签的各分词；

步骤 130，根据预置规则将标注有第一类型标签的各分词转换为对应的第一数值，并将标注有第一类型标签的各分词与其对应的第一数值存储于预置的包括分词与数值的对应转换关系的字典中；

步骤 140，根据所述词典，将标注有第一类型标签的各分句转换成数值向量格式的分句，形成第一数值向量分句集合；

步骤 150，从第一数值向量分句集合中选取第一预置数量的分句作为第一训练数据；

步骤 160，将第一训练数据经深度学习模型进行训练，以获得第一深度学习模型。

13、根据权利要求 9 所述的计算机设备，其特征在于，第二深度学习模型的训练过程，包括如下步骤：

步骤 210，确定第二数据集，所述第二数据集包括预先对预置篇数的证券研报的摘要部分进行分句处理后得到的多个分句，其中每个分句具有预先标注的第一类型标签，第一类型标签包括观点类和非观点类，且标注有观点类标签的分句同时具有预先标注的第二类型标签，第二类型标签包括行业观点类和非行业观点类；

步骤 220，将标注有第二类型标签的各分句进行分词处理，得到标注有第二类型标签的各分词；

步骤 230，根据预置规则将标注有第二类型标签的各分词转换为对应的第二数值，并将标注有第二类型标签的各分词与其对应的第二数值存储于预置的包括分词与数值的对应转换关系的字典中；

步骤 240，根据所述字典，将标注有第二类型标签的各分句转换成数值向量格式的分句，形成第二数值向量分句集合；

步骤 250，从第二数值向量分句集合中选取第二预置数量的分句作为第二训练数据；

步骤 260，将第二训练数据经深度学习模型进行训练，以获得第二深度学习模型。

14、根据权利要求 12 或 13 所述的计算机设备，其特征在于，所述深度学习模型为长短期记忆网络机器学习模型。

15、一种计算机可读存储介质，其上存储有计算机程序，其特征在于：所述程序被处理器执行时实现利用深度学习模型的证券研报核心观点提取方法的以下步骤：

S01，接收输入的待分析证券研报；

S02，对待分析证券研报进行分句处理，得到待分析证券研报中的各分句；

S03，将待分析证券研报中的各分句采用预先训练好的第一深度学习模型进行评分，以得到针对待分析证券研报中的各分句的第一评分，并基于第一评分从待分析证券研报中的各分句中确定观点类分句；

S04，将待分析证券研报中的各观点类分句采用预先训练好的第二深度学习模型进行评分，以得到针对待分析证券研报中的各观点类分句的第二评分，并基于第二评分从待分析证券研报中的各观点类分句中确定行业观点类分句；

S05，提取第二评分最高的行业观点类分句作为待分析证券研报的核心观点进行输出。

16、根据权利要求 15 所述的计算机可读存储介质，其特征在于，S01 接收输入的待分析证券研报，包括：

判断接收到的待分析证券研报的格式是否为文本类型；

若否，则将所述待分析证券研报的格式转换为文本类型。

17、根据权利要求 15 所述的计算机可读存储介质，其特征在于，S02 对待分析证券研报进行分句处理，得到待分析证券研报中的各分句，包括：

根据预置类型的符号对待分析证券研报进行分句处理，得到待分析证券研报中的各分句；

对待分析证券研报中的各分句进行分词处理，得到待分析证券研报中的各分词；

基于预置的包括分词与数值的对应转换关系的字典，确定待分析证券研报中的各分词对应的数值；

根据确定结果，将待分析证券研报中的各分句转换成数值向量格式的分句。

18、根据权利要求 15 所述的计算机可读存储介质，其特征在于，第一深度学习模型的训练过程，包括如下步骤：

步骤 110，确定第一数据集，所述第一数据集包括预先对预置篇数的证券研报的摘要部分进行分句处理后得到的多个分句，其中每个分句具有预先标注的第一类型标签，第一类型标签包括观点类和非观点类；

步骤 120，将标注有第一类型标签的各分句进行分词处理，得到标注有第一类型标签的各分词；

步骤 130，根据预置规则将标注有第一类型标签的各分词转换为对应的第一数值，并将标注有第一类型标签的各分词与其对应的第一数值存储于预置的包括分词与数值的对应转换关系的字典中；

步骤 140，根据所述词典，将标注有第一类型标签的各分句转换成数值向量格式的分句，形成第一数值向量分句集合；

步骤 150，从第一数值向量分句集合中选取第一预置数量的分句作为第一训练数据；

步骤 160，将第一训练数据经深度学习模型进行训练，以获得第一深度学习模型。

19、根据权利要求 15 所述的计算机可读存储介质，其特征在于，第二深度学习模型的训练过程，包括如下步骤：

步骤 210，确定第二数据集，所述第二数据集包括预先对预置篇数的证券研报的摘要部分进行分句处理后得到的多个分句，其中每个分句具有预先标注的第一类型标签，第一类型标签包括观点类和非观点类，且标注有观点类标签的分句同时具有预先标注的第二类型标签，第二类型标签包括行业观点类和非行业观点类；

步骤 220，将标注有第二类型标签的各分句进行分词处理，得到标注有第二类型标签的各分词；

步骤 230，根据预置规则将标注有第二类型标签的各分词转换为对应的第二数值，并将标注有第二类型标签的各分词与其对应的第二数值存储于预置的包括分词与数值的对应转换关系的字典中；

步骤 240，根据所述字典，将标注有第二类型标签的各分句转换成数值向量格式的分句，形成第二数值向量分句集合；

步骤 250，从第二数值向量分句集合中选取第二预置数量的分句作为第二训练数据；

步骤 260，将第二训练数据经深度学习模型进行训练，以获得第二深度学习模型。

20、根据权利要求 18 或 19 所述的计算机可读存储介质，其特征在于，所述深度学习模型为长短期记忆网络机器学习模型。

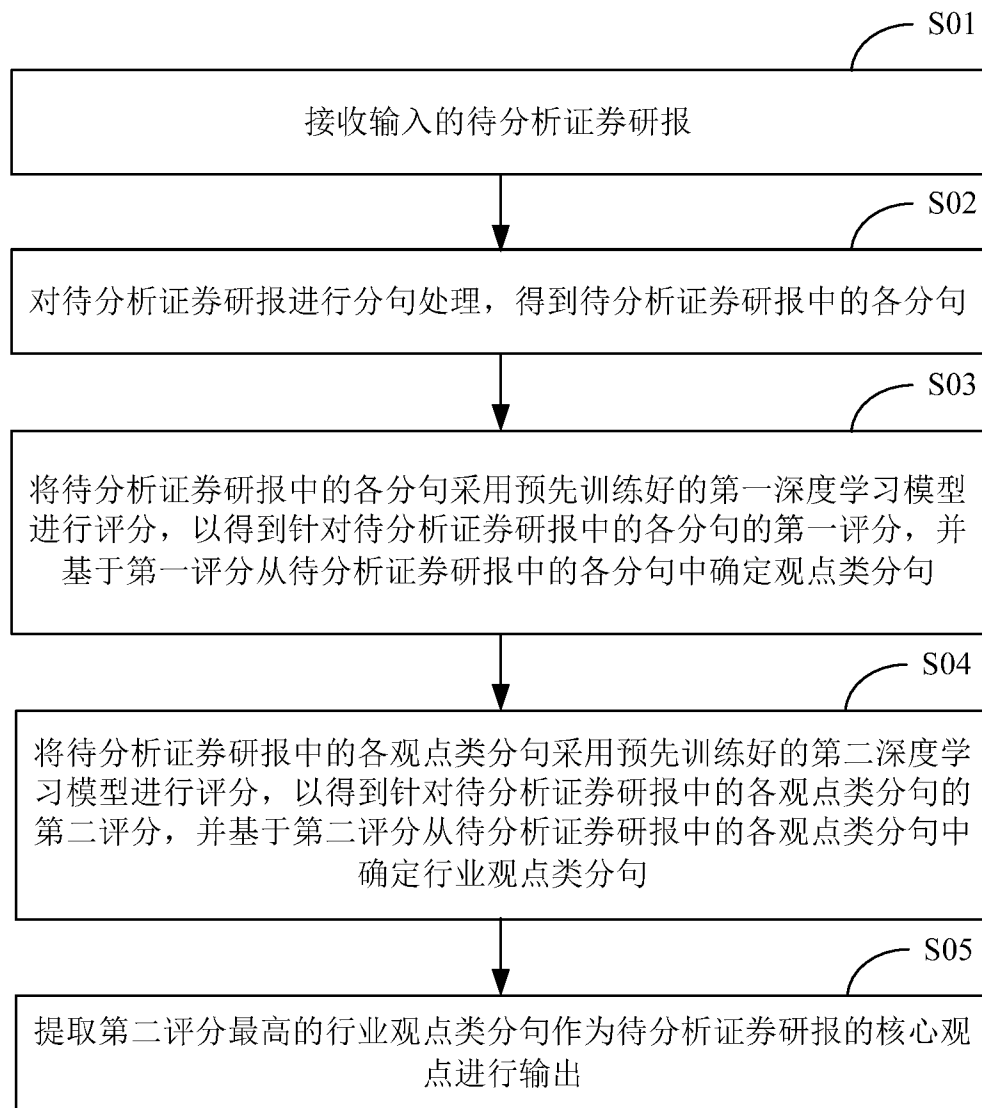


图 1

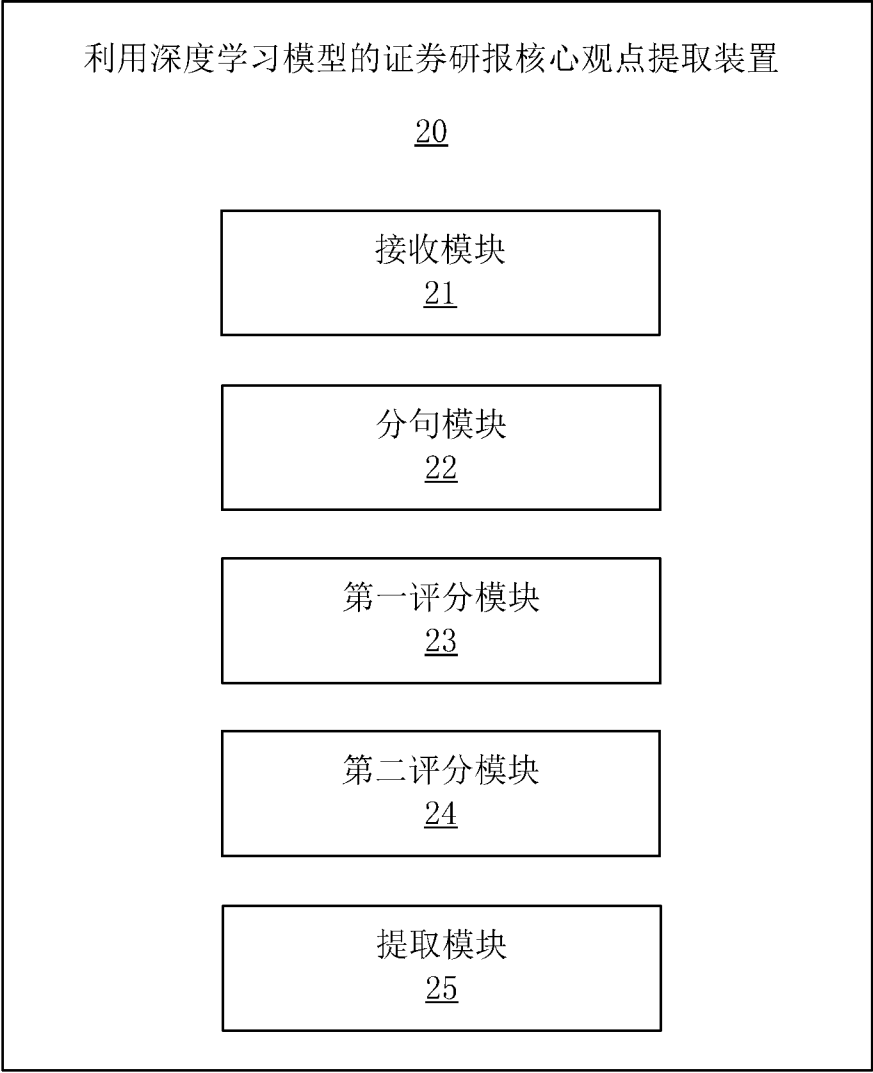


图 2

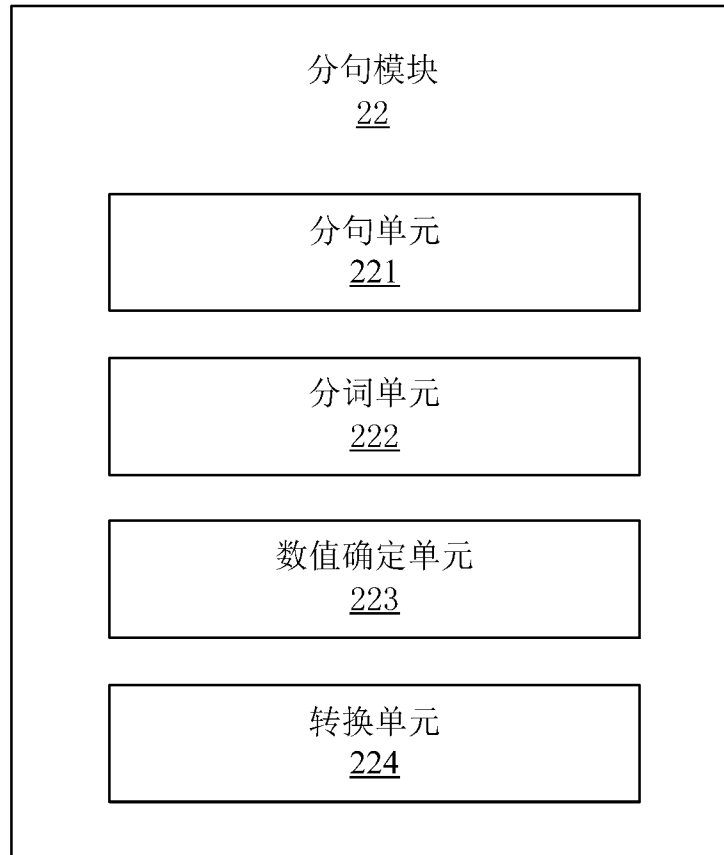


图 3

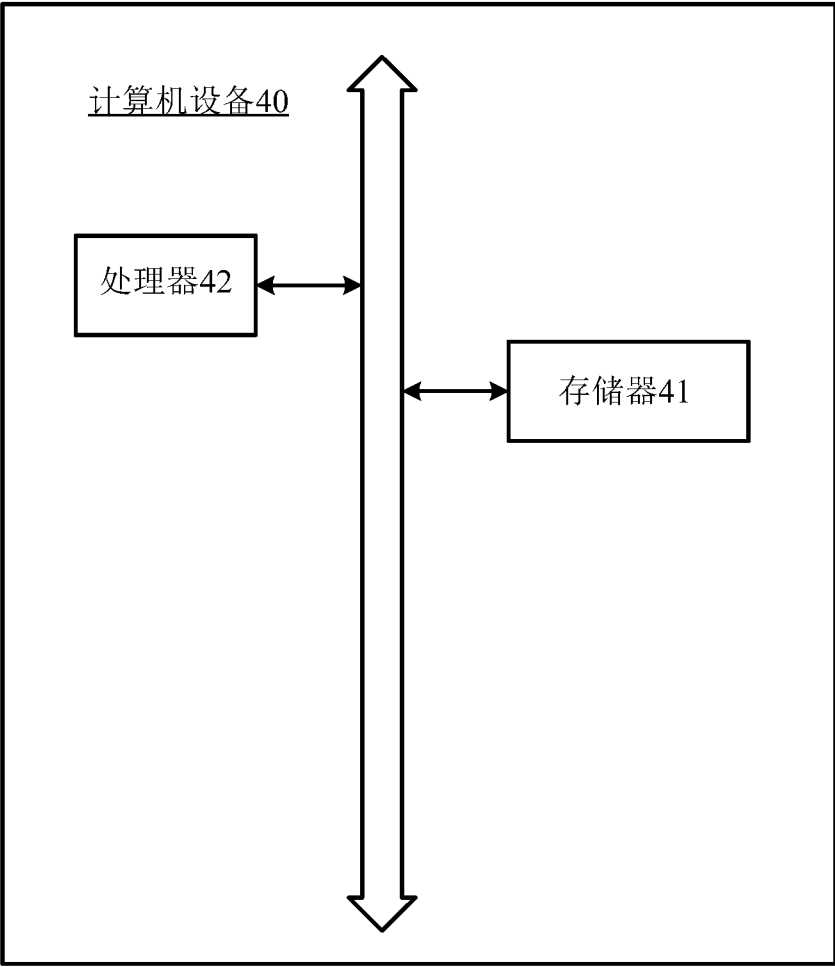


图 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/070292

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/27(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WPI; EPODOC; CNKI; CNPAT: 证券, 观点, 分句, 分类, 深度学习, 机器学习, view, stock, detach

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 106776551 A (GUILIN UNIVERSITY OF ELECTRONIC TECHNOLOGY) 31 May 2017 (2017-05-31) description, paragraphs [0003]-[0086]	1-20
A	CN 108491406 A (SHENZHEN AXMTEC CO., LTD.) 04 September 2018 (2018-09-04) entire document	1-20
A	US 2016283583 A1 (TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED) 29 September 2016 (2016-09-29) entire document	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

05 July 2019

Date of mailing of the international search report

22 July 2019

Name and mailing address of the ISA/CN

China National Intellectual Property Administration
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/070292

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	106776551	A	31 May 2017	None			
CN	108491406	A	04 September 2018	None			
US	2016283583	A1	29 September 2016	CN	104915327	A	16 September 2015
				WO	2015135452	A1	17 September 2015

国际检索报告

国际申请号

PCT/CN2019/070292

A. 主题的分类

G06F 17/27 (2006.01) i

按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献 (标明分类系统和分类号)

G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用))

WPI; EPDOC; CNKI; CNPAT: 证券, 观点, 分句, 分类, 深度学习, 机器学习, view, stock, detach

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	CN 106776551 A (桂林电子科技大学) 2017年 5月 31日 (2017 - 05 - 31) 说明书第[0003]-[0086]段	1-20
A	CN 108491406 A (深圳市阿西莫夫科技有限公司) 2018年 9月 4日 (2018 - 09 - 04) 全文	1-20
A	US 2016283583 A1 (TENCENT TECHNOLOGY SHENZHEN COMPANY LIMITED) 2016年 9月 29日 (2016 - 09 - 29) 全文	1-20

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2019年 7月 5日

国际检索报告邮寄日期

2019年 7月 22日

ISA/CN的名称和邮寄地址

中国国家知识产权局 (ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

丛磊

电话号码 (86-10)-53961305

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/070292

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	106776551	A	2017年 5月 31日	无	
CN	108491406	A	2018年 9月 4日	无	
US	2016283583	A1	2016年 9月 29日	CN 104915327 A	2015年 9月 16日
				WO 2015135452 A1	2015年 9月 17日