

(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



(10) 国际公布号

WO 2020/082612 A1

(43) 国际公布日

2020 年 4 月 30 日 (30.04.2020)

(51) 国际专利分类号:

G06F 17/27 (2006.01)

(21) 国际申请号:

PCT/CN2019/070291

(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:

201811228240.1 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)。 蒋逸文(JIANG, Yiwen); 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)。 孙葛亮(SUN, Geliang); 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)。 刘琼(LIU, Qiong); 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)。

(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)。

(54) Title: METHOD AND APPARATUS FOR SENTIMENT ANALYSIS ON SECURITY RESEARCH REPORT USING BIG DATA, AND COMPUTER DEVICE

(54) 发明名称: 利用大数据的证券研报情感分析方法、装置及计算机设备

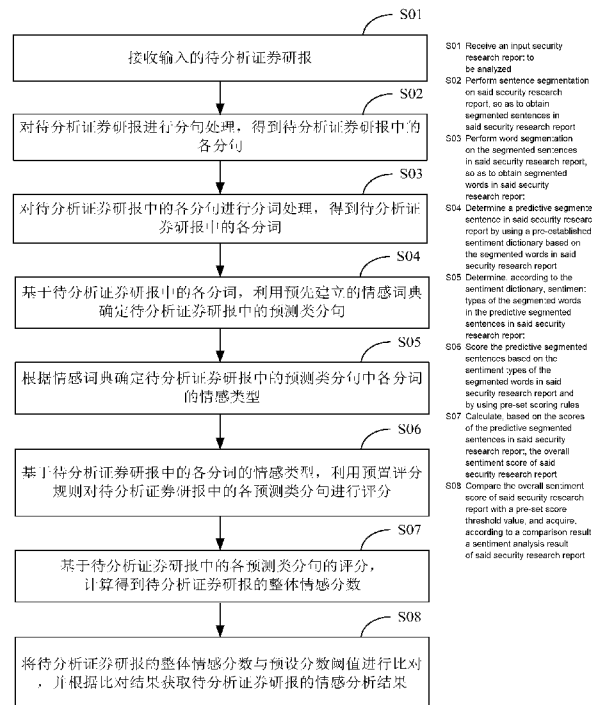


图 1

(57) Abstract: A method for sentiment analysis on a security research report using big data. The method comprises: receiving an input security research report to be analyzed (S01); performing sentence segmentation on said security research report, so as to obtain segmented sentences (S02); performing word segmentation on the segmented sentences to obtain segmented words (S03); determining a predictive segmented sentence by using a sentiment dictionary based on the segmented words (S04); determining sentiment types of the segmented words in the predictive segmented sentence according to the sentiment dictionary (S05); scoring the predictive segmented

(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))。

sentences based on the sentiment types of the segmented words and by using pre-set scoring rules (S06); calculating, based on scoring calculation, the overall sentiment score of said security research report (S07); and according to a result of comparison between the overall sentiment score of said security research report and a pre-set score threshold value, acquiring a sentiment analysis result of said security research report (S08). The sentiment analysis on a security research report is implemented using a big data analysis and intelligent scoring method, which can solve the problem of low efficiency and accuracy of a sentiment analysis solution for a security research report in the prior art, and improve the efficiency and accuracy of sentiment analysis on the security research report.

(57) 摘要: 一种利用大数据的证券研报情感分析方法, 包括: 接收输入的待分析证券研报(S01); 对待分析证券研报进行分句以得到各分句(S02); 对各分句进行分词以得到各分词(S03); 基于各分词利用情感词典确定预测类分句(S04); 根据情感词典确定预测类分句中各分词的情感类型(S05); 基于各分词的情感类型并利用预置评分规则对各预测类分句进行评分(S06); 基于评分计算得到待分析证券研报的整体情感分数(S07); 根据待分析证券研报的整体情感分数与预设分数阈值的比对结果获取待分析证券研报的情感分析结果(S08)。利用大数据分析及智能化评分的方式实现对证券研报的情感分析, 可解决现有技术针对证券研报情感分析方案的效率及准确率较低的问题, 提高证券研报情感分析的效率及准确率。

利用大数据的证券研报情感分析方法、装置及计算机设备

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 CN2018112282401、名称为“利用大数据的证券研报情感分析方法、装置及计算机设备”的中国专利申请的优先权，该中国专利申请的整

技术领域

10 本申请涉及计算机技术领域，具体涉及一种利用大数据的证券研报情感分析方法、装置及计算机设备。

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。以针对证券研报的情感分析为例，目前主要还是通过人工方式对证券研报进行阅读，以分析出作者的情感趋势，但此种方式需要耗费大量人力，且效率及准确率都较低。此外，也有通过情感词典进行分析的方案，比如利用现有可用来判断积极和消极情感方面的词典资源，如知网汉语词典 Hownet、台湾大学简体中文情感极性词典 NTUSD 等，但是，此种方式大多是针对短句进行分析，相对割裂了语句的前后关系，很难对整篇文章做出较为完善准确的分析，尤其在针对具有因果、转折类等关联关系的大量语句时，分析的准确率更不理想。

25 相关技术中针对证券研报情感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

30 本申请的目的在于提供一种利用大数据的证券研报情感分析方法、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，可提高对证券研报情感分析的效率及准确率。

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，提供了一种利用大数据的证券研报情感分析方法，包括如下步骤：

- 35 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，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；
- S04，基于待分析证券研报中的各分词，利用预先建立的情感词典确定待分析证券研报中的预测类分句；
- S05，根据所述情感词典确定待分析证券研报中的预测类分句中各分词的情感类型；
- 40 S06，基于待分析证券研报中的各分词的情感类型，利用预置评分规则对待分析证券研报中的各预测类分句进行评分；
- S07，基于待分析证券研报中的各预测类分句的评分，计算得到待分析证券研报的整体

情感分数；

S08，将待分析证券研报的整体情感分数与预设分数阈值进行比对，并根据比对结果获取待分析证券研报的情感分析结果。

为了实现上述目的，本申请还提供一种利用大数据的证券研报情感分析装置，包括：

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；

分词模块，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；

预测类分句确定模块，用于基于待分析证券研报中的各分词，利用预先建立的情感词典确定待分析证券研报中的预测类分句；

情感类型确定模块，用于根据所述情感词典确定待分析证券研报中的预测类分句中各分词的情感类型；

评分模块，用于基于待分析证券研报中的预测类分句中各分词的情感类型，利用预置评分规则对待分析证券研报中的各预测类分句进行评分；

计算模块，用于基于待分析证券研报中的各预测类分句的评分，计算得到待分析证券研报的整体情感分数；

情感分析结果获取模块，用于将待分析证券研报的整体情感分数与预设分数阈值进行比对，并根据比对结果获取待分析证券研报的情感分析结果。

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大数据的证券研报情感分析方法的以下步骤：

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，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；

S04，基于待分析证券研报中的各分词，利用预先建立的情感词典确定待分析证券研报中的预测类分句；

S05，根据所述情感词典确定待分析证券研报中的预测类分句中各分词的情感类型；

S06，基于待分析证券研报中的各分词的情感类型，利用预置评分规则对待分析证券研报中的各预测类分句进行评分；

S07，基于待分析证券研报中的各预测类分句的评分，计算得到待分析证券研报的整体情感分数；

S08，将待分析证券研报的整体情感分数与预设分数阈值进行比对，并根据比对结果获取待分析证券研报的情感分析结果。

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大数据的证券研报情感分析方法的以下步骤：

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，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；

S04，基于待分析证券研报中的各分词，利用预先建立的情感词典确定待分析证券研报中的预测类分句；

S05，根据所述情感词典确定待分析证券研报中的预测类分句中各分词的情感类型；

S06，基于待分析证券研报中的各分词的情感类型，利用预置评分规则对待分析证券研报中的各预测类分句进行评分；

S07，基于待分析证券研报中的各预测类分句的评分，计算得到待分析证券研报的整体情感分数；

S08，将待分析证券研报的整体情感分数与预设分数阈值进行比对，并根据比对结果获

取待分析证券研报的情感分析结果。

本申请提供的利用大数据的证券研报情感分析方法、装置、计算机设备及可读存储介质，可预先对大量证券研报中的分词进行提取以生成包括所需情感类型词库的情感词典，基于此，可先将待分析证券研报进行分句以及分词处理，然后将得到的各分词与情感词典匹配的方式先确定出预测类分句再确定出预测类分句中各分词的情感类型，接下来基于预测类分句中各分词的情感类型并利用预置评分规则对各预测类分句进行评分，并可基于评分结果计算得到待分析证券研报的整体情感分数，最后根据待分析证券研报的整体情感分数与预置分数阈值的比对结果获得待分析证券研报的情感分析结果。通过本方案，可基于大数据分析及智能化评分的方式实现对待分析证券研报的情感分析，不但可大大节省人力，而且可提高分析效率及分析结果的准确率。

附图说明

图 1 是根据本申请实施例的利用大数据的证券研报情感分析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大数据的证券研报情感分析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大数据的证券研报情感分析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大数据的证券研报情感分析方法进行说明。

图 1 为本申请利用大数据的证券研报情感分析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，比如可以一篇待分析证券研报为单位进行情感分析，当接收到一篇待分析证券研报后，可先对该篇待分析证券研报的格式进行判断。具体的，可判断该待分析证券研报的格式是否为文本类型，比如“.txt”格式、“.doc”格式等。若判断结果为否，也就是说该待分析证券研报的格式不是文本类型，则可将该待分析证券研报的格式转换为文本类型，比如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。

S03，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。

在具体实现时，在得到待分析证券研报中的各分句后，可进一步采用现有技术中的分词模块，比如采用 jieba 分词模块对上述各分句进行分词处理，以得到待分析证券研报中的

各分词。

S04, 基于待分析证券研报中的各分词, 利用预先建立的情感词典确定待分析证券研报中的预测类分句。

首先, 可先对情感词典的建立过程进行说明, 该过程可包括如下步骤:

步骤 110, 对预置篇数的证券研报进行分句处理, 得到证券研报中的各分句 (在本实施例中, 可称为第一分句)。

预先选取预置篇数 (比如 300 篇等) 涉及不同行业的证券研报, 然后对各篇证券研报进行分句处理, 以得到所有证券研报中的各第一分句。

步骤 120, 对各第一分句进行分词处理, 得到证券研报中的各第一分词。

在得到各第一分句后, 可利用现有分词模块, 比如采用 jieba 分词模块对各第一分句进行分词处理, 以得到所有证券研报中的各分词 (在本实施例中, 可称为第一分词)。

步骤 130, 从所有第一分词中提取用于表达正面情感的分词、用于表达负面情感的分词、用于表达否定情感的分词及用于表达预测的分词。

在得到所有证券研报中的各第一分词后, 我们可利用人工提取与机器提取相结合的方式, 从各第一分词中提取用于表达正面情感的分词 (主要包括对市场和行业看多的词语, 比如“领先大市”、“价值低估”等)、用于表达负面情感的分词 (主要包括对市场和行业看空的词语, 比如“经济衰退”、“差强人意”等)、用于表达否定情感的分词 (主要指可对句子含义起到反转作用的词语, 比如“极难”、“无理由”等)、用于表达预测的分词 (主要指与作者的预判、建议等相关的词语, 主要用于后续定位预测类分句, 比如“建议”、“预计”、“预测”等), 此外, 在实际应用中, 还可以进一步提取与行业相关的分词 (主要用于定位行业, 比如“有色金属”、“化工行业”等)。

当然, 在提取完成后, 还可以加入人工检验的步骤, 对提取结果进行校验, 以得到更为准确的提取结果, 从而为后续建立各相关词库做充分的数据准备。

步骤 140, 根据提取结果分别生成正面情感词库、负面情感词库、否定情感词库及预测类词库。

也就是说, 可根据提取的用于表达正面情感的分词生成正面情感词库, 根据提取的用于表达负面情感的分词生成负面情感词库, 根据提取的用于表达否定情感的分词生成否定情感词库, 根据提取的用于表达预测的分词生成预测类库, 根据提取的行业相关的分词生成行业相类词库。

步骤 150, 基于正面情感词库、负面情感词库、否定情感词库及预测类词库建立情感词典。

将生成的正面情感词库、负面情感词库、否定情感词库、预测类词库、行业相类词库等整合起来以建立情感词典, 以便后续在对证券研报进行情感分析时, 可根据情感词典中的各词库快速且准确的确定分词类型、分句类型等, 从而实现对语句、文章等做出更为准确的分析 (尤其是情感分析)。

在本实施例中, 在 S03 得到待分析证券研报中的各分词后, 可将待分析证券研报中的各分词与情感词典中的预测类词库中的分词进行匹配, 若待分析证券研报中的任一分词能够与情感词典中预测类词库中的分词相匹配, 则可将该任一分词确定为预测类分词, 进而可将该任一分词所属的分句确定为预测类分句。

由于在证券研报中, 预测类语句 (也可理解为本方案中的预测类分句) 出现的地方, 通常对应为作者用来表达对未来市场行情发展的研判 (比如可包括看涨、看空等情感趋势), 因此, 先确定出预测类分句, 再对预测类分句中的分词进行情感分析, 可准确且有效的分析出作者的情感趋势。

S05, 根据情感词典确定待分析证券研报中的预测类分句中各分词的情感类型。

在具体实现时, 在确定了各预测类分句后, 可将预测类分句中包含的各分词分别与情感词典中正面情感词库、负面情感词库、否定情感词库中的分词进行匹配, 若任一分词能

够与正面情感词库中的分词相匹配，或者能够与负面情感词库中的分词相匹配，或者能够与否定情感词库中的分词相匹配，则可将该任一分词的情感类型分别确定为正面分词、负面分词、否定分词。

S06，基于待分析证券研报中的预测类分句中各分词的情感类型，利用预置评分规则对待分析证券研报中的各预测类分句进行评分。

在确定了预测类分句中各分词的情感类型后，则可基于各分词的情感类型确定其所属各分句的情感趋势，在本实施例中，可以智能化评分的方式对各预测类分句的情感趋势进行评价，并可将所得评分应用于后续步骤以计算得到证券研报的整体情感分数。

在具体实现时，可先统计待分析证券研报中每个预测类分句中的正面分词、负面分词的初始数量（由于该数量可能会被调整，因此，在此处称为初始数量），和否定分词的数量。

以一个预测类分句为例，比如可在统计出该分句的正面分词、负面分词的初始数量（比如分别为 3 和 2）后，判断否定分词的数量是否为奇数。

如果否定分词的数量为奇数（比如，否定分词的数量为 1 个，为“很难”），则可进一步判断该奇数个否定分词前面相邻的或后面相邻的是正面分词还是负面分词。

如果该奇数个否定分词前面相邻的或后面相邻的是正面分词（比如“领先大市”），则将该否定分词与该正面分词相结合表达的是负面的意思，此时，需要将该正面分词的初始数量减一（即 $3-1=2$ ）并将该负面分词的初始数量加一（即 $2+1=3$ ），以分别作为调整后的正面分词数量及负面分词数量。

如果该奇数个否定分词前面相邻的或后面相邻的是负面分词（比如“下跌”），则该否定分词与该负面分词相结合表达的是正面的意思，此时，需要将该负面分词的初始数量减一（即 $2-1=1$ ）并将该正面分词的初始数量加一（即 $3+1=4$ ），以分别作为调整后的负面分词数量及正面分词数量。

以上述奇数个否定分词后面相邻的是负面分词的情况为例，在确定了调整后的正面分词数量及负面分词数量后，可将该预测类分句中正面分词数量与负面分词数量的差（即 $4-1=3$ ），除以正面分词数量与负面分词数量的和（即 $4+1=5$ ），以得到对该预测类分句的评分（即 $3/5=0.6$ ）。

通常在证券研报中，预测类分句的情感趋势与证券研报整体的情感趋势是息息相关的，因此，可基于各预测类分句的评分在后续步骤中得到证券研报的整体情感趋势。

S07，基于待分析证券研报中的各预测类分句的评分，计算得到待分析证券研报的整体情感分数。

在本实施例中，待分析证券研报中的各预测类分句的评分包括正分和负分，分别对应正面预测类分句和负面预测类分句。

在 S06 计算得到待分析证券研报中的各预测类分句的评分后，可以对该评分进行赋权累加计算。

具体的，可分别计算待分析证券研报中的所有正面预测类分句的平均分，以及所有负面预测类分句的平均分，由于在实际情况下，绝大多数证券研报都是偏向正面的观点，因此，在本方案中，我们对负面预测类分句赋予更高的权重，比如可设置为正面预测类分句的二倍。

具体计算时，可将所有正面预测类分句的平均分与所有负面预测类分句的平均分的二倍进行求和计算，以得到该待分析证券研报的整体情感分数，具体可通过如下公式进行计算：

整体情感分数=所有正面预测类分句的平均分+（所有负面预测类分句的平均分 $\times$ 2）。

S08，将待分析证券研报的整体情感分数与预设分数阈值进行比对，并根据比对结果获取待分析证券研报的情感分析结果。

理论上，该预设分数阈值可设置为 0，但考虑到在实际情况下，大多数证券研报都是偏向正面的观点，因此，在本实施例中，该预设分数阈值可设置为包括第一预设分数阈值

和第二预设分数阈值，且第一预设分数阈值小于第二预设分数阈值，比如可分别设置为 0 和 0.2。

在具体实现时，可将整体情感分数低于第一预设分数阈值（也就是低于 0）的待分析证券研报的情感分析结果确定为看空；将整体情感分数介于第一预设分数阈值与第二预设分数阈值之间（也就是在 0 至 0.2 之间）的待分析证券研报的情感分析结果确定为看平；将整体情感分数高于第二预设分数阈值（也就是高于 0.2）的待分析证券投资研报的情感分析结果确定为看多（也称为看涨）。

以此，可通过将客观的评分结果与预定指标相比对，以得到待分析证券研报的情感分析结果，从而可保证该情感分析结果的客观性及准确性。

根据本实施例的各个实施方式，可预先对大量证券研报中的分词进行提取以生成包括所需情感类型词库的情感词典，基于此，可先将待分析证券研报进行分句以及分词处理，然后通过将得到的各分词与情感词典匹配的方式先确定出预测类分句再确定出预测类分句中各分词的情感类型，接下来基于预测类分句中各分词的情感类型并利用预置评分规则对各预测类分句进行评分，并可基于评分结果计算得到待分析证券研报的整体情感分数，最后根据待分析证券研报的整体情感分数与预置分数阈值的比对结果获得待分析证券研报的情感分析结果。通过本方案，可基于大数据分析及智能化评分的方式实现对待分析证券研报的情感分析，不但可大大节省人力，而且可提高分析效率及分析结果的准确率。

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 中提供的利用大数据的证券研报情感分析方法，本实施例中提供一种利用大数据的证券研报情感分析装置，具体地，图 2 至 3 示出了该利用大数据的证券研报情感分析装置的可选的结构框图，该利用大数据的证券研报情感分析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大数据的证券研报情感分析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，该利用大数据的证券研报情感分析装置 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；

分词模块 23，可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；

预测类分句确定模块 24，可用于基于待分析证券研报中的各分词，利用预先建立的情感词典确定待分析证券研报中的预测类分句；

情感类型确定模块 25，可用于根据所述情感词典确定待分析证券研报中的预测类分句中各分词的情感类型；

评分模块 26，可用于基于待分析证券研报中的预测类分句中各分词的情感类型，利用预置评分规则对待分析证券研报中的各预测类分句进行评分；

计算模块 27，可用于基于待分析证券研报中的各预测类分句的评分，计算得到待分析证券研报的整体情感分数；

情感分析结果获取模块 28，可用于将待分析证券研报的整体情感分数与预设分数阈值进行比对，并根据比对结果获取待分析证券研报的情感分析结果。

在本实施例中，建立情感词典，可包括如下步骤：

步骤 110, 对预置篇数的证券研报进行分句处理, 得到证券研报中的各第一分句;

步骤 120, 对各第一分句进行分词处理, 得到证券研报中的各第一分词;

步骤 130, 从所有第一分词中提取用于表达正面情感的分词、用于表达负面情感的分词、用于表达否定情感的分词及用于表达预测的分词;

5 步骤 140, 根据提取结果分别生成正面情感词库、负面情感词库、否定情感词库及预测类词库;

步骤 150, 基于正面情感词库、负面情感词库、否定情感词库及预测类词库建立情感词典。

进一步的, 预测类分句确定模块 24, 可具体用于:

10 将待分析证券研报中的各分词与所述情感词典中预测类词库中的分词进行匹配;

若待分析证券研报中的任一分词能够与所述情感词典中预测类词库中的分词相匹配, 则将该任一分词所属的分句确定为预测类分句。

进一步的, 情感类型确定模块 25, 可具体用于:

15 将待分析证券研报中的预测类分句中的各分词分别与所述情感词典中正面情感词库、负面情感词库、否定情感词库中的分词进行匹配;

若待分析证券研报中的任一分词能够与正面情感词库、负面情感词库或者否定情感词库中的分词相匹配, 则将该任一分词的情感类型确定为正面分词、负面分词或者否定分词。

在具体实现时, 参看图 3 所示, 评分模块 26, 可具体包括:

20 统计单元 261, 可用于统计待分析证券研报中每个预测类分句中的正面分词、负面分词的初始数量及否定分词的数量;

第一判断单元 262, 可用于判断否定分词的数量是否为奇数;

第二判断单元 263, 可用于当第一判断单元 262 的判断结果为是时, 则判断奇数个否定分词前面或后面相邻的是正面分词还是负面分词;

25 分词数量调整单元 264, 可用于当奇数个否定分词前面或后面相邻的是正面分词时, 则将正面分词的初始数量减一并并负面分词的初始数量加一, 以分别作为调整后的正面分词数量及负面分词数量; 当奇数个否定分词前面或后面相邻的是负面分词时, 则将负面分词的初始数量减一并并正面分词的初始数量加一, 以分别作为调整后的负面分词数量及正面分词数量;

30 评分单元 265, 可用于将待分析证券研报中的各预测类分句中正面分词数量与负面分词数量的差除以正面分词数量与负面分词数量的和, 以得到待分析证券研报中的各预测类分句的评分。

在具体实现时, 待分析证券研报中的各预测类分句的评分包括正分和负分, 分别对应正面预测类分句和负面预测类分句。

基于此, 计算模块 27, 可具体用于:

35 分别计算待分析证券研报中的所有正面预测类分句的平均分及所有负面预测类分句的平均分;

将待分析证券研报中的所有正面预测类分句的平均分与所有负面预测类分句的平均分的二倍进行求和计算, 以得到待分析证券研报的整体情感分数。

40 进一步的, 预设分数阈值可包括第一预设分数阈值和第二预设分数阈值, 第一预设分数阈值小于第二预设分数阈值。

基于此, 情感分析结果获取模块 28, 可具体用于:

将整体情感分数低于第一预设分数阈值、介于第一预设分数阈值与第二预设分数阈值之间、高于第二预设分数阈值的待分析证券投资研报的情感分析结果, 分别确定为看空、看平、看多。

45 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，可预先对大量证券研报中的分词进行提取以生成包括所需情感类型词库的情感词典，基于此，可先将待分析证券研报进行分句以及分词处理，然后通过将得到的各分词与情感词典匹配的方式先确定出预测类分句再确定出预测类分句中各分词的情感类型，接下来基于预测类分句中各分词的情感类型并利用预置评分规则对
5 各预测类分句进行评分，并可基于评分结果计算得到待分析证券研报的整体情感分数，最后根据待分析证券研报的整体情感分数与预置分数阈值的比对结果获得待分析证券研报的情感分析结果。通过本方案，可基于大数据分析及智能化评分的方式实现对待分析证券研报的情感分析，不但可大大节省人力，而且可提高分析效率及分析结果的准确率。

10 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
15 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）、
20 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
25 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 的利用大数据的证券研报情感分析装置的程序代码等。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
30 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，例如利用大数据的证券研报情感分析装置等。

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
35 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大数据的证券研报情感分析装置，被处理器执行时实现实施例 1 的利用大数据的证券研报情感分析方法。

40 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。
45

权利要求书

1、一种利用大数据的证券研报情感分析方法，其特征在于，包括如下步骤：

S01，接收输入的待分析证券研报；

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，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；

S04，基于待分析证券研报中的各分词，利用预先建立的情感词典确定待分析证券研报中的预测类分句；

S05，根据所述情感词典确定待分析证券研报中的预测类分句中各分词的情感类型；

S06，基于待分析证券研报中的预测类分句中各分词的情感类型，利用预置评分规则对待分析证券研报中的各预测类分句进行评分；

S07，基于待分析证券研报中的各预测类分句的评分，计算得到待分析证券研报的整体情感分数；

S08，将待分析证券研报的整体情感分数与预设分数阈值进行比对，并根据比对结果获取待分析证券研报的情感分析结果。

2、根据权利要求1所述的利用大数据的证券研报情感分析方法，其特征在于，建立所述情感词典，包括如下步骤：

步骤110，对预置篇数的证券研报进行分句处理，得到证券研报中的各第一分句；

步骤120，对各第一分句进行分词处理，得到证券研报中的各第一分词；

步骤130，从所有第一分词中提取用于表达正面情感的分词、用于表达负面情感的分词、用于表达否定情感的分词及用于表达预测的分词；

步骤140，根据提取结果分别生成正面情感词库、负面情感词库、否定情感词库及预测类词库；

步骤150，基于正面情感词库、负面情感词库、否定情感词库及预测类词库建立所述情感词典。

3、根据权利要求2所述的利用大数据的证券研报情感分析方法，其特征在于，S04基于待分析证券研报中的各分词，利用预先建立的情感词典确定待分析证券研报中的预测类分句，包括：

将待分析证券研报中的各分词与所述情感词典中预测类词库中的分词进行匹配；

若待分析证券研报中的任一分词能够与所述情感词典中预测类词库中的分词相匹配，则将任一分词所属的分句确定为预测类分句。

4、根据权利要求2所述的利用大数据的证券研报情感分析方法，其特征在于，S05根据所述情感词典确定待分析证券研报中的预测类分句中各分词的情感类型，包括：

将待分析证券研报中的预测类分句中的各分词分别与所述情感词典中正面情感词库、负面情感词库、否定情感词库中的分词进行匹配；

若待分析证券研报中的任一分词能够与正面情感词库、负面情感词库或者否定情感词库中的分词相匹配，则将任一分词的情感类型确定为正面分词、负面分词或者否定分词。

5、根据权利要求4所述的利用大数据的证券研报情感分析方法，其特征在于，S06基于待分析证券研报中的预测类分句中各分词的情感类型，利用预置评分规则对待分析证券研报中的各预测类分句进行评分，包括：

统计待分析证券研报中每个预测类分句中的正面分词、负面分词的初始数量及否定分词的数量；

判断否定分词的数量是否为奇数；

若是，则判断奇数个否定分词前面或后面相邻的是正面分词还是负面分词；

若奇数个否定分词前面或后面相邻的是正面分词，则将正面分词的初始数量减一并将

负面分词的初始数量加一，以分别作为调整后的正面分词数量及负面分词数量；若奇数个否定分词前面或后面相邻的是负面分词，则将负面分词的初始数量减一并将正面分词的初始数量加一，以分别作为调整后的负面分词数量及正面分词数量；

将待分析证券研报中的各预测类分句中正面分词数量与负面分词数量的差除以正面分词数量与负面分词数量的和，以得到待分析证券研报中的各预测类分句的评分。

6、根据权利要求1所述的利用大数据的证券研报情感分析方法，其特征在于，待分析证券研报中的各预测类分句的评分包括正分和负分，分别对应正面预测类分句和负面预测类分句；

S07 基于待分析证券研报中的各预测类分句的评分，计算得到待分析证券研报的整体情感分数，包括：

分别计算待分析证券研报中的所有正面预测类分句的平均分及所有负面预测类分句的平均分；

将待分析证券研报中的所有正面预测类分句的平均分与所有负面预测类分句的平均分的二倍进行求和计算，以得到待分析证券研报的整体情感分数。

7、根据权利要求1所述的利用大数据的证券研报情感分析方法，其特征在于，所述预设分数阈值包括第一预设分数阈值和第二预设分数阈值，第一预设分数阈值小于第二预设分数阈值；

S08 将待分析证券研报的整体情感分数与预设分数阈值进行比对，并根据比对结果获取待分析证券研报的情感分析结果，包括：

将整体情感分数低于第一预设分数阈值、介于第一预设分数阈值与第二预设分数阈值之间、高于第二预设分数阈值的待分析证券投资研报的情感分析结果，分别确定为看空、看平、看多。

8、一种利用大数据的证券研报情感分析装置，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；

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；

分词模块，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；

预测类分句确定模块，用于基于待分析证券研报中的各分词，利用预先建立的情感词典确定待分析证券研报中的预测类分句；

情感类型确定模块，用于根据所述情感词典确定待分析证券研报中的预测类分句中各分词的情感类型；

评分模块，用于基于待分析证券研报中的各预测类分句中各分词的情感类型，利用预设评分规则对待分析证券研报中的各预测类分句进行评分；

计算模块，用于基于待分析证券研报中的各预测类分句的评分，计算得到待分析证券研报的整体情感分数；

情感分析结果获取模块，用于将待分析证券研报的整体情感分数与预设分数阈值进行比对，并根据比对结果获取待分析证券研报的情感分析结果。

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大数据的证券研报情感分析方法的以下步骤：

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，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；

S04，基于待分析证券研报中的各分词，利用预先建立的情感词典确定待分析证券研报中的预测类分句；

S05，根据所述情感词典确定待分析证券研报中的预测类分句中各分词的情感类型；

S06, 基于待分析证券研报中的预测类分句中各分词的情感类型, 利用预置评分规则对待分析证券研报中的各预测类分句进行评分;

S07, 基于待分析证券研报中的各预测类分句的评分, 计算得到待分析证券研报的整体情感分数;

5 S08, 将待分析证券研报的整体情感分数与预设分数阈值进行比对, 并根据比对结果获取待分析证券研报的情感分析结果。

10、根据权利要求 9 所述的计算机设备, 其特征在于, 建立所述情感词典, 包括如下步骤:

步骤 110, 对预置篇数的证券研报进行分句处理, 得到证券研报中的各第一分句;

10 步骤 120, 对各第一分句进行分词处理, 得到证券研报中的各第一分词;

步骤 130, 从所有第一分词中提取用于表达正面情感的分词、用于表达负面情感的分词、用于表达否定情感的分词及用于表达预测的分词;

步骤 140, 根据提取结果分别生成正面情感词库、负面情感词库、否定情感词库及预测类词库;

15 步骤 150, 基于正面情感词库、负面情感词库、否定情感词库及预测类词库建立所述情感词典。

11、根据权利要求 10 所述的计算机设备, 其特征在于, S05 根据所述情感词典确定待分析证券研报中的预测类分句中各分词的情感类型, 包括:

20 将待分析证券研报中的预测类分句中的各分词分别与所述情感词典中正面情感词库、负面情感词库、否定情感词库中的分词进行匹配;

若待分析证券研报中的任一分词能够与正面情感词库、负面情感词库或者否定情感词库中的分词相匹配, 则将该任一分词的情感类型确定为正面分词、负面分词或者否定分词。

25 12、根据权利要求 11 所述的计算机设备, 其特征在于, S06 基于待分析证券研报中的预测类分句中各分词的情感类型, 利用预置评分规则对待分析证券研报中的各预测类分句进行评分, 包括:

统计待分析证券研报中每个预测类分句中的正面分词、负面分词的初始数量及否定分词的数量;

判断否定分词的数量是否为奇数;

若是, 则判断奇数个否定分词前面或后面相邻的是正面分词还是负面分词;

30 若奇数个否定分词前面或后面相邻的是正面分词, 则将正面分词的初始数量减一并将负面分词的初始数量加一, 以分别作为调整后的正面分词数量及负面分词数量; 若奇数个否定分词前面或后面相邻的是负面分词, 则将负面分词的初始数量减一并将正面分词的初始数量加一, 以分别作为调整后的负面分词数量及正面分词数量;

35 将待分析证券研报中的各预测类分句中正面分词数量与负面分词数量的差除以正面分词数量与负面分词数量的和, 以得到待分析证券研报中的各预测类分句的评分。

13、根据权利要求 9 所述的计算机设备, 其特征在于, 待分析证券研报中的各预测类分句的评分包括正分和负分, 分别对应正面预测类分句和负面预测类分句;

S07 基于待分析证券研报中的各预测类分句的评分, 计算得到待分析证券研报的整体情感分数, 包括:

40 分别计算待分析证券研报中的所有正面预测类分句的平均分及所有负面预测类分句的平均分;

将待分析证券研报中的所有正面预测类分句的平均分与所有负面预测类分句的平均分的二倍进行求和计算, 以得到待分析证券研报的整体情感分数。

45 14、根据权利要求 9 所述的计算机设备, 其特征在于, 所述预设分数阈值包括第一预设分数阈值和第二预设分数阈值, 第一预设分数阈值小于第二预设分数阈值;

S08 将待分析证券研报的整体情感分数与预设分数阈值进行比对, 并根据比对结果获

取待分析证券研报的情感分析结果，包括：

将整体情感分数低于第一预设分数阈值、介于第一预设分数阈值与第二预设分数阈值之间、高于第二预设分数阈值的待分析证券投资研报的情感分析结果，分别确定为看空、看平、看多。

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大数据的证券研报情感分析方法的以下步骤：

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，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；

S04，基于待分析证券研报中的各分词，利用预先建立的情感词典确定待分析证券研报中的预测类分句；

S05，根据所述情感词典确定待分析证券研报中的预测类分句中各分词的情感类型；

S06，基于待分析证券研报中的预测类分句中各分词的情感类型，利用预置评分规则对待分析证券研报中的各预测类分句进行评分；

S07，基于待分析证券研报中的各预测类分句的评分，计算得到待分析证券研报的整体情感分数；

S08，将待分析证券研报的整体情感分数与预设分数阈值进行比对，并根据比对结果获取待分析证券研报的情感分析结果。

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，建立所述情感词典，包括如下步骤：

步骤110，对预置篇数的证券研报进行分句处理，得到证券研报中的各第一分句；

步骤120，对各第一分句进行分词处理，得到证券研报中的各第一分词；

步骤130，从所有第一分词中提取用于表达正面情感的分词、用于表达负面情感的分词、用于表达否定情感的分词及用于表达预测的分词；

步骤140，根据提取结果分别生成正面情感词库、负面情感词库、否定情感词库及预测类词库；

步骤150，基于正面情感词库、负面情感词库、否定情感词库及预测类词库建立所述情感词典。

17、根据权利要求16所述的计算机可读存储介质，其特征在于，S05根据所述情感词典确定待分析证券研报中的预测类分句中各分词的情感类型，包括：

将待分析证券研报中的预测类分句中的各分词分别与所述情感词典中正面情感词库、负面情感词库、否定情感词库中的分词进行匹配；

若待分析证券研报中的任一分词能够与正面情感词库、负面情感词库或者否定情感词库中的分词相匹配，则将该任一分词的情感类型确定为正面分词、负面分词或者否定分词。

18、根据权利要求17所述的计算机可读存储介质，其特征在于，S06基于待分析证券研报中的预测类分句中各分词的情感类型，利用预置评分规则对待分析证券研报中的各预测类分句进行评分，包括：

统计待分析证券研报中每个预测类分句中的正面分词、负面分词的初始数量及否定分词的数量；

判断否定分词的数量是否为奇数；

若是，则判断奇数个否定分词前面或后面相邻的是正面分词还是负面分词；

若奇数个否定分词前面或后面相邻的是正面分词，则将正面分词的初始数量减一并将负面分词的初始数量加一，以分别作为调整后的正面分词数量及负面分词数量；若奇数个否定分词前面或后面相邻的是负面分词，则将负面分词的初始数量减一并将正面分词的初始数量加一，以分别作为调整后的负面分词数量及正面分词数量；

将待分析证券研报中的各预测类分句中正面分词数量与负面分词数量的差除以正面分

词数量与负面分词数量的和，以得到待分析证券研报中的各预测类分句的评分。

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，待分析证券研报中的各预测类分句的评分包括正分和负分，分别对应正面预测类分句和负面预测类分句；

5 S07 基于待分析证券研报中的各预测类分句的评分，计算得到待分析证券研报的整体情感分数，包括：

分别计算待分析证券研报中的所有正面预测类分句的平均分及所有负面预测类分句的平均分；

将待分析证券研报中的所有正面预测类分句的平均分与所有负面预测类分句的平均分的二倍进行求和计算，以得到待分析证券研报的整体情感分数。

10 20、根据权利要求 15 所述的计算机可读存储介质，其特征在于，所述预设分数阈值包括第一预设分数阈值和第二预设分数阈值，第一预设分数阈值小于第二预设分数阈值；

S08 将待分析证券研报的整体情感分数与预设分数阈值进行比对，并根据比对结果获取待分析证券研报的情感分析结果，包括：

15 将整体情感分数低于第一预设分数阈值、介于第一预设分数阈值与第二预设分数阈值之间、高于第二预设分数阈值的待分析证券投资研报的情感分析结果，分别确定为看空、看平、看多。

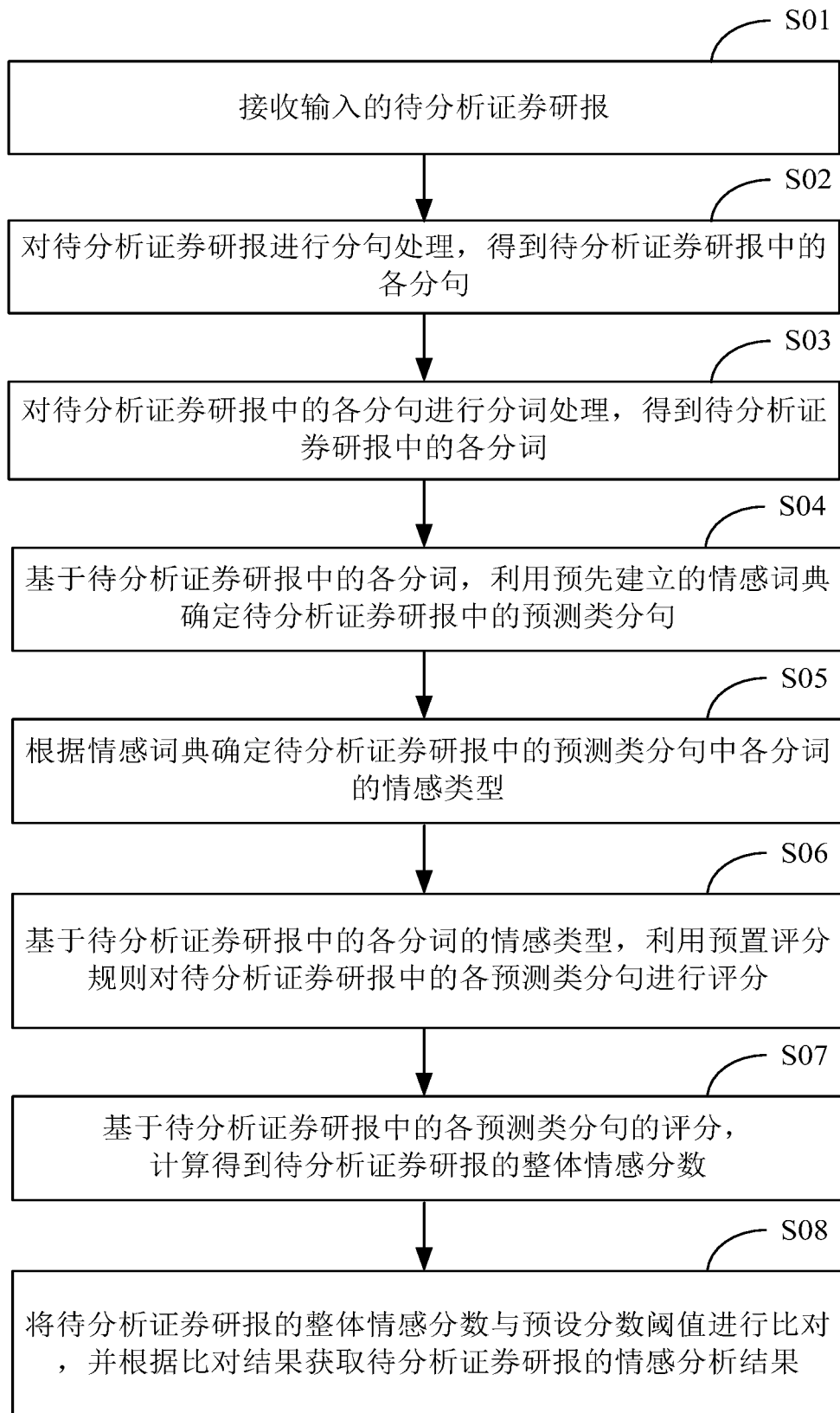


图 1

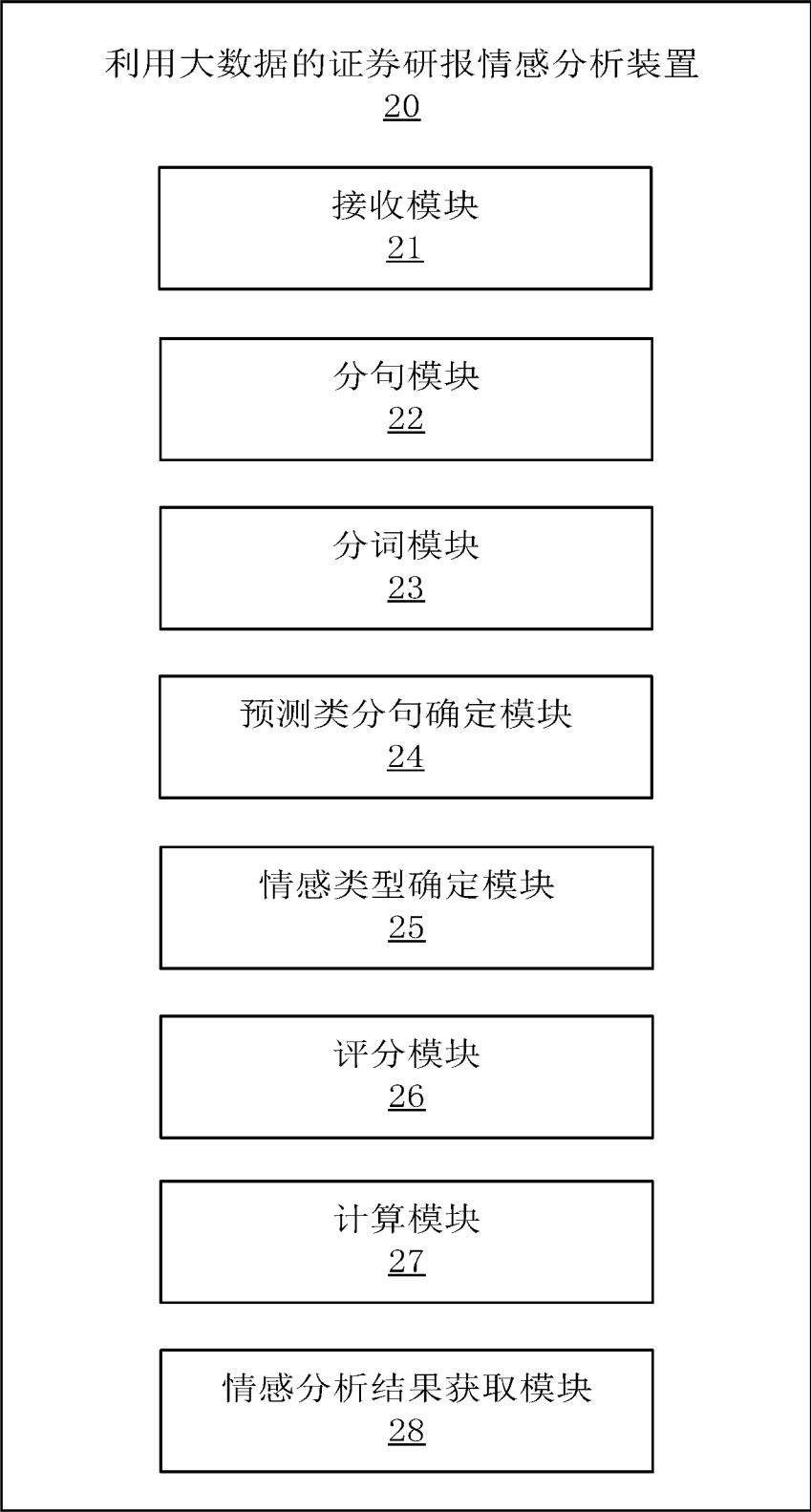


图 2



图 3

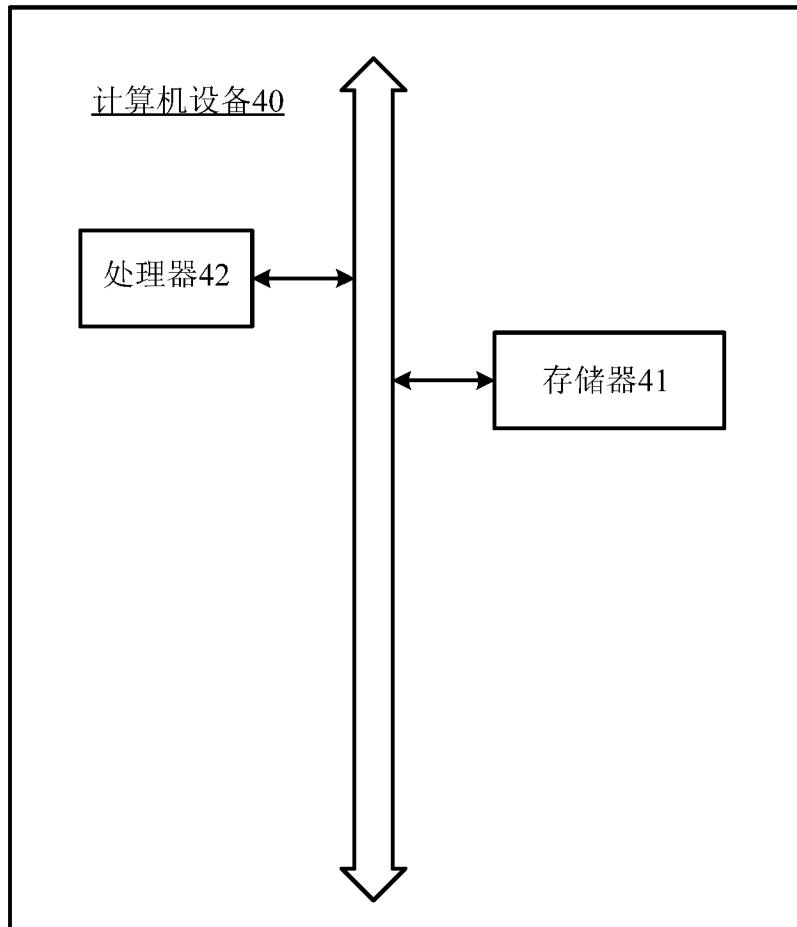


图 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/070291

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, CNPAT, CNKI, IEEE: 平安科技(深圳)有限公司, 证券, 股票, 文本, 情感, 分句, 句子, 分词, 词语, 情感词典, 分类, 褒义, 贬义, 正面, 负面, 正向, 负向, 训练模型, 比对, 比较, 得分, 分数, stock, share, text, emotion+, dictionary, sentence, term?, classif+, positive, negative, training model, compar+, score

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 108090040 A (BEIJING GRIDSUM TECHNOLOGY CO., LTD.) 29 May 2018 (2018-05-29) description, paragraphs [0050]-[0071] and [0109]-[0118]	1-20
Y	CN 108629693 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 09 October 2018 (2018-10-09) description, paragraphs [0053]-[0058] and [0063]-[0067]	1-20
A	CN 103970864 A (TSINGHUA UNIVERSITY) 06 August 2014 (2014-08-06) entire document	1-20
A	US 2014343923 A1 (EDUCATIONAL TESTING SERVICE) 20 November 2014 (2014-11-20) 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

13 June 2019

Date of mailing of the international search report

01 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/CN2019/070291

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	108090040	A	29 May 2018	None	
CN	108629693	A	09 October 2018	None	
CN	103970864	A	06 August 2014	CN 103970864	B 22 September 2017
US	2014343923	A1	20 November 2014	None	

国际检索报告

国际申请号

PCT/CN2019/070291

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, CNPAT, CNKI, IEEE: 平安科技 (深圳) 有限公司, 证券, 股票, 文本, 情感, 分句, 句子, 分词, 词语, 情感词典, 分类, 褒义, 贬义, 正面, 负面, 正向, 负向, 训练模型, 比对, 比较, 得分, 分数, stock, share, text, emotion+, dictionary, sentence, term?, classif+, positive, negative, training model, compar+, score

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
Y	CN 108090040 A (北京国双科技有限公司) 2018年 5月 29日 (2018 - 05 - 29) 说明书第[0050]-[0071], [0109]-[0118]段	1-20
Y	CN 108629693 A (平安科技深圳有限公司) 2018年 10月 9日 (2018 - 10 - 09) 说明书第[0053]-[0058], [0063]-[0067]段	1-20
A	CN 103970864 A (清华大学) 2014年 8月 6日 (2014 - 08 - 06) 全文	1-20
A	US 2014343923 A1 (EDUCATIONAL TESTING SERVICE) 2014年 11月 20日 (2014 - 11 - 20) 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 6月 13日

国际检索报告邮寄日期

2019年 7月 1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

李艳红

电话号码 86-(10)-53961221

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/070291

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	108090040	A	2018年 5月 29日	无	
CN	108629693	A	2018年 10月 9日	无	
CN	103970864	A	2014年 8月 6日	CN 103970864 B	2017年 9月 22日
US	2014343923	A1	2014年 11月 20日	无	