

(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
2018 年 10 月 11 日 (11.10.2018)



(10) 国际公布号
WO 2018/184518 A1

(51) 国际专利分类号:
G06F 17/27 (2006.01) **G06Q 50/00** (2012.01)

(21) 国际申请号: PCT/CN2018/081697

(22) 国际申请日: 2018 年 4 月 3 日 (03.04.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201710225681.5 2017 年 4 月 7 日 (07.04.2017) 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八卦岭八卦三路平安大厦六楼, Guangdong 518052 (CN)。

(72) 发明人: 王健宗 (WANG, Jianzong); 中国广东省深圳市福田区八卦岭八卦三路平安大厦六楼, Guangdong 518052 (CN)。 黄章成 (HUANG, Zhangcheng); 中国广东省深圳市福田区八卦岭八卦三路平安大厦六楼, Guangdong 518052 (CN)。 吴天博 (WU, Tianbo); 中国广东省深圳

市福田区八卦岭八卦三路平安大厦六楼, Guangdong 518052 (CN)。 肖京 (XIAO, Jing); 中国广东省深圳市福田区八卦岭八卦三路平安大厦六楼, Guangdong 518052 (CN)。

(74) 代理人: 广州华进联合专利商标代理有限公司 (ADVANCE CHINA IP LAW OFFICE); 中国广东省广州市天河区花城大道 85 号 3901 房, Guangdong 510623 (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。

(54) Title: MICROBLOG DATA PROCESSING METHOD AND DEVICE, COMPUTER DEVICE AND STORAGE MEDIUM

(54) 发明名称: 微博数据处理方法、装置、计算机设备及存储介质

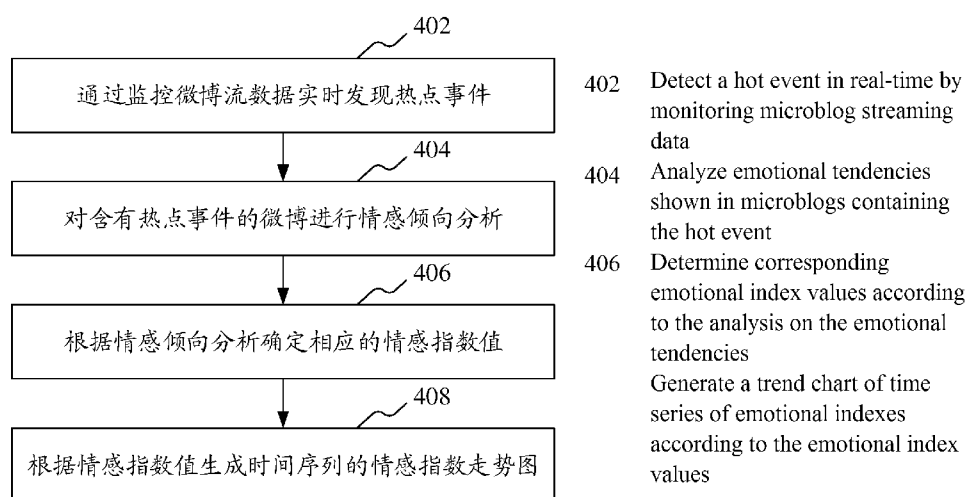


图 4

(57) Abstract: A microblog data processing method comprises: detecting a hot event in real-time by monitoring microblog streaming data (402); analyzing emotional tendencies shown in microblogs containing the hot event (404); determining corresponding emotional index values according to the analysis on the emotional tendencies (406); and generating a trend chart of time series of emotional indexes according to the emotional index values (408).

(57) 摘要: 一种微博数据处理方法, 包括: 通过监控微博流数据实时发现热点事件 (402); 对含有所述热点事件的微博进行情感倾向分析 (404); 根据所述情感倾向分析确定相应的情感指数值 (406); 根据所述情感指数值生成时间序列的情感指数走势图 (408)。

(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))。

发明名称：微博数据处理方法、装置、计算机设备及存储介质

本申请要求于 2017 年 4 月 7 日提交中国专利局、申请号为 2017102256815、发明名称为“微博数据处理方法、装置、计算机设备及存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

5

技术领域

本申请涉及计算机处理领域，特别是涉及一种微博数据处理方法、装置、计算机设备及存储介质。

10 **背景技术**

随着社交媒体的发展，社交网站、在线社区、微博等已逐渐成为人们生活中不可或缺的一部分，也是当今时代信息传播的主要渠道。信息传播形态的巨大变革，冲击着投资者原有的信息利用模式与投资信念，并深刻影响了资本市场的信息传递与金融生态。行为金融理论认为投资者的投资决策行为受到投资者注意力、情绪等因素的共同影响，而传统的股市预测模型中虽然也有涉及到情感倾向分析，但是大多数情感分析是基于新闻网页，其并不能真正反映投资者注意力和情绪，导致在进行股市预测时往往会出现较大偏差。

发明内容

20 根据本申请的各种实施例，提供了一种微博数据处理方法、装置、计算机设备以及存储介质。

一种微博数据处理方法，包括：

通过监控微博流数据实时发现热点事件；

对含有所述热点事件的微博进行情感倾向分析；

25 根据所述情感倾向分析确定相应的情感指数值；及

根据所述情感指数值生成时间序列的情感指数走势图。

一种微博数据处理装置，包括：

发现模块，用于通过监控微博流数据实时发现热点事件；

30 分析模块，用于对含有所述热点事件的微博进行情感倾向分析；

确定模块，用于根据所述情感倾向分析确定相应的情感指数值；及

生成模块，用于根据所述情感指数值生成时间序列的情感指数走势图。

一种计算机设备，包括存储器和一个或多个处理器，所述存储器中存储有计算机可读指令，所述计算机可读指令被所述一个或多个处理器执行时，使得所述一个或多个处理器执行以下步骤：

通过监控微博流数据实时发现热点事件；

5 对含有所述热点事件的微博进行情感倾向分析；

根据所述情感倾向分析确定相应的情感指数值；及

根据所述情感指数值生成时间序列的情感指数走势图。

10 一个或多个存储有计算机可读指令的非易失性计算机可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行以下步骤：

通过监控微博流数据实时发现热点事件；

对含有所述热点事件的微博进行情感倾向分析；

根据所述情感倾向分析确定相应的情感指数值；及

15 根据所述情感指数值生成时间序列的情感指数走势图。

本申请的一个或多个实施例的细节在下面的附图和描述中提出。本申请的其它特征、目的和优点将从说明书、附图以及权利要求书变得明显。

20 附图说明

为了更清楚地说明本申请实施例中的技术方案，下面将对实施例描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

25 图 1 为根据一个或多个实施例中微博数据处理方法的应用场景图。

图 2 为根据一个或多个实施例中终端的内部结构框图。

图 3 为根据一个或多个实施例中服务器的框图。

图 4 为根据一个或多个实施例中微博数据处理方法的流程图。

30 图 5 为根据一个或多个实施例中通过监控微博流数据实时发现热点事件的方法流程图。

图 6 为根据一个或多个实施例中将含有热点事件的微博进行情感倾向分析的方法流程图。

图 7 为根据一个或多个实施例中根据情感倾向分析确定相应的情感指数值的方法流程图。

图 8 为根据一个或多个实施例中微博数据处理装置的框图。

图 9 为根据一个或多个实施例中发现模块的框图。

图 10 为根据一个或多个实施例中分析模块的框图。

图 11 为根据一个或多个实施例中确定模块的框图。

5

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仅用以解释本申请，并不用于限定本申请。

10 本申请提供的微博数据处理方法，可以应用于如图 1 所示的应用场景中。终端 102 与服务器 104 通过网络进行通信。终端 102 可以但不限于各种个人计算机、笔记本电脑、智能手机、平板电脑和便携式可穿戴设备，服务器 104 可以用独立的服务器或者是多个服务器组成的服务器集群来实现。首先，终端 102 将微博流数据上传到服务器 104，服务器 104 通过监控微博流数据
15 实时发现热点事件，对含有所述热点事件的微博进行情感倾向分析，根据所述情感倾向分析确定相应的情感指数值；及根据所述情感指数值生成时间序列的情感指数走势图。

如图 2 所示，在其中一个实施例中，终端 102 的内部结构如图 2 所示，包括通过系统总线连接的处理器、存储器、网络接口、显示屏和输入装置。
20 其中，该终端的处理器用于提供计算和控制能力，支撑整个终端的运行。该终端的存储器包括非易失性计算机可读存储介质、内存储器。终端 102 的非易失性计算机可读存储介质存储有操作系统和计算机可读指令，该计算机可读指令被执行时，可使得处理器执行一种微博数据处理方法。该内存储器为非易失性计算机可读存储介质中的操作系统和计算机可读指令的运行提供环
25 境。网络接口用于连接到网络进行通信。终端 102 的显示屏可以是液晶显示屏或者电子墨水显示屏等，输入装置可以是显示屏上覆盖的触摸层，也可以是电子设备外壳上设置的按键、轨迹球或触控板，也可以是外接的键盘、触控板或鼠标等。该终端可以是平板电脑、笔记本电脑、台式计算机等。本领域技术人员可以理解，图 2 中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的终端的限定，具体的终端可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。
30

如图 3 所示，在其中一个实施例中，服务器 104 的内部结构如图 3 所示，包括通过系统总线连接的处理器、存储器和网络接口。其中，该服务器的处理器用于提供计算和控制能力、支撑整个服务器的运行。该服务器的存储器
35

包括非易失性计算机可读存储介质、内存储器。该非易失性计算机可读存储介质可存储操作系统和计算机可读指令。该计算机指令被执行时，可使得处理器执行一种微博数据处理方法。该服务器的网络接口用于与外部的服务器和终端通过网络连接通信。本领域技术人员可以理解，图 3 中示出的结构，
5 仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的服务器的限定，具体的服务器可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

如图 4 所示，在其中一个实施例中，提出了一种微博数据处理方法，该方法可应用于终端或服务器中，具体包括以下步骤：

10 步骤 402，通过监控微博流数据实时发现热点事件。

在其中一个实施例中，热点事件是指当前关注度或影响力比较大的事件。通过监控微博平台的数据流数据可以实时发现当前的热点话题，热点话题对应的事件就是热点事件。热点事件能够反映出投资者的注意力和情感倾向，所以及时发现热点事件有利于发现投资者的注意力和情感倾向。

15 在其中一个实施例中，通过社交媒体实时热点发现算法发掘微博平台实时热点话题。具体地，首先，利用动态更新微博流算法，获取微博平台实时主微博流。其中，主微博流是指从大量实时微博数据流中抽样出的最具代表性的那部分微博数据流。动态更新微博流算法实际上是指对监控的微博账号的更新规则，该更新规则用于保证通过监控账号获取到的微博数据流能够全面反映整体微博流。举个例子，比如，微博数据流最初是通过采集选定的 10
20 万个微博账号中的微博构建的，但是这 10 万个账号随着时间推移，可能会出现不同的运营状况，比如，停止更新、被盗号之类的，或者这 10 万个微博账号的实时微博流不能很好地代表整个实时微博数据流，那么就需要动态更新选择的微博账号，比如，删除超过 5 天不更新的账号，加入新的活跃账号。

25 其次，将主微博流向量化，计算一阶词频和二阶词频出现的加速度。其中，向量化是指将获取到的主微博流以词向量的形式进行表示，具体可以采用 word2vec 方法将该主微博流表示为向量的形式，word2vec 是一个将单词转换成向量形式的工具，可以把文本内容的处理简化为向量空间中的向量运算。

30 因为自然语言理解的问题要转换为机器学习的问题，首先需要将这些文本数学化，即将文本转换为数学的表示形式。举个例子，话筒表示为[0.792, -0.177, -0.107, 0.109, -0.542,]。即将文本以数学化的向量进行表示。可以通过计算出向量空间上的相似度就可以表示文本语义上的相似度。一阶词频是指单个词出现的频率，二阶词频是指同时两个词出现的频率。最后，根据预先设置的各个加速度阈值，实时监控发现微博热点话题。比如，可以设置一
35 阶词频和二阶词频对应的加速度阈值，当达到该加速度阈值时，将对应的话

题确定为热点话题，相应的事件也就是热点事件。此外，还可以通过关注关键词的变化来获取热点话题内关注度的迁徙，通过监控热点事件关注点的迁移，可以实时获取与之相关的受到影响的股票。

在另一个实施例中，首先，实时获取微博流数据，将微博流数据向量化；其次，监控向量化后的微博数据流，记录每个词出现的频率和次数；最后，根据每个词出现的频率和次数确定当前的热点事件。在该实施例中，为了对微博流进行实时监控，需要将微博流数据转换为向量的表示形式来进行监控，所以当获取到微博流数据后，将每一条微博转换为向量的形式来表示。然后监控向量化的微博数据，记录每个词出现的频率和次数，当某个词出现的频率和次数都比较高时，将对应的事件确定为热点事件。

步骤 404，对含有热点事件的微博进行情感倾向分析。

在其中一个实施例中，为了获取投资者的情感倾向，当获取到包含热点事件的微博后，对每一条微博进行情感倾向分析。在情感倾向划分方面，根据心理学最新的一份研究结果，将人类情感分为四类，分别为快乐、悲伤、愤怒和惊奇。具体地，为了对每条微博进行情感倾向分析，需要建立一个情感倾向分析模型。在情感倾向分析模型建立之前，首先需要确定训练集，用于对情感倾向分析模型进行训练。训练集是经过情感标注后的四分类微博数据集，其中，情感标注是基于动态情感词典的机器标注和人工标注相结合的混合标注方法。该混合标注方法的优势是，一方面可以避免单纯采用一类标注方法的误差，另一方面，可以节省人工标注的时间。其次，从进行情感标注后的微博文本中抽取特征用于对情感倾向分析模型进行训练。在其中一个实施例中，采用 LDA 模型和 word2vec 模型进行特征的抽取，其中，LDA 模型用于预测每条微博的主题分布；word2vec 模型用于获取每条微博的词向量表示。在另一个实施例中，也可以采用主成分分析法进行特征的抽取。最后，采用情感倾向分析模型进行情感倾向分析，其中，若情感倾向分析模型在训练时是基于抽取的 LDA 主题特征和 word2vec 词向量特征进行训练的，那么该情感倾向分析模型训练完成后，将获取到的新的微博对应的 LDA 主题特征和 word2vec 词向量特征一起作为情感倾向分析模型的输入向量，经过该情感倾向分析模型输出每条微博对应的情感倾向。情感倾向也分为四种，快乐、悲伤、愤怒和惊奇。

步骤 406，根据情感倾向分析确定相应的情感指数值。

在其中一个实施例中，在根据情感倾向分析模型对包含热点事件的微博进行情感倾向分析后，统计每一类情感对应的微博数目，按照情感类别，可以将情感分为四类，分别是快乐、悲伤、愤怒和惊奇。根据统计出来的每一类情感对应的微博数目来计算热点事件对应的情感指数值。其中，情感指数

包括快乐情感占比、悲伤情感占比、愤怒情感占比、惊奇情感占比以及情感效价中的一种或多种。其中，情感效价是指整体微博情绪的反映值，情感效价计算公式为：

SentmentValence= $\log\{(1+P)/(1+N)\}$ ，其中，P 为积极情感微博总数，N 为消极情感微博总数。积极情感微博总数为快乐和惊奇情感微博数量之和，消极情感微博总数为悲伤和愤怒情感微博数量之和。具体地，比如当天的热点事件相关的总微博数目为 10 万条，其中，快乐情感的微博条数为 3 万，悲伤情感的微博条数为 1 万，愤怒情感的微博条数为 1 万，惊奇情感的微博条数为 5 万，那么对应的快乐情感占比为 0.3，悲伤和愤怒情感占比都是 0.1，惊奇情感占比为 0.5。情感效价为 $\log 3$ 。

步骤 408，根据情感指数值生成时间序列的情感指数走势图。

在其中一个实施例中，通过情感倾向分析确定相应的情感指数值后，将该情感指数值与对应的统计时间点或时间段进行对应存储，通过记录一段时间内的每一个时间点对应的情感指数值后，按照时间的顺序生成时间序列的情感指数走势图，该情感指数走势图能够充分的且直观的反映出一段时间内投资者情绪的波动。在其中一个实施例中，通过将该情感指数的走势与股票大盘走势进行相关性分析，可以确定出影响股市走势的情感因子。后续可以将该情感因子和其他股市数据一起作为股市预测模型的预测因子进行相应的股市预测，由于充分考虑了投注者的注意力和情绪，有利于提高股市预测的准确率。

在其中一个实施例中，通过监控微博流数据实时发现热点事件，对含有热点事件的微博进行情感倾向分析，进而根据情感倾向分析确定相应的情感指数值，根据所述情感指数值生成时间序列的情感指数走势图。该微博数据处理方法选出了能够反应投资者注意力和情绪的微博热点事件，通过该微博热点事件进行情感倾向分析，并根据情感指数值生成时间序列的情感指数走势图，该情感指数走势图能够充分直观的反映投资者的注意力和情绪，若应用于股市预测有利于提升股市预测的准确率。

如图 5 所示，在其中一个实施例中，通过监控微博流数据实时发现热点事件的步骤 402 包括：

步骤 402A，实时获取微博流数据，将微博流数据向量化。

在其中一个实施例中，为了能够及时的获取到投资者的注意和情绪，实时获取微博流数据，最初获取到的微博数据流为文本数据，文本数据与普通的数值数据或者类属数据不同，文本数据是一种半结构化数据，在进行文本数据分析之前，需要对文本数据进行预处理，采用向量化的数值来表达这些半结构的文本数据。在其中一个实施例中，在进行向量化之前，先对文本数

据进行分词处理，经过分词处理后，一个文本数据就可以表示为由若干关键词来表示的多维向量，便于后续进行特征词的监控。

步骤 402B，监控向量化后的微博流数据，记录每个特征词出现的频率和次数。

5 在其中一个实施例中，特征词是指能够代表某个事件的词汇，一条微博文本中并不是所有的词语都需要记录，因为有一些常用词在任何事件中都可能出现，所以记录这些常用词并没有意义，比如，助词“的”、“地”等。将微博流数据向量化后，实时监控向量化后的微博流数据，监控每个特征词出现的频率和次数，频率是指单位时间内该词语出现的次数。次数是指该词语出现的总数量。通过监控每个特征词出现的频率可以发现短时间内出现的热点话题，通过记录每个特征词出现的次数，可以考察较长一段时间内人们对某个话题的关注度。

步骤 402C，根据每个特征词出现的频率和次数确定当前的热点事件。

15 在其中一个实施例中，监听并记录每个特征词出现的频率和次数，当一个或多个特征词的频率或次数达到了预设阈值时，则判定该一个或多个特征词对应的事件为热点事件。可以分别设置频率阈值和次数阈值，一个实施例中，当特征词出现的频率达到了频率阈值或者当特征词出现的次数达到了次数阈值，则将特征词对应的事件确定为热点事件。在另一个实施例中，当特征词出现的频率达到了频率阈值，且，次数也达到了次数阈值时，则将该特征词对应的事件确定为热点事件。根据监控到的一个或多个特征词出现的频率可以发现短时间内的热点事件，而通过监控一个或多个特征词出现的次数则可以考察一段时间内的热点话题，进而确定该话题的热度和持续时间。该方法有利于及时发现并记录热点事件。

25 如图 6 所示，在其中一个实施例中，将含有热点事件的微博进行情感倾向分析的步骤包括：

步骤 404A，抽取获取到的含有热点事件的每条微博的 LDA 主题特征和 word2vec 词向量特征。

30 在其中一个实施例中，当通过监控微博流数据发现热点事件后，获取含有该热点事件的所有微博，对获取到的含有热点事件的每条微博分别抽取相应的 LDA 主体特征和 word2vec 词向量特征。预先训练好 LDA 主题模型和 word2vec 词向量模型，其中，通过将每条微博的向量文本输入该 LDA 主题模型可以预测该条微博的主题分布，比如，将前 250 维的概率分布作为该条微博的 250 维特征；通过将每条微博的向量文本输入该 word2vec 模型可以获取任一微博词语的向量表示，同样可以选取前 250 维作为特征，通过简单的同维相加，可以获取该条微博的 500 维词向量表示。其中，word2vec 是一款

35

将词表征为实数值向量的高效工具，其利用深度学习的思想，可以通过训练，把对文本内容的处理简化为 k 维向量空间中的向量运算。

步骤 404B，将 LDA 主题特征和 word2vec 词向量特征代入情感倾向分析模型，输出每条微博的情感倾向。

5 在其中一个实施例中，在获取到每条微博对应的 LDA 主题特征和 word2vec 词向量特征后，该获取到的特征代入情感倾向分析模型，输出每条微博的情感倾向。其中，情感倾向分析模型可以采用 Boosted tree 算法建立，其中，Boosted tree 算法主要用于多分类问题的预测。在其中一个实施例中，情感倾向分为四类，分别是快乐、悲伤、愤怒和惊奇。在采用情感倾向分析
10 模型之前，首先需要建立情感倾向分析模型，具体地，采用四分类标注的方法对微博数据集进行情感标注，将进行情感标注后的微博数据集作为训练集，从训练集中的每条微博中抽取 LDA 主题特征和 word2vec 词向量特征代入初始化的情感倾向分析模型进行训练，得到相应的模型参数，从而得到最终的情感倾向分析模型。其中，初始化的情感倾向分析模型是采用 Boosted tree 方法
15 进行建立的。在训练之前需要对每个特征进行标准化处理，指标标准化的公式为： $ZX_t = (X_t - X) / \sigma$ ，其中， X_t 是每维特征具体的数值， ZX_t 是对应的每维数据标准化后的值， X 为该列特征所有数值的均值， σ 是该列特征所有数值的标准差，利用该公式对每维特征做标准化处理，便于后续进行机器学习得到情感倾向分析模型。所以对含有热点事件的每条微博进行情感倾向分析时，
20 需要抽取每条微博中的 LDA 主题特征和 word2vec 词向量作为情感倾向分析模型的输入向量，然后输出每条微博对应的情感倾向，比如，是快乐的还是悲伤的，是愤怒的还是惊奇的。

如图 7 所示，在其中一个实施例中，根据情感倾向分析确定相应的情感指数值的步骤包括：

25 步骤 406A，根据情感倾向分析将含有热点事件的微博按照情感类别进行分类，分别统计每一类情感对应的微博数。

在其中一个实施例中，对包含热点事件的每条微博进行情感性分析以后，将含有热点事件的微博按照情感类别进行分类，分别统计每一类情感对应的微博数。其中，情感类别分为四类，分别是快乐、悲伤、愤怒和惊奇。为了
30 获取投资者的情绪，分别统计每一类情感对应的微博数，便于后续进行情感指数的计算。

步骤 406B，根据统计出的每一类情感对应的微博数确定热点事件对应的情感指数值。

35 在其中一个实施例中，情感指数包括快乐情感占比、悲伤情感占比、愤怒情感占比、惊奇情感占比以及情感效价中的一种或多种。其中，情感效价

是指整体微博情绪的反映值，情感效价计算公式为：
$$\text{SentmentValence} = \log\{(1+P)/(1+N)\}$$
，其中，P为积极情感微博总数，N为消极情感微博总数。积极情感微博总数为快乐和惊奇情感微博数量之和，消极情感微博总数为悲伤和愤怒情感微博数量之和。情感指数值就是相应的情感指数的值。比如，当天的热点事件相关的总微博数目为10万条，其中，快乐情感的微博条数为3万，悲伤情感的微博条数为1万，愤怒情感的微博条数为1万，惊奇情感的微博条数为5万，那么情感指数值包括：快乐情感占比0.3，悲伤和愤怒情感占比都是0.1，惊奇情感占比0.5，情感效价 $\log 3$ 。后续根据计算得到的情感指数值进行股市的预测。需要说明的是，这里的热点事件并不是指一个热点事件，而是一段时间内的所有热点事件。

在其中一个实施例中，在根据情感指数值生成时间序列的情感指数走势图的步骤之后包括：根据所述情感指数走势图确定影响股市数据的情感因子，将情感因子和其他股市数据一起作为股市预测模型的预测因子进行相应的股市预测。

在其中一个实施例中，在获取到时间序列的情感指数走势图后，通过情感指数走势图确定影响股市数据的情感因子。通过将时间序列的情感指数走势图与大盘日收益率、价格和交易额进行相关性分析，将存在相关性的情感指数抽取为股市大盘的预测因子，即确定影响股市数据的情感因子。其中，情感指数包括情感效价、快乐情感占比、悲伤情感占比、愤怒情感占比、惊奇情感占比等。所以在分析时，需要将每个因素都进行分析，筛选出对股市数据有影响的一个或多个因素，共同作为股市数据的情感因子。然后将确定的情感因子和其他传统的股市因子一起作为股市预测模型的预测因子进行相应的股市预测。其他股市数据包括开盘价、收盘价、最高价、最低价、成交额、收益率等数据。具体地，首先，需要通过对情感指数与大盘日收益率、价格和交易额进行相关性分析，存在相关性的情感指数抽取为股市大盘的预测特征，采用机器学习算法，训练得到包含情感特征的股市预测模型，进而根据计算得到的情感指数值（即情感因子对应的情感指数值）和获取的其他股市数据采用该股市预测模型进行相应的股市预测。这里的股市预测，可以预测大盘的收益状况，也可以预测单支股票的收益状况，不同的预测目的，采用的数据不同。通过考虑投资者注意力和情绪进行相应的股市预测，有利于提高股市预测的准确率和可靠性，为决策者提供可靠的依据。

此外，在其中一个实施例中，通过监控微博流数据还可以发现对社会影响巨大，且持续时间较久的热点事件，通过分析该影响大的热点事件有利于股票的选择。因为影响大的热点事件会引起投资大众的过度关注，造成当前价格的非理性波动，而未来一段时间后会回归基本价值，所以这中间会存在

一些买入时机。具体地，首先，确定热点事件的事件影响力，当事件影响力大于预设阈值时，选择股市中股票名称或者股票行业分类名称中包含事件关键字的股票或行业，构成候选股票池；根据股票信息从候选股票池中选出基本面好股票；根据事件周期和大盘走势确定对相关股票的影响度，然后根据相应的影响度反馈给选股择时模型，用于进行股票的选择。比如，美国大选对于黄金、石油等价格引起的短时间内的非理性波动，可以选择合适的时机抄底买入。

应该理解的是，虽然图 4 至 7 的流程图中的各个步骤按照箭头的指示依次显示，但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明，这些步骤的执行并没有严格的顺序限制，这些步骤可以以其它的顺序执行。而且，图 4 至 7 中的至少一部分步骤可以包括多个子步骤或者多个阶段，这些子步骤或者阶段并不必然是在同一时刻执行完成，而是可以在不同的时刻执行，这些子步骤或者阶段的执行顺序也不必然是依次进行，而是可以与其它步骤或者其它步骤的子步骤或者阶段的至少一部分轮流或者交替地执行。

如图 8 所示，在其中一个实施例中，提出了一种微博数据处理装置，该装置包括：

发现模块 802，用于通过监控微博流数据实时发现热点事件。

分析模块 804，用于对含有所述热点事件的微博进行情感倾向分析。

确定模块 806，用于根据所述情感倾向分析确定相应的情感指数值。

生成模块 808，用于根据情感指数值生成时间序列的情感指数走势图。

如图 9 所示，在其中一个实施例中，发现模块 702 包括：

获取模块 802A，用于实时获取微博流数据，将所述微博流数据向量化。

记录模块 802B，用于监控向量化后的微博流数据，记录每个特征词出现的频率和次数。

热点事件确定模块 802C，用于根据每个特征词出现的频率和次数确定当前的热点事件。

如图 10 所示，在其中一个实施例中，分析模块 704 包括：

抽取模块 804A，用于抽取获取到的含有热点事件的每条微博的 LDA 主题特征和 word2vec 词向量特征。

输出模块 804B，用于将所述 LDA 主题特征和 word2vec 词向量特征代入情感倾向分析模型，输出每条微博的情感倾向。

如图 11 所示，在其中一个实施例中，确定模块 706 包括：

统计模块 806A，用于根据情感倾向分析将含有热点事件的微博按照情感类别进行分类，分别统计每一类情感对应的微博数。

情感指数值确定模块 806B, 用于根据统计出的所述每一类情感对应的微博数确定所述热点事件对应的情感指数值。

在其中一个实施例中, 上述微博数据处理装置还包括: 股市预测模块, 用于根据所述情感指数走势图确定影响股市数据的情感因子, 将所述情感因子和其他股市数据一起作为股市预测模型的预测因子进行相应的股市预测。

上述微博数据处理装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。其中, 网络接口可以是以网卡或无线网卡等。上述各模块可以硬件形式内嵌于或独立于服务器中的处理器中, 也可以以软件形式存储于服务器中的存储器中, 以便于处理器调用执行以上各个模块对应的操作。该处理器可以为中央处理单元 (CPU)、微处理器、单片机等。

上述控制承保处理的装置可以实现为一种计算机可读指令的形式, 计算机可读指令可以在如图 2 或 3 所示的终端或服务器上运行。

在其中一个实施例中, 提出了一种计算机设备, 计算机设备的内部结构可对应于如图 2 或 3 所示的结构, 即该计算机设备既可以是服务器也可以是终端, 其包括一系列存储于存储器上的计算机可读指令, 当该计算机可读指令被处理器执行时, 可以实现本申请各实施例提出的微博数据处理方法。本申请实施例提出了一种计算机设备。该计算机设备包括一系列存储于存储器上的计算机可读指令, 当该计算机可读指令被处理器执行时, 可以实现本申请各实施例提出的微博数据处理方法。计算机设备包括存储器、处理器及存储于所述存储器上并可在所述处理器上运行的计算机可读指令, 所述处理器执行所述计算机可读指令时实现以下步骤: 通过监控微博流数据实时发现热点事件; 对含有所述热点事件的微博进行情感倾向分析; 根据所述情感倾向分析确定相应的情感指数值; 根据情感指数值生成时间序列的情感指数走势图。

在其中一个实施例中, 所述处理器所述执行的所述通过监控微博流数据实时发现热点事件包括: 实时获取微博流数据, 将所述微博流数据向量化; 监控向量化后的微博流数据, 记录每个特征词出现的频率和次数; 根据每个特征词出现的频率和次数确定当前的热点事件。

在其中一个实施例中, 所述处理器所述执行的所述将含有所述热点事件的微博进行情感倾向分析包括: 抽取获取到的含有热点事件的每条微博的 LDA 主题特征和 word2vec 词向量特征; 将所述 LDA 主题特征和 word2vec 词向量特征代入情感倾向分析模型, 输出每条微博的情感倾向。

在其中一个实施例中, 所述处理器所述执行的所述根据所述情感倾向分析确定相应的情感指数值包括: 按照情感类别, 分别统计每一类情感对应的微博数; 根据统计出的所述每一类情感对应的微博数确定所述热点事件对应

的情感指数值。

在其中一个实施例中，所述处理器还用于执行以下步骤：根据所述情感指数走势图确定影响股市数据的情感因子，将所述情感因子和其他股市数据一起作为股市预测模型的预测因子进行相应的股市预测。

5

在其中一个实施例中，提出了一个或多个存储有计算机可执行指令的非易失性计算机可读存储介质，所述计算机可执行指令被一个或多个处理器执行时，使得所述一个或多个处理器执行以下步骤：通过监控微博流数据实时发现热点事件；对含有所述热点事件的微博进行情感倾向分析；根据所述情感倾向分析确定相应的情感指数值；根据情感指数值生成时间序列的情感指数走势图。

在其中一个实施例中，所述处理器所述执行的所述通过监控微博流数据实时发现热点事件包括：实时获取微博流数据，将所述微博流数据向量化；监控向量化后的微博流数据，记录每个特征词出现的频率和次数；根据每个特征词出现的频率和次数确定当前的热点事件。

在其中一个实施例中，所述处理器所述执行的所述将含有所述热点事件的微博进行情感倾向分析包括：抽取获取到的含有热点事件的每条微博的 LDA 主题特征和 word2vec 词向量特征；将所述 LDA 主题特征和 word2vec 词向量特征代入情感倾向分析模型，输出每条微博的情感倾向。

在其中一个实施例中，所述处理器所述执行的所述根据所述情感倾向分析确定相应的情感指数值包括：按照情感类别，分别统计每一类情感对应的微博数；根据统计出的所述每一类情感对应的微博数确定所述热点事件对应的情感指数值。

在其中一个实施例中，所述处理器还用于执行以下步骤：根据所述情感指数走势图确定影响股市数据的情感因子，将所述情感因子和其他股市数据一起作为股市预测模型的预测因子进行相应的股市预测。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机程序来指令相关的硬件来完成，该计算机程序可存储于一计算机可读取存储介质中，该程序在执行时，可包括如上述各方法的实施例的流程。其中，前述的存储介质可为磁碟、光盘、只读存储记忆体（Read-Only Memory, ROM）等非易失性存储介质，或随机存储记忆体（Random Access Memory, RAM）等。

以上所述实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的各个技术特征所有可能的组合都进行描述，然而，只要这

35

些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

以上所述实施例仅表达了本申请的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对发明专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本申请构思的前提下，还可以做出若干变形和改进，这些都属于本申请的保护范围。因此，本申请专利的保护范围应以所附权利要求为准。

权利要求书

1、一种微博数据处理方法，包括：

通过监控微博流数据实时发现热点事件；

对含有所述热点事件的微博进行情感倾向分析；

根据所述情感倾向分析确定相应的情感指数值；及

5 根据所述情感指数值生成时间序列的情感指数走势图。

2、根据权利要求1所述的方法，所述通过监控微博流数据实时发现热点事件，包括：

实时获取微博流数据，将所述微博流数据向量化；

监控向量化后的微博流数据，记录每个特征词出现的频率和次数；及

10 根据每个特征词出现的频率和次数确定当前的热点事件。

3、根据权利要求1所述的方法，其特征在于，所述将含有所述热点事件的微博进行情感倾向分析，包括：

抽取获取到的含有热点事件的每条微博的 LDA 主题特征和 word2vec 词向量特征；及

15 将所述 LDA 主题特征和 word2vec 词向量特征代入情感倾向分析模型，输出每条微博的情感倾向。

4、根据权利要求1所述的方法，其特征在于，所述根据所述情感倾向分析确定相应的情感指数值，包括：

根据所述情感倾向分析将含有热点事件的微博按照情感类别进行分类，

20 分别统计每一类情感对应的微博数；及

根据统计出的所述每一类情感对应的微博数确定所述热点事件对应的情感指数值。

5、根据权利要求1所述的方法，其特征在于，在所述根据所述情感指数值生成时间序列的情感指数走势图之后，还包括：

25 根据所述情感指数走势图确定影响股市数据的情感因子；及

将所述情感因子和其他股市数据一起作为股市预测模型的预测因子进行相应的股市预测。

6、一种微博数据处理装置，包括：

发现模块，用于通过监控微博流数据实时发现热点事件；

30 分析模块，用于对含有所述热点事件的微博进行情感倾向分析；

确定模块，用于根据所述情感倾向分析确定相应的情感指数值；及

生成模块，用于根据所述情感指数值生成时间序列的情感指数走势图。

7、根据权利要求6所述的装置，所述发现模块包括：

获取模块，用于实时获取微博流数据，将所述微博流数据向量化；

记录模块, 用于监控向量化后的微博流数据, 记录每个特征词出现的频率和次数; 及

热点事件确定模块, 用于根据每个特征词出现的频率和次数确定当前的热点事件。

5 8、根据权利要求 6 所述的装置, 其特征在于, 所述分析模块包括:

抽取模块, 用于抽取获取到的含有热点事件的每条微博的 LDA 主题特征和 word2vec 词向量特征; 及

输出模块, 用于将所述 LDA 主题特征和 word2vec 词向量特征代入情感倾向分析模型, 输出每条微博的情感倾向。

10 9、根据权利要求 6 所述的装置, 其特征在于, 所述确定模块包括:

统计模块, 用于根据情感倾向分析将含有热点事件的微博按照情感类别进行分类, 分别统计每一类情感对应的微博数; 及

情感指数值确定模块, 用于根据统计出的所述每一类情感对应的微博数确定所述热点事件对应的情感指数值。

15 10、根据权利要求 6 所述的装置, 其特征在于, 所述装置还包括:

股市预测模块, 用于根据所述情感指数走势图确定影响股市数据的情感因子, 将所述情感因子和其他股市数据一起作为股市预测模型的预测因子进行相应的股市预测。

20 11、一种计算机设备, 包括存储器及一个或多个处理器, 所述存储器中存储有计算机可读指令, 所述计算机可读指令被所述一个或多个处理器执行时, 使得所述一个或多个处理器执行以下步骤:

通过监控微博流数据实时发现热点事件;

对含有所述热点事件的微博进行情感倾向分析;

根据所述情感倾向分析确定相应的情感指数值; 及

25 根据所述情感指数值生成时间序列的情感指数走势图。

12、根据权利要求 11 所述的计算机设备, 其特征在于, 所述通过监控微博流数据实时发现热点事件, 包括:

实时获取微博流数据, 将所述微博流数据向量化;

监控向量化后的微博流数据, 记录每个特征词出现的频率和次数; 及

30 根据每个特征词出现的频率和次数确定当前的热点事件。

13、根据权利要求 11 所述的计算机设备, 其特征在于, 所述将含有所述热点事件的微博进行情感倾向分析, 包括:

抽取获取到的含有热点事件的每条微博的 LDA 主题特征和 word2vec 词向量特征; 及

35 将所述 LDA 主题特征和 word2vec 词向量特征代入情感倾向分析模型,

输出每条微博的情感倾向。

14、根据权利要求 11 所述的计算机设备，其特征在于，所述根据所述情感倾向分析确定相应的情感指数值，包括：

根据所述情感倾向分析将含有热点事件的微博按照情感类别进行分类，
5 分别统计每一类情感对应的微博数；及

根据统计出的所述每一类情感对应的微博数确定所述热点事件对应的情感指数值。

15、根据权利要求 11 所述的计算机设备，其特征在于，在所述根据所述情感指数值生成时间序列的情感指数走势图之后，所述处理器还用于执行以下步骤：
10

根据所述情感指数走势图确定影响股市数据的情感因子；及

将所述情感因子和其他股市数据一起作为股市预测模型的预测因子进行相应的股市预测。

16、一个或多个存储有计算机可读指令的非易失性计算机可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行以下步骤：
15

通过监控微博流数据实时发现热点事件；

对含有所述热点事件的微博进行情感倾向分析；

20 根据所述情感倾向分析确定相应的情感指数值；及

根据所述情感指数值生成时间序列的情感指数走势图。

17、根据权利要求 16 所述的存储介质，其特征在于，所述通过监控微博流数据实时发现热点事件，包括：

实时获取微博流数据，将所述微博流数据向量化；

25 监控向量化后的微博流数据，记录每个特征词出现的频率和次数；及

根据每个特征词出现的频率和次数确定当前的热点事件。

18、根据权利要求 16 所述的存储介质，其特征在于，所述将含有所述热点事件的微博进行情感倾向分析，包括：

抽取获取到的含有热点事件的每条微博的 LDA 主题特征和 word2vec 词
30 向量特征；及

将所述 LDA 主题特征和 word2vec 词向量特征代入情感倾向分析模型，输出每条微博的情感倾向。

19、根据权利要求 16 所述的存储介质，其特征在于，所述根据所述情感倾向分析确定相应的情感指数值，包括：

35 根据所述情感倾向分析将含有热点事件的微博按照情感类别进行分类，

分别统计每一类情感对应的微博数；及

根据统计出的所述每一类情感对应的微博数确定所述热点事件对应的情感指数值。

20、根据权利要求 16 所述的存储介质，其特征在于，在所述根据所述情感指数值生成时间序列的情感指数走势图之后，所述处理器还用于执行以下步骤：

根据所述情感指数走势图确定影响股市数据的情感因子；及

将所述情感因子和其他股市数据一起作为股市预测模型的预测因子进行相应的股市预测。

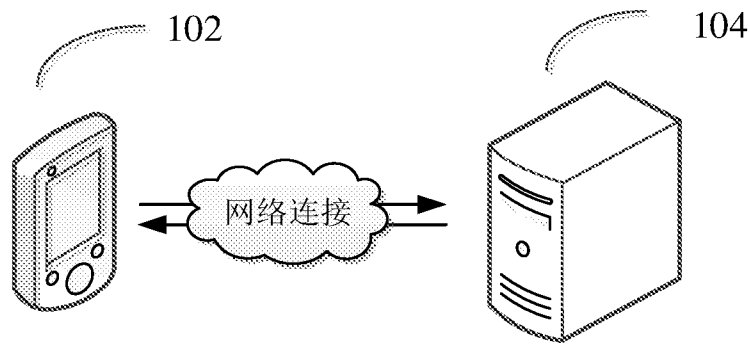


图 1

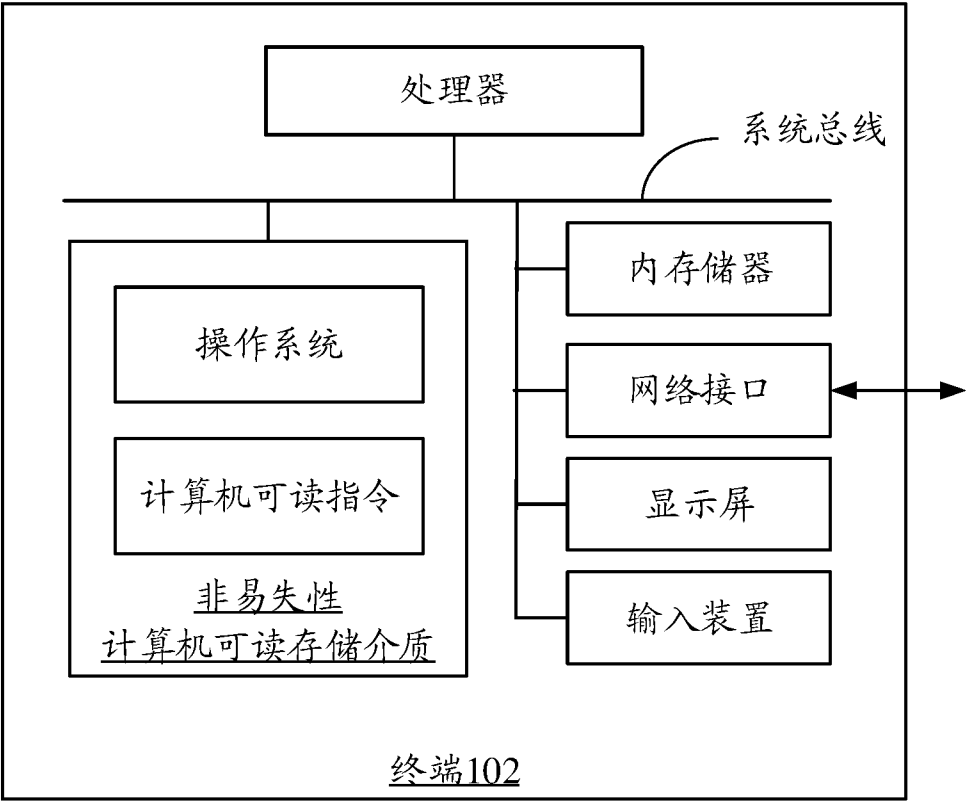


图 2

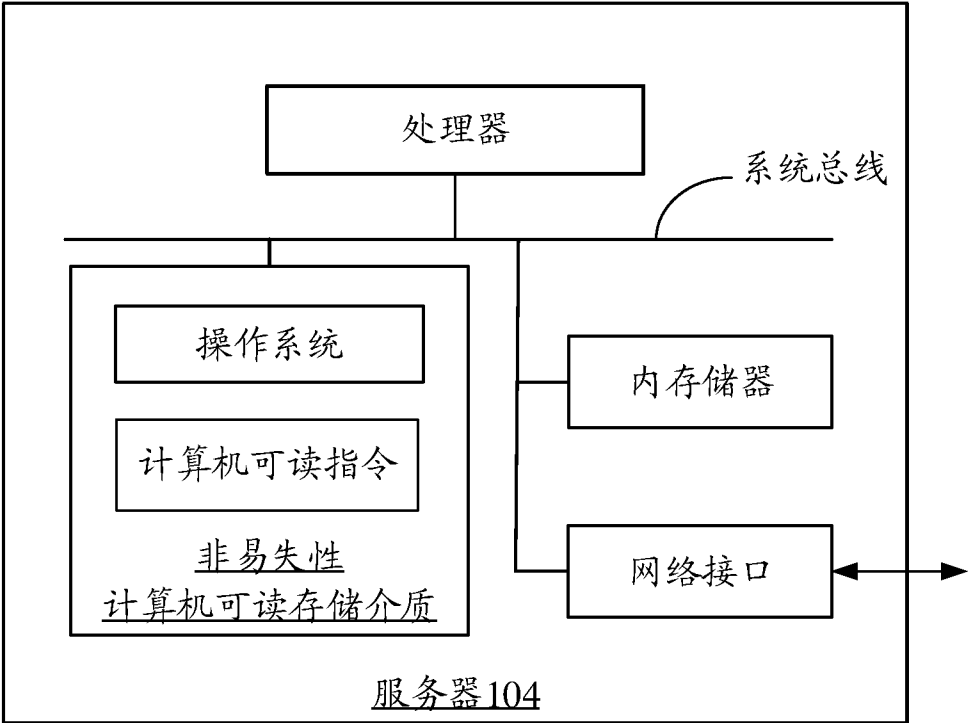


图 3

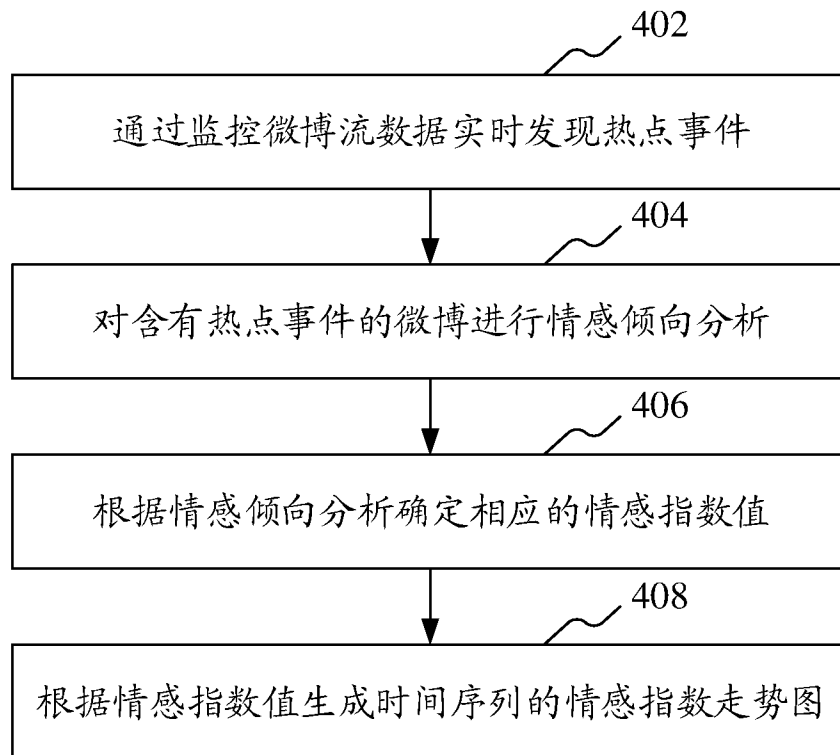


图 4

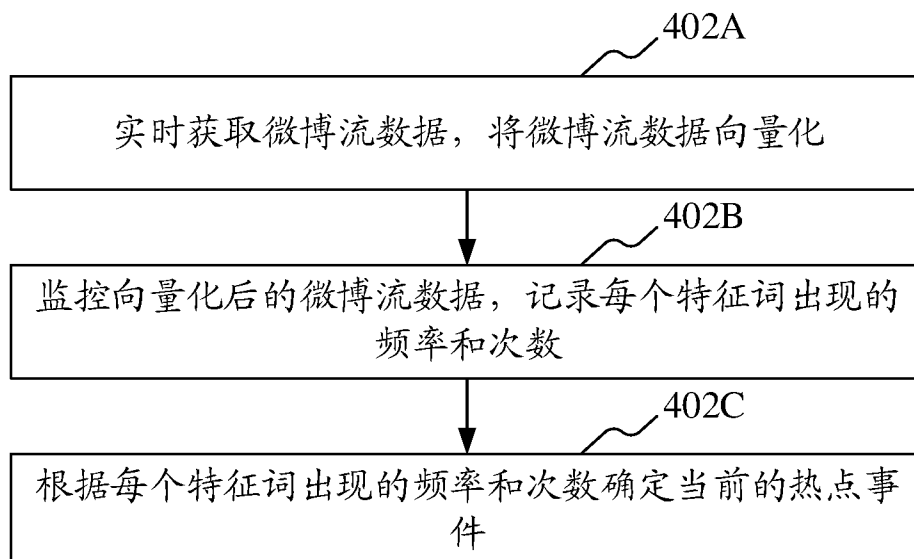


图 5

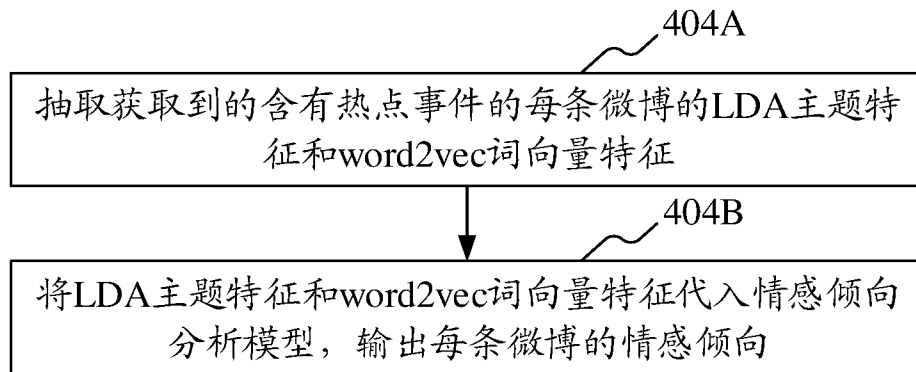


图 6

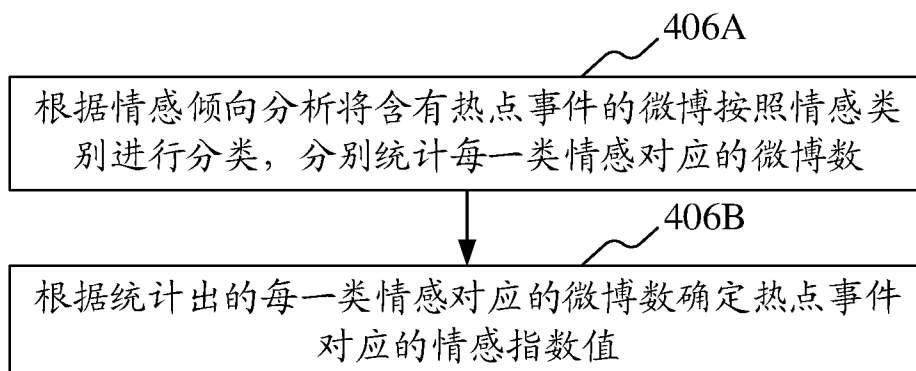


图 7

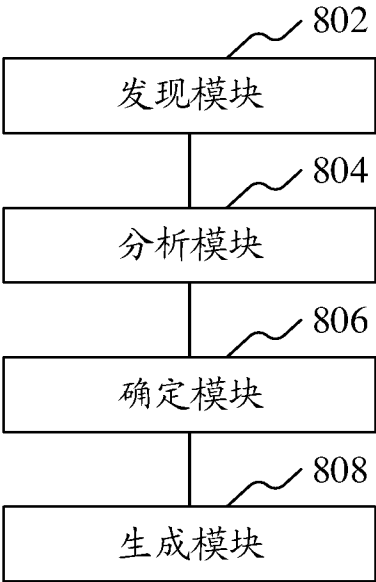


图 8

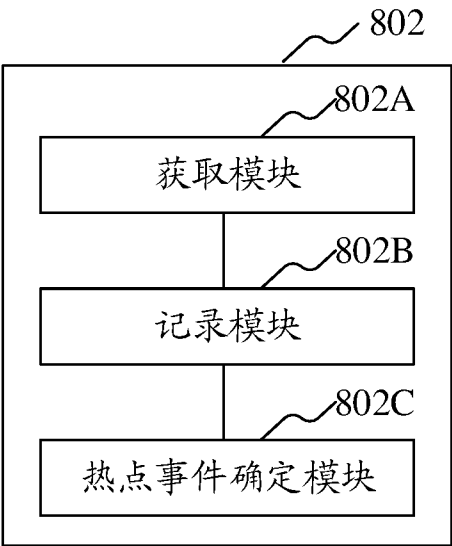


图 9

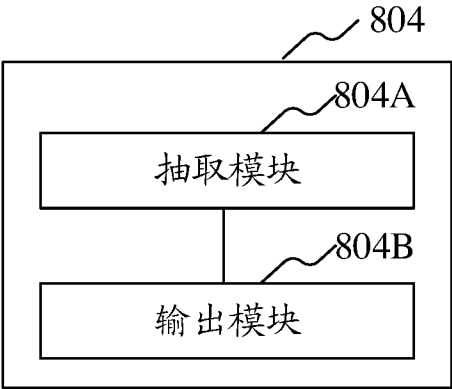


图 10

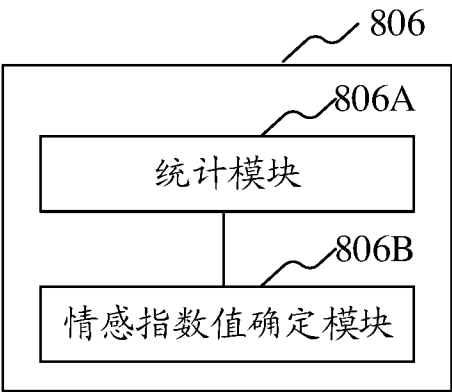


图 11

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2018/081697

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/27 (2006.01) i; G06Q 50/00 (2012.01) 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; G06Q

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)

CNABS; CNTXT; DWPI; SIPOABS; CNKI: 微博, 监控, 热点, 事件, 情感, 倾向, 指数, 走势, 序列, 主题, 模型, 向量, 频率, micro-blog, monitor, hot, event, polarity, trend, exponent, sensibility, topic, model, vector, frequency

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 107797983 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 13 March 2018 (13.03.2018), claims 1-10	1-20
X	CN 105740228 A (YUNNAN UNIVERSITY) 06 July 2016 (06.07.2016), description, paragraphs [0062]-[0089]	1, 4-6, 9-11, 14-16, 19, 20
Y	CN 105740228 A (YUNNAN UNIVERSITY) 06 July 2016 (06.07.2016), description, paragraphs [0062]-[0089]	2, 3, 7, 8, 12, 13, 17, 18
Y	CN 105589941 A (BEIJING PERCENT CORPORATION) 18 May 2016 (18.05.2016), description, paragraphs [0068]-[0096]	2, 3, 7, 8, 12, 13, 17, 18
A	CN 104537097 A (CHENGDU BLTSAFE INFORMATION TECHNOLOGY CO., LTD.) 22 April 2015 (22.04.2015), entire document	1-20

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	“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
“A” document defining the general state of the art which is not considered to be of particular relevance	“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
“E” earlier application or patent but published on or after the international filing date	“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
“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)	“&” document member of the same patent family
“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	

Date of the actual completion of the international search 20 April 2018	Date of mailing of the international search report 07 June 2018
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer ZHANG, Jianfeng Telephone No. (86-512) 88995721

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2018/081697

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 107797983 A	13 March 2018	None	
CN 105740228 A	06 July 2016	None	
CN 105589941 A	18 May 2016	None	
CN 104537097 A	22 April 2015	CN 104537097 B	11 August 2017

国际检索报告

国际申请号

PCT/CN2018/081697

A. 主题的分类 G06F 17/27(2006.01)i; G06Q 50/00(2012.01)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; G06Q 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)) CNABS;CNTXT;DWPI;SIPOABS;CNKI; 微博, 监控, 热点, 事件, 情感, 倾向, 指数, 走势, 序列, 主题, 模型, 向量, 频率, micro-blog, monitor, hot, event, polarity, trend, exponent, sensibility, topic, model, vector, frequency																				
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 107797983 A (平安科技深圳有限公司) 2018年 3月 13日 (2018 - 03 - 13) 权利要求1-10</td> <td>1-20</td> </tr> <tr> <td>X</td> <td>CN 105740228 A (云南大学) 2016年 7月 6日 (2016 - 07 - 06) 说明书第[0062]-[0089]段</td> <td>1、4-6、9-11、 14-16、19、20</td> </tr> <tr> <td>Y</td> <td>CN 105740228 A (云南大学) 2016年 7月 6日 (2016 - 07 - 06) 说明书第[0062]-[0089]段</td> <td>2、3、7、8、 12、13、17、18</td> </tr> <tr> <td>Y</td> <td>CN 105589941 A (北京百分点信息科技有限公司) 2016年 5月 18日 (2016 - 05 - 18) 说明书第[0068]-[0096]段</td> <td>2、3、7、8、 12、13、17、18</td> </tr> <tr> <td>A</td> <td>CN 104537097 A (成都布林特信息技术有限公司) 2015年 4月 22日 (2015 - 04 - 22) 全文</td> <td>1-20</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 107797983 A (平安科技深圳有限公司) 2018年 3月 13日 (2018 - 03 - 13) 权利要求1-10	1-20	X	CN 105740228 A (云南大学) 2016年 7月 6日 (2016 - 07 - 06) 说明书第[0062]-[0089]段	1、4-6、9-11、 14-16、19、20	Y	CN 105740228 A (云南大学) 2016年 7月 6日 (2016 - 07 - 06) 说明书第[0062]-[0089]段	2、3、7、8、 12、13、17、18	Y	CN 105589941 A (北京百分点信息科技有限公司) 2016年 5月 18日 (2016 - 05 - 18) 说明书第[0068]-[0096]段	2、3、7、8、 12、13、17、18	A	CN 104537097 A (成都布林特信息技术有限公司) 2015年 4月 22日 (2015 - 04 - 22) 全文	1-20
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
PX	CN 107797983 A (平安科技深圳有限公司) 2018年 3月 13日 (2018 - 03 - 13) 权利要求1-10	1-20																		
X	CN 105740228 A (云南大学) 2016年 7月 6日 (2016 - 07 - 06) 说明书第[0062]-[0089]段	1、4-6、9-11、 14-16、19、20																		
Y	CN 105740228 A (云南大学) 2016年 7月 6日 (2016 - 07 - 06) 说明书第[0062]-[0089]段	2、3、7、8、 12、13、17、18																		
Y	CN 105589941 A (北京百分点信息科技有限公司) 2016年 5月 18日 (2016 - 05 - 18) 说明书第[0068]-[0096]段	2、3、7、8、 12、13、17、18																		
A	CN 104537097 A (成都布林特信息技术有限公司) 2015年 4月 22日 (2015 - 04 - 22) 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 2018年 4月 20日		国际检索报告邮寄日期 2018年 6月 7日																		
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局(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 张剑峰 电话号码 (86-512) 88995721																		

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/081697

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	107797983	A	2018年 3月 13日	无	
CN	105740228	A	2016年 7月 6日	无	
CN	105589941	A	2016年 5月 18日	无	
CN	104537097	A	2015年 4月 22日	CN 104537097 B	2017年 8月 11日