

(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 2020/082560 A1

(51) 国际专利分类号:
G06F 17/27 (2006.01)

(21) 国际申请号: PCT/CN2018/122813

(22) 国际申请日: 2018 年 12 月 21 日 (21.12.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201811254895.6 2018年10月25日 (25.10.2018) 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: 金戈(JIN, Ge); 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)。 徐亮(XU, Liang); 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)。 肖京(XIAO, Jing); 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: 深圳市世纪恒程知识产权代理事务所(CENFO INTELLECTUAL PROPERTY AGENCY); 中国广东省深圳市南山区粤海街道高新技术产业园北区松坪山路 3 号奥特讯电力大厦 201, Guangdong 518057 (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,

(54) **Title:** METHOD, APPARATUS AND DEVICE FOR EXTRACTING TEXT KEYWORD, AS WELL AS COMPUTER READ-ABLE STORAGE MEDIUM

(54) 发明名称: 文本关键词提取方法、装置、设备及计算机可读存储介质

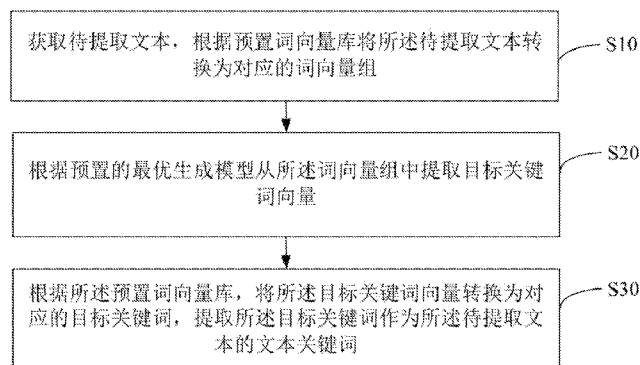


图 2

- S10 Obtaining a text to be extracted, and converting the text to be extracted into a corresponding word vector group according to a preset word vector library
- S20 Extracting a target keyword vector from the word vector group according to a preset optimal generation model
- S30 Converting the target keyword vector into a corresponding target keyword according to the preset word vector library, and extracting the target keyword as a text keyword for the text to be extracted

(57) **Abstract:** Provided by the present application are a method, apparatus and device for extracting text keyword on the basis of a classification model and a prediction model, as well as a computer readable storage medium. The method comprises: obtaining a text to be extracted, and converting same into a corresponding word vector group according to a preset word vector library; extracting a target keyword vector from the word vector group according to a preset optimal generation model; converting the target keyword vector into a corresponding target keyword according to the preset word vector library, and extracting the target keyword as a text keyword

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))。

for the text to be extracted.

(57) 摘要: 本申请提供一种基于分类模型及预测模型的文本关键词提取方法、装置、设备及计算机可读存储介质, 该方法包括: 获取待提取文本, 根据预置词向量库将所述待提取文本转换为对应的词向量组; 根据预置的最优生成模型从所述词向量组中提取目标关键词向量; 根据所述预置词向量库, 将所述目标关键词向量转换为对应的目标关键词, 提取所述目标关键词作为所述待提取文本的文本关键词。

文本关键词提取方法、装置、设备及计算机可读存储介质

本申请要求于2018年10月25日提交中国专利局、申请号为201811254895.6、发明名称为“文本关键词提取方法、装置、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。

技术领域

本申请涉及关键词提取技术领域，尤其涉及一种文本关键词提取方法、装置、设备及计算机可读存储介质。

背景技术

关键词抽取在文本处理的许多领域中应用广泛，如：文本聚类、文本摘要和信息检索。在当下大数据时代，关键词抽取更是在NLP领域扮演着重要角色，为情感分析、语义分析、知识图谱等热点问题提供了基石。目前该领域主流代表的方法有基于隐含主题模型的关键词抽取（LDA）、基于TF-IDF词频统计的关键词抽取和基于词图模型的关键词抽取（TextRank）。

本申请提出一种新的关键词抽取方法。

发明内容

本申请的主要目的在于提供一种文本关键词提取方法，旨在解决现有文本关键词提取效率不高的技术问题。

为实现上述目的，本申请提供一种文本关键词提取方法，其中，所述文本关键词提取方法包括以下步骤：

获取待提取文本，根据预置词向量库将所述待提取文本转换为对应的词向量组；

根据预置的最优生成模型从所述词向量组中提取目标关键词向量；

根据所述预置词向量库，将所述目标关键词向量转换为对应的目标关键词，提取所述目标关键词作为所述待提取文本的文本关键词。

此外，为实现上述目的，本申请还提供一种文本关键词提取装置，所述文本关键词提取装置包括：

第一向量转换模块，设置为获取待提取文本，根据预置词向量库将所述待提取文本转换为对应的词向量组；关键词生成模块，设置为根据预置的最优生成模型从所述词向量组中提取目标关键词向量；第二向量转换模块，设置为根据所述预置词向量库，将所述目标关键词向量转换为对应的目标关键词，提取所述目标关键词作为所述待提取文本的文本关键词。

此外，为实现上述目的，本申请还提供一种文本关键词提取设备，所述文本关键词提取设备包括处理器、存储器、以及存储在所述存储器上并可被所述处理器执行的计算机可读指令，其中所述计算机可读指令被所述处理器执行时，实现如上述的文本关键词提取方法的步骤。

此外，为实现上述目的，本申请还提供一种计算机可读存储介质，所述计算机可读存储介质上存储有计算机可读指令，其中所述计算机可读指令被处理器执行时，实现如上述的文本关键词提取方法的步骤。

本申请实施例通过获取待提取文本，根据预置词向量库将所述待提取文本转换为对应的词向量组；根据预置的最优生成模型从所述词向量组中提取目标关键词向量，即通过将待提取文本转换为向量化数据，并将其作为生成模型的输入，可减少模型运算量，提高文本关键词提取效率；根据所述预置词向量库，将所述目标关键词向量转换为对应的目标关键词，提取所述目标关键词作为所述待提取文本的文本关键词，实现对待提取文本的文本关键词的提取。

附图说明

图1是本申请实施例方案涉及的硬件运行环境的文本关键词提取设备结构示意图；

图2为本申请文本关键词提取方法第一实施例的流程示意图；

图3为本申请文本关键词提取方法第二实施例的流程示意图；

图4为本申请文本关键词提取装置第一实施例的功能模块示意图。

本申请目的的实现、功能特点及优点将结合实施例，参照附图做进一步说明。

具体实施方式

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。

请参见图 1，图 1 为本申请所提供的文本关键词提取设备的硬件结构示意图。

所述文本关键词提取设备可以是 PC，也可以是智能手机、平板电脑、便携计算机、台式计算机等具有显示功能的设备，可选地，所述文本关键词提取设备可以是服务器设备，存在文本关键词提取的后端管理系统，用户通过所述后端管理系统对文本关键词提取设备进行管理。

所述文本关键词提取设备可以包括：处理器 101 以及存储器 201 等部件。在所述文本关键词提取设备中，所述处理器 101 与所述存储器 201 连接，所述存储器 201 上存储有计算机可读指令，处理器 101 可以调用存储器 201 中存储的计算机可读指令，并实现如下述文本关键词提取方法各实施例的步骤。

所述存储器 201，可用于存储软件程序以及各种数据。存储器 201 可主要包括存储程序区和存储数据区，其中，存储程序区可存储操作系统、至少一个功能所需的应用程序（比如计算机可读指令）等；存储数据区可包括数据库，例如关联网络的节点的节点信息等。此外，存储器 201 可以包括高速随机存取存储器，还可以包括非易失性存储器，例如至少一个磁盘存储器件、闪存器件、或其他易失性固态存储器件。

处理器 101，是文本关键词提取设备的控制中心，利用各种接口和线路连接整个文本关键词提取设备的各个部分，通过运行或执行存储在存储器 201 内的软件程序和/或模块，以及调用存储在存储器 201 内的数据，执行文本关键词提取设备的各种功能和处理数据，从而对文本关键词提取设备进行整体监控。处理器 101 可包括一个或多个处理单元；可选地，处理器 101 可集成应用处理器和调制解调处理器，其中，应用处理器主要处理操作系统、用户界面和应用程序等，调制解调处理器主要处理无线通信。可以理解的是，上述调制解调处理器也可以不集成到处理器 101 中。

本领域技术人员可以理解,图 1 中示出的文本关键词提取设备结构并不构成对文本关键词提取设备的限定,可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件布置。

基于上述硬件结构,提出本申请方法各个实施例,下文中的“提取设备”为文本关键词提取设备的简称。

本申请提供一种文本关键词提取方法。

参照图 2,图 2 为本申请文本关键词提取方法第一实施例的流程示意图。

本实施例中,所述文本关键词提取方法包括以下步骤:

步骤 S10,获取待提取文本,根据预置词向量库将所述待提取文本转换为对应的词向量组;

待提取文本,指待进行关键词提取的文本数据,待提取文本是多个字符按照特定语义顺序构成的字符串,待提取文本可以为一篇文章或一段文字。

对于获取待提取文本,具体地,提取设备可提供输入界面,以获取用户通过输入界面输入需要进行关键词提取的文本数据,将获取的文本数据作为待提取文本。提取设备也可接收其他设备发送的需要进行关键词提取的文本数据,将接收到的文本数据作为待提取文本。提取设备还可提供可选文本列表,以获取用户从可选文本列表中选择待提取文本。

预置词向量库,存储有预置的语料词语与对应的词向量,词向量指词语被映射到实数的向量,比如,文本形式的“话筒”表示为数学形式“[00010000000..]”,此时,“[00010000000..]”即为“话筒”的词向量。可以理解,这里不限定将文本形式的语料词语转换为何种表示的词向量,只要能够将文本形式的语料词语进行数学化表示即可。在执行步骤 10 中的根据预置词向量库将所述待提取文本转换为对应的词向量组这一步骤之前,需建立预置词向量库,具体地,包括:

步骤 S11,获取待转换的语料文本,将所述语料文本进行分词,获得分词后的待转换词组;

待转换的语料文本,即待进行向量转换的语料文本,提取设备可

直接从互联网上拉取语料文本，比如新闻或者文章等，语料文本也可从语料库获得。

待转换词组指将语料文本进行分词后，得到的组成语料文本的词组，本实施例中的“词组”指多个词，待转换词组包含多个待转换词。

分词是将连续的字符序列切分成多个单独的字符或者字符序列的操作。提取设备可再对语料文本按照标点符号进行语句分割，得到若干语句，再对各语句进行分词，得到组成语料文本的各词。提取设备可采用预设的分词方式对待提取文本进行分词处理，得到多个字符（无顺序词组）或者字符序列（具有特定排列顺序的词组，如与语料文本顺序相同的词组），提取设备可再根据词汇表确定分词处理后获得的待转换词组中各待转换词的词性，还可统计各词相应的词长，其中，词性是反映词的内容所属类型的数据，词性包括形容词、介词、谓词和名词等 12 种词性。词长是词所包含的字符的数量。预设的分词方式可以是基于字符匹配、基于语义理解或者基于统计的分词方式。提取设备可设置分词得到的各待转换词的词长阈值，使得分词得到的各待转换词的词长均不超过词长阈值。

举例说明，提取设备对词序列“我/今天/很/高兴”确定各词相应的词性后得到“黎明 a/今天 b/很 c/高兴 d”，其中 a 表示人名，b 表示状语，c 表示副词，d 表示谓语等。提取设备对词序列“我/今天/很/高兴”确定各词相应的词长后得到“我 1/今天 2/很 1/高兴 2”，其中数字表示词长。

步骤 S12，将所述待转换词组中各待转换词转换成对应的词向量，将各所述待转换词与对应的词向量关联存储在预置词向量库。

提取设备根据待转换词组中各待转换词的内容、词性与词长，将该词向量化，得到该词相应的词向量，从而得到待转换词相应的词向量。其中，提取设备可以利用机器学习模型将词转化为词向量，机器学习模型可以为 word2vec 模型等。

具体地，提取设备可以预先设置编码方式，通过该编码方式将词性编码为词性向量，将词长编码为词长向量，然后将内容向量、词性向量和词长向量进行组合得到相应词所对应的词向量，得到词向量序列。其中，编码方式比如 One-Hot 编码或者整数编码等。内容向量、

词性向量和词长向量组合的方式可以是直接拼接或者是通过连接向量间接拼接。可以理解，在拼接过程中对内容向量、词性向量和词长向量的拼接顺序不作限定。

将各待转换词与对应的词向量关联存储在预置词向量库，可根据待转换词在预置词向量库查找到对应的词向量，也可以根据词向量查找到对应的待转换词。

在本实施例中，通过获取待转换的语料文本，将所述语料文本进行分词，获得分词后的待转换词组；将所述待转换词组中各待转换词转换成对应的词向量，将各所述待转换词与对应的词向量关联存储在预置词向量库，为后续将待提取文本转换为向量化数据并将其作为生成模型的输入以减少模型运算量，提高文本关键词提取效率提供保障。

提取设备获取待提取文本后，对待提取文本进行分词，得到组成待提取文本的各词，再将各词向量化，得到各词各自对应的词向量，从而得到对应的词向量组，其中，待提取文本的分词方式与待转换的语料文本的分词方式一致，相关分词方式已在前文解释，此处不赘述。可通过查询预置词向量库，获取组成待提取文本的各词对应的词向量，将待提取文本的各词转换为对应的词向量；也可以通过与待转换词相同的词向量转化方式对待提取文本进行向量化，此处不赘述。

步骤 S20，根据预置的最优生成模型从所述词向量组中提取目标关键词向量；

预置的最优生成模型指包含最优模型参数的生成模型，即训练好的生成模型，在生成模型的训练阶段，以生成模型和判别模型共同组成生成式对抗网络，在生成式对抗网络实现生成模型和判别模型的模型训练。生成式对抗网络要解决的问题是如何从训练样本中学习出新样本，常见的应用是根据真实图片生成新图片。

本实施例中的生成模型为经过训练后具有关键词向量提取功能的机器学习模型，判别模型为经过训练后具有将真实关键词向量与生成模型提取的预测关键词向量进行区分的判别功能的机器学习模型，机器学习模型可通过样本学习具备前述提取功能或判别功能，机器学习模型可采用神经网络模型、支持向量机或者逻辑回归模型等。

本实施例中，提取设备将词向量组输入最优生成模型，利用最优生成模型中隐层的模型参数对词向量组进行运算，得到运算结果，即生成最优生成模型提取的目标关键词向量，其中，利用隐层的模型参数对词向量进行的运算可以是线性变换、非线性变换或者卷积变换等。

一实施方式中，提取设备可以按照词向量组中各词向量的先后顺序，通过最优生成模型的隐层依次地对词向量组中各词向量进行运算，循环地将前次的运算结果和当前词向量作为当次运算的输入值，进行当次运算，直至最后一次运算为止。可以理解，由于第一次处理时不存在前次运算，所以第一次运算时的输入值为第一个词向量。举例说明，待提取文本对应的词向量组为 X_1 、 X_2 、 X_3 、 X_4 、 X_5 。最优生成模型的隐层可按照 X_1 - X_5 的顺序或者 X_5 - X_1 的顺序依次地对各词向量进行运算。比如，先将 X_1 作为输入进行运算，得到运算结果 Y_1 ，然后将 Y_1 和 X_2 作为输入进行运算，得到运算结果 Y_2 ，然后将 Y_2 和 X_3 作为输入进行运算，得到运算结果 Y_3 ，依次类推，直至得到最后一个词向量 X_5 相应的运算结果 Y_5 。

步骤 S30，根据所述预置词向量库，将所述目标关键词向量转换为对应的目标关键词，提取所述目标关键词作为所述待提取文本的文本关键词。

目标关键词向量即最优生成模型从输入的词向量组中提取的/预测的待提取文本的关键词向量，目标关键词即最优生成模型提取的/预测的待提取文本的关键词。

一实施方式中，通过查询预置词向量库，获取组成待提取文本的各词对应的词向量，将待提取文本的各词转换为对应的词向量，此时，根据所述预置词向量库，将所述目标关键词向量转换为对应的目标关键词这一步骤具体包括：查询预置词向量库，基于预置词向量库存储的词与对应向量的关联关系，从预置词向量库中获取目标关键词向量对应的目标关键词，完成目标关键词向量的转化。

另一实施方式中，通过与待转换词相同的词向量转化方式对待提取文本进行向量化，其中，待转换词与待提取文本转化后的词向量表示方式为分布式表示 (Distributed representation) 词向量表示方法，

此时，根据所述预置词向量库，将所述目标关键词向量转换为对应的目标关键词这一步骤具体包括：

步骤 S31，遍历所述预置词向量库中所有的预置词向量，分别计算各所述预置词向量与所述目标关键词向量的欧氏距离；

Distributed representation 词向量表示方法表示的词向量使得相关或者相似的词，其数学含义表现为向量距离的接近。例如，“麦克”和“话筒”的距离会远远小于“麦克”和“天气”。

Distributed representation 词向量表示方法的原理包括：通过训练将某种语言的特定文本中的每一个词映射成一个固定长度的向量，将所有这些向量放在一起形成一个词向量空间，而每一向量则为该空间中的一个点，在这个空间上引入“距离”，则可根据词向量之间的距离来判断它们对应的词之间的（词法、语义上的）相似性。

本实施方式中，以欧氏距离来衡量向量之间的距离，间接衡量向量对应的词的语义的相似性，即语义相同或相似的词语词向量距离相近。通过分别计算目标关键词向量与各预置词向量间的欧氏距离，确定预置词向量库中与目标关键词向量最接近的一个或多个预置词向量，进而确定目标关键词向量对应的目标关键词。

欧氏距离是指词向量在各维度上差值的平方和的算术平方根，以公式表示为：

$$D(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

其中， $D(X, Y)$ 指 X 词向量与 Y 词向量的欧氏距离，n 为向量维度，x、y 为各维度的 X 词向量与 Y 词向量。

步骤 S32，从所有的预置词向量中获取与所述目标关键词向量之间欧氏距离最小的匹配词向量，并从所述预置词向量库中获取所述匹配词向量对应的匹配词，所述匹配词为目标关键词。

欧氏距离最小，向量越接近，与目标关键词向量之间欧氏距离最小的预置词向量为与目标关键词向量最接近的词向量，其对应的词即为目标关键词。

本实施例通过获取待提取文本，根据预置词向量库将所述待提取

文本转换为对应的词向量组；根据预置的最优生成模型从所述词向量组中提取目标关键词向量，即通过将待提取文本转换为向量化数据，并将其作为生成模型的输入，可减少模型运算量，提高文本关键词提取效率；根据所述预置词向量库，将所述目标关键词向量转换为对应的目标关键词，提取所述目标关键词作为所述待提取文本的文本关键词，实现对待提取文本的文本关键词的提取。

进一步地，如图 3，在本申请文本关键词提取方法的第二实施例中，所述步骤 S20 之前包括：

步骤 S21，根据预置词向量库将预置的训练文本转换为对应的训练词向量组，并获取所述训练词向量组中的真实关键词向量；

预置的训练文本，即预置的用于训练生成模型和判别模型的训练样本，提取设备可直接从互联网上拉取训练样本，也可从语料库获得训练样本。提取设备获取训练文本后，对训练文本进行分词，得到组成训练文本的各训练词，再将各训练词向量化，得到各训练词各自对应的训练词向量，从而得到对应的训练词向量组，其中，训练文本的分词方式与待转换的语料文本的分词方式一致，相关分词方式已在前文解释，此处不赘述。

可通过查询预置词向量库，获取组成训练文本的各训练词对应的训练词向量，将训练文本的各训练词转换为对应的训练词向量，获得训练词向量组；也可通过对待转换词相同的词向量转化方式对训练文本进行向量化，此处不赘述。

训练词向量组是实际输入到生成模型和判别模型以进行训练模型的样本数据，训练词向量组包括多个训练词向量。

可由用户输入训练样本的真实关键词，提取设备将真实关键词向量化获得真实关键词向量；也可从在抓取/获取训练样本时，获取关键词标签，作为训练样本的真实关键词，提取设备将真实关键词向量化获得真实关键词向量。

生成模型用于从文本中提取关键词，即预测文本关键词，判别模型用于判断生成模型输出的是不是真实关键词。生成模型和判别模型均为神经网络模型，初始模型参数随机设定，并没有进行优化，随后，

这两个模型一起对抗训练,生成模型产生预测的文本关键词给判别模型进行判别,判别模型判断生成模型输出的是不是真实关键词,在这两个模型训练的过程中,不断优化模型参数,两个模型的能力越来越强,最终达到稳态。

步骤 S22,将所述训练词向量组输入到最新生成模型中,并由所述最新生成模型输出从所述训练词向量组中提取预测关键词向量;

在模型训练过程中,模型参数不断优化更新,最新生成模型指本次输入训练词向量时、具有最新模型参数的生成模型,最新判别模型指本次输入训练词向量时、具有最新模型参数的判别模型。

初始化模型的模型参数随机设定,并没有进行优化,因此生成模型第一次经过内部神经网络计算出的预测关键词向量是随机的。预测关键词向量即生成模型经过内部运算从训练词向量组中选取的一个或多个关键词向量。

步骤 S23,将所述真实关键词向量和所述预测关键词向量输入到最新判别模型中,并由所述最新判别模型输出所述预测关键词向量与所述真实关键词向量的匹配概率;

判别模型的训练数据中包括两类输入,一类是训练文本对应的训练词向量组及训练文本的真实关键词向量,另一类是训练文本及生成模型生成的预测关键词向量,判别模型的目标是将真实关键词向量与预测关键词向量进行区分。

最新判别模型对预测关键词向量与真实关键词向量的匹配概率进行计算,具体地,所述步骤 S23 中将所述真实关键词向量和所述预测关键词向量输入到最新判别模型中,与由所述最新判别模型输出所述预测关键词向量与所述真实关键词向量的匹配概率之间包括:

步骤 S231,分别计算各预测关键词向量与每个真实关键词向量的欧氏距离;

欧氏距离是指词向量在各维度上差值的平方和的算术平方根,以公式表示为:

$$D(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

其中, $D(X, Y)$ 指 X 词向量与 Y 词向量的欧氏距离, n 为向量维度, x 、 y 为各维度的 X 词向量与 Y 词向量。

用欧氏距离表征词向量对应词语的相似度, 欧氏距离越小, 预测关键词向量与真实关键词向量对应的词语语义更接近, 则预测关键词向量与真实关键词向量更匹配。

步骤 S232, 统计与预设个数真实关键词向量的欧氏距离小于预设值的匹配预测词向量的数目, 所述预设个数至少为一个;

以预设个数为一个为例对本实施例进行解释, 即统计与任一真实关键词向量的欧氏距离小于预设值的匹配预测词向量的数目。对于一个待提取文本/训练文本, 真实关键词向量和预测关键词向量都可能有多, 各个预测关键词向量只要与任意一个真实关键词向量匹配, 就相当于该预测关键词向量与真实关键词向量匹配。

预设值可以在模型训练过程中, 经内部损失函数及参数优化运算获得, 也可以为模型初始预设的值。

欧氏距离小于预设值, 即预测关键词向量与真实关键词向量匹配的阈值条件, 匹配预测词向量即与预设个数真实关键词向量的欧氏距离小于预设值的预测关键词向量, 在本实施例中, 匹配预测词向量与真实关键词向量匹配。

步骤 S233, 基于所述匹配预测词向量的数目计算所述预测关键词向量与所述真实关键词向量的匹配概率。

在一实施方式中, 匹配预测词向量的数目与所有预测关键词向量数目的比值为匹配概率; 在另一实施方式中, 匹配预测词向量的数目与所有真实关键词向量数目的比值为匹配概率。

步骤 S24, 若所述匹配概率大于预设阈值, 则所述最新生成模型为预置的最优生成模型。

若所述匹配概率大于预设阈值, 则说明最新生成模型的模型参数已取得最优参数, 则所述最新生成模型为预置的最优生成模型, 用于后续的待提取文本的关键词提取。

进一步地, 所述步骤 S23 之后包括:

步骤 S25, 若所述匹配概率小于预设阈值, 则根据所述匹配概率

计算获得所述最新生成模型和所述最新判别模型各自的损失函数；

步骤 S26，根据所述最新生成模型和所述最新判别模型各自的损失函数对所述最新生成模型和所述最新判别模型各自的模型参数进行优化，以获得模型参数优化更新后的最新生成模型和最新判别模型；

最新判别模型的损失函数如下所示：

$$-((1-y)\log(1-D(G(z))))-y\log(D(x))$$

其中， y 为生成模型输出的匹配概率， $G(z)$ 是生成模型的输出， $D(x)$ 为判别模型的输出。

最新判别模型的损失函数的含义在于：尽可能将与真实关键词向量匹配的预测关键词向量标为 $y=1$ ，将与真实关键词向量不匹配的预测关键词向量标为 $y=0$ ，通过上述函数对最新判别模型中神经网络的参数进行优化。

当更新完判别模型的参数后，再更新生成模型的参数。

生成模型的损失函数如下所示：

$$(1-y)\log(1-D(G(z)))$$

其中， y 为生成模型输出的匹配概率， $G(z)$ 是生成模型的输出。

生成模型需要尽可能生成预测关键词向量，使得判别模型无法将其判别为假。在这种情况下，生成模型便能够生成具有较高可信度的预测关键词向量。在得到生成模型的损失函数后，通过生成模型的损失函数对生成模型的神经网络中的参数进行优化。损失函数用于描述模型的生成能力或判别能力，损失函数越小，模型的生成能力或判别能力越高，通过损失函数对神经网络中的参数求导，使损失函数最小化，以求得较优的模型参数。

其中，步骤 S26 中所述根据所述最新生成模型和所述最新判别模型各自的损失函数对所述最新生成模型和所述最新判别模型各自的模型参数进行优化的步骤包括：

步骤 S261，根据所述最新生成模型和所述最新判别模型各自的损失函数，通过 ADAM 算法对所述最新生成模型和所述最新判别模型各自的模型参数进行优化。

ADAM 算法 (Adaptive Moment Estimation) 即自适应时刻估计

方法,通过计算梯度的一阶矩估计和二阶矩估计而为不同的参数设计独立的自适应性学习率,可实现基于训练数据迭代地更新最新生成模型和最新判别模型中神经网络的权重。ADAM 算法具体步骤包括:

分别确定 α (学习率)、 β_1 (一阶矩估计的指数衰减率,如 0.9); β_2 (二阶矩估计的指数衰减率,如 0.999); 随机目标函数 (e) (本实施例中即最新生成模型和最新判别模型); 在确定了参数 α 、 β_1 、 β_2 和随机目标函数 (e) 之后,初始化参数向量、一阶矩向量、二阶矩向量和时间步; 当参数 e 没有收敛时,循环迭代地更新各个部分: 即时间步 t 加 1、更新目标函数在该时间步上对参数 e 所求的梯度、更新偏差的一阶矩估计和二阶原始矩估计,再计算偏差修正的一阶矩估计和偏差修正的二阶矩估计,然后再用以上计算出来的值更新最新生成模型和最新判别模型的参数 e 。

ADAM 算法与其他自适应学习率算法相比,其收敛速度更快,学习效果更为有效,可纠正其他优化技术中存在的问题,如学习率消失、收敛过慢或是高方差的参数更新导致损失函数波动较大等问题。

步骤 S27,执行所述将所述训练词向量组输入到最新生成模型中的步骤。

在对生成模型和判别模型进行优化后,利用优化后的生成模型和判别模型执行所述将所述训练词向量组输入到最新生成模型中的步骤,直至判别模型输出匹配概率大于预设阈值,迭代终止。

此外,本申请还提供一种与上述文本关键词提取方法各步骤对应的文本关键词提取装置。

参照图 4,图 4 为本申请文本关键词提取装置第一实施例的功能模块示意图。

在本实施例中,本申请文本关键词提取装置包括:

第一向量转换模块 10,设置为获取待提取文本,根据预置词向量库将所述待提取文本转换为对应的词向量组;

关键词生成模块 20,设置为根据预置的最优生成模型从所述词向量组中提取目标关键词向量;

第二向量转换模块 30,设置为根据所述预置词向量库,将所述

目标关键词向量转换为对应的目标关键词,提取所述目标关键词作为所述待提取文本的文本关键词。

进一步地,所述第二向量转换模块 30,还设置为遍历所述预置词向量库中所有的预置词向量,分别计算各所述预置词向量与所述目标关键词向量的欧氏距离;从所有的预置词向量中获取与所述目标关键词向量之间欧氏距离最小的匹配词向量,并从所述预置词向量库中获取所述匹配词向量对应的匹配词,所述匹配词为目标关键词。

进一步地,所述文本关键词提取装置包括:

训练模块,设置为根据预置词向量库将预置的训练文本转换为对应的训练词向量组,并获取所述训练词向量组中的真实关键词向量;将所述训练词向量组输入到最新生成模型中,并由所述最新生成模型输出从所述训练词向量组中提取预测关键词向量;将所述真实关键词向量和所述预测关键词向量输入到最新判别模型中,并由所述最新判别模型输出所述预测关键词向量与所述真实关键词向量的匹配概率;若所述匹配概率大于预设阈值,则所述最新生成模型为预置的最优生成模型。

进一步地,所述训练模块,还设置为若所述匹配概率小于预设阈值,则根据所述匹配概率计算获得所述最新生成模型和所述最新判别模型各自的损失函数;根据所述最新生成模型和所述最新判别模型各自的损失函数对所述最新生成模型和所述最新判别模型各自的模型参数进行优化,以获得模型参数优化更新后的最新生成模型和最新判别模型;执行所述将所述训练词向量组输入到最新生成模型中的步骤。

进一步地,所述训练模块,还设置为根据所述最新生成模型和所述最新判别模型各自的损失函数,通过 ADAM 算法对所述最新生成模型和所述最新判别模型各自的模型参数进行优化。

进一步地,所述文本关键词提取装置还包括:

分词模块,设置为获取待转换的语料文本,将所述语料文本进行分词,获得分词后的待转换词组;

向量转换模块,设置为将所述待转换词组中各待转换词转换成对应的词向量,将各所述待转换词与对应的词向量关联存储在预置词向

量库。

进一步地,所述训练模块,还设置为分别计算各预测关键词向量与每个真实关键词向量的欧氏距离;统计与预设个数真实关键词向量的欧氏距离小于预设值的匹配预测词向量的数目,所述预设个数至少为一个;基于所述匹配预测词向量的数目计算所述预测关键词向量与所述真实关键词向量的匹配概率。

本申请还提出一种计算机可读存储介质,该计算机可读存储介质可以为非易失性可读计算机可读存储介质,其上存储有计算机程序。所述计算机可读存储介质可以是图 1 的文本关键词提取设备中的存储器 201,也可以是如 ROM(Read-Only Memory,只读存储器)/RAM(Random Access Memory,随机存取存储器)、磁碟、光盘中的至少一种,所述计算机可读存储介质包括若干指令用以使得一台具有处理器的设备(可以是手机,计算机,服务器,网络设备或本申请实施例中的文本关键词提取设备等)执行本申请各个实施例所述的方法。

需要说明的是,在本文中,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者服务端不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者服务端所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括该要素的过程、方法、物品或者服务端中还存在另外的相同要素。

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、一种文本关键词提取方法，其中，所述文本关键词提取方法包括以下步骤：

获取待提取文本，根据预置词向量库将所述待提取文本转换为对应的词向量组；

根据预置的最优生成模型从所述词向量组中提取目标关键词向量；

根据所述预置词向量库，将所述目标关键词向量转换为对应的目标关键词，提取所述目标关键词作为所述待提取文本的文本关键词。

2、如权利要求 1 所述的文本关键词提取方法，其中，所述根据所述预置词向量库，将所述目标关键词向量转换为对应的目标关键词的步骤包括：

遍历所述预置词向量库中所有的预置词向量，分别计算各所述预置词向量与所述目标关键词向量的欧氏距离；

从所有的预置词向量中获取与所述目标关键词向量之间欧氏距离最小的匹配词向量，并从所述预置词向量库中获取所述匹配词向量对应的匹配词，所述匹配词为目标关键词。

3、如权利要求 1 所述的文本关键词提取方法，其中，所述根据预置的最优生成模型从所述词向量组中提取目标关键词向量的步骤之前包括：

根据预置词向量库将预置的训练文本转换为对应的训练词向量组，并获取所述训练词向量组中的真实关键词向量；

将所述训练词向量组输入到最新生成模型中，并由所述最新生成模型输出从所述训练词向量组中提取预测关键词向量；

将所述真实关键词向量和所述预测关键词向量输入到最新判别模型中，并由所述最新判别模型输出所述预测关键词向量与所述真实关键词向量的匹配概率；

若所述匹配概率大于预设阈值，则所述最新生成模型为预置的最优生成模型。

4、如权利要求 3 所述的文本关键词提取方法，其中，所述由所述最新判别模型输出所述预测关键词向量与所述真实关键词向量的匹配概率的步骤之后包括：

若所述匹配概率小于预设阈值，则根据所述匹配概率计算获得所述最新生成模型和所述最新判别模型各自的损失函数；

根据所述最新生成模型和所述最新判别模型各自的损失函数对所述最新生成模型和所述最新判别模型各自的模型参数进行优化，以获得模型参数优化更新后的最新生成模型和最新判别模型；

执行所述将所述训练词向量组输入到最新生成模型中的步骤。

5、如权利要求 4 所述的文本关键词提取方法，其中，所述根据所述最新生成模型和所述最新判别模型各自的损失函数对所述最新生成模型和所述最新判别模型各自的模型参数进行优化的步骤包括：

根据所述最新生成模型和所述最新判别模型各自的损失函数，通过 ADAM 算法对所述最新生成模型和所述最新判别模型各自的模型参数进行优化。

6、如权利要求 1 所述的文本关键词提取方法，其中，所述根据预置词向量库将所述待提取文本转换为对应的词向量组的步骤之前包括：

获取待转换的语料文本，将所述语料文本进行分词，获得分词后的待转换词组；

将所述待转换词组中各待转换词转换成对应的词向量，将各所述待转换词与对应的词向量关联存储在预置词向量库。

7、如权利要求 3 所述的文本关键词提取方法，其中，所述将所述真实关键词向量和所述预测关键词向量输入到最新判别模型中，与由所述最新判别模型输出所述预测关键词向量与所述真实关键词向量的匹配概率之间包括：

分别计算各预测关键词向量与每个真实关键词向量的欧氏距离；

统计与预设个数真实关键词向量的欧氏距离小于预设值的匹配预测词向量的数目，所述预设个数至少为一个；

基于所述匹配预测词向量的数目计算所述预测关键词向量与所

述真实关键词向量的匹配概率。

8、一种文本关键词提取装置，其中，所述文本关键词提取装置包括：

第一向量转换模块，设置为获取待提取文本，根据预置词向量库将所述待提取文本转换为对应的词向量组；

关键词生成模块，设置为根据预置的最优生成模型从所述词向量组中提取目标关键词向量；

第二向量转换模块，设置为根据所述预置词向量库，将所述目标关键词向量转换为对应的目标关键词，提取所述目标关键词作为所述待提取文本的文本关键词。

9、如权利要求 8 所述的文本关键词提取装置，其中，所述第二向量转换模块，还设置为遍历所述预置词向量库中所有的预置词向量，分别计算各所述预置词向量与所述目标关键词向量的欧氏距离；从所有的预置词向量中获取与所述目标关键词向量之间欧氏距离最小的匹配词向量，并从所述预置词向量库中获取所述匹配词向量对应的匹配词，所述匹配词为目标关键词。

10、如权利要求 8 所述的文本关键词提取装置，其中，所述文本关键词提取装置包括：

训练模块，设置为根据预置词向量库将预置的训练文本转换为对应的训练词向量组，并获取所述训练词向量组中的真实关键词向量；将所述训练词向量组输入到最新生成模型中，并由所述最新生成模型输出从所述训练词向量组中提取预测关键词向量；将所述真实关键词向量和所述预测关键词向量输入到最新判别模型中，并由所述最新判别模型输出所述预测关键词向量与所述真实关键词向量的匹配概率；若所述匹配概率大于预设阈值，则所述最新生成模型为预置的最优生成模型。

11、如权利要求 10 所述的文本关键词提取装置，其中，所述训练模块，还设置为若所述匹配概率小于预设阈值，则根据所述匹配概率计算获得所述最新生成模型和所述最新判别模型各自的损失函数；根据所述最新生成模型和所述最新判别模型各自的损失函数对所述

最新生成模型和所述最新判别模型各自的模型参数进行优化,以获得模型参数优化更新后的最新生成模型和最新判别模型;执行所述将所述训练词向量组输入到最新生成模型中的步骤。

12、如权利要求 8 所述的文本关键词提取装置,其中,所述文本关键词提取装置包括:

分词模块,设置为获取待转换的语料文本,将所述语料文本进行分词,获得分词后的待转换词组;

向量转换模块,设置为将所述待转换词组中各待转换词转换成对应的词向量,将各所述待转换词与对应的词向量关联存储在预置词向量库。

13、一种文本关键词提取设备,其中,所述文本关键词提取设备包括处理器、存储器、以及存储在所述存储器上并可被所述处理器执行的计算机可读指令,其中所述计算机可读指令被所述处理器执行时,实现如下步骤:

获取待提取文本,根据预置词向量库将所述待提取文本转换为对应的词向量组;

根据预置的最优生成模型从所述词向量组中提取目标关键词向量;

根据所述预置词向量库,将所述目标关键词向量转换为对应的目标关键词,提取所述目标关键词作为所述待提取文本的文本关键词。

14、如权利要求 13 所述的文本关键词提取设备,其中,所述根据所述预置词向量库,将所述目标关键词向量转换为对应的目标关键词的步骤包括:

遍历所述预置词向量库中所有的预置词向量,分别计算各所述预置词向量与所述目标关键词向量的欧氏距离;

从所有的预置词向量中获取与所述目标关键词向量之间欧氏距离最小的匹配词向量,并从所述预置词向量库中获取所述匹配词向量对应的匹配词,所述匹配词为目标关键词。

15、如权利要求 13 所述的文本关键词提取设备,其中,所述根据预置的最优生成模型从所述词向量组中提取目标关键词向量的步

骤之前包括：

根据预置词向量库将预置的训练文本转换为对应的训练词向量组，并获取所述训练词向量组中的真实关键词向量；

将所述训练词向量组输入到最新生成模型中，并由所述最新生成模型输出从所述训练词向量组中提取预测关键词向量；

将所述真实关键词向量和所述预测关键词向量输入到最新判别模型中，并由所述最新判别模型输出所述预测关键词向量与所述真实关键词向量的匹配概率；

若所述匹配概率大于预设阈值，则所述最新生成模型为预置的最优生成模型。

16、如权利要求 15 所述的文本关键词提取设备，其中，所述由所述最新判别模型输出所述预测关键词向量与所述真实关键词向量的匹配概率的步骤之后包括：

若所述匹配概率小于预设阈值，则根据所述匹配概率计算获得所述最新生成模型和所述最新判别模型各自的损失函数；

根据所述最新生成模型和所述最新判别模型各自的损失函数对所述最新生成模型和所述最新判别模型各自的模型参数进行优化，以获得模型参数优化更新后的最新生成模型和最新判别模型；

执行所述将所述训练词向量组输入到最新生成模型中的步骤。

17、如权利要求 13 所述的文本关键词提取设备，其中，所述所述根据预置词向量库将所述待提取文本转换为对应的词向量组的步骤之前包括：

获取待转换的语料文本，将所述语料文本进行分词，获得分词后的待转换词组；

将所述待转换词组中各待转换词转换成对应的词向量，将各所述待转换词与对应的词向量关联存储在预置词向量库。

18、一种计算机可读存储介质，其中，所述计算机可读存储介质上存储有计算机可读指令，其中所述计算机可读指令被处理器执行时，实现如下步骤：

获取待提取文本，根据预置词向量库将所述待提取文本转换为对

应的词向量组；

根据预置的最优生成模型从所述词向量组中提取目标关键词向量；

根据所述预置词向量库，将所述目标关键词向量转换为对应的目标关键词，提取所述目标关键词作为所述待提取文本的文本关键词。

19、如权利要求 18 所述的计算机可读存储介质，其中，所述根据所述预置词向量库，将所述目标关键词向量转换为对应的目标关键词的步骤包括：

遍历所述预置词向量库中所有的预置词向量，分别计算各所述预置词向量与所述目标关键词向量的欧氏距离；

从所有的预置词向量中获取与所述目标关键词向量之间欧氏距离最小的匹配词向量，并从所述预置词向量库中获取所述匹配词向量对应的匹配词，所述匹配词为目标关键词。

20、如权利要求 18 所述的计算机可读存储介质，其中，所述根据预置的最优生成模型从所述词向量组中提取目标关键词向量的步骤之前包括：

根据预置词向量库将预置的训练文本转换为对应的训练词向量组，并获取所述训练词向量组中的真实关键词向量；

将所述训练词向量组输入到最新生成模型中，并由所述最新生成模型输出从所述训练词向量组中提取预测关键词向量；

将所述真实关键词向量和所述预测关键词向量输入到最新判别模型中，并由所述最新判别模型输出所述预测关键词向量与所述真实关键词向量的匹配概率；

若所述匹配概率大于预设阈值，则所述最新生成模型为预置的最优生成模型。

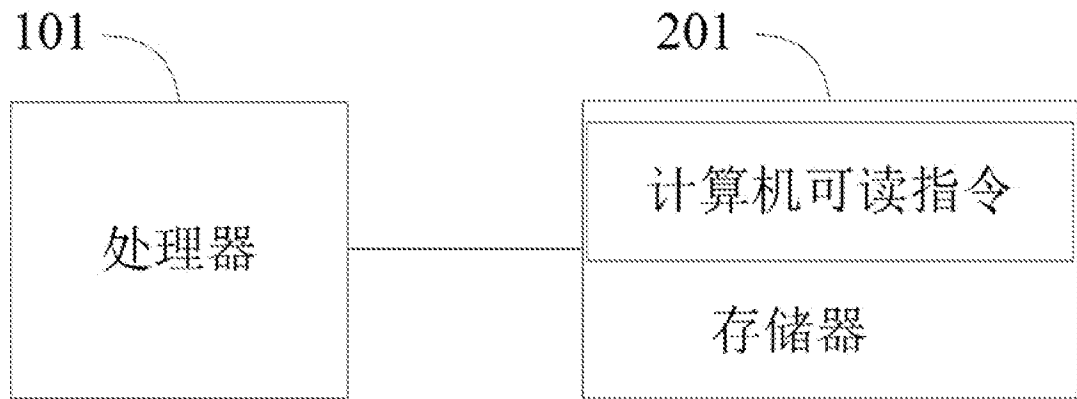


图 1

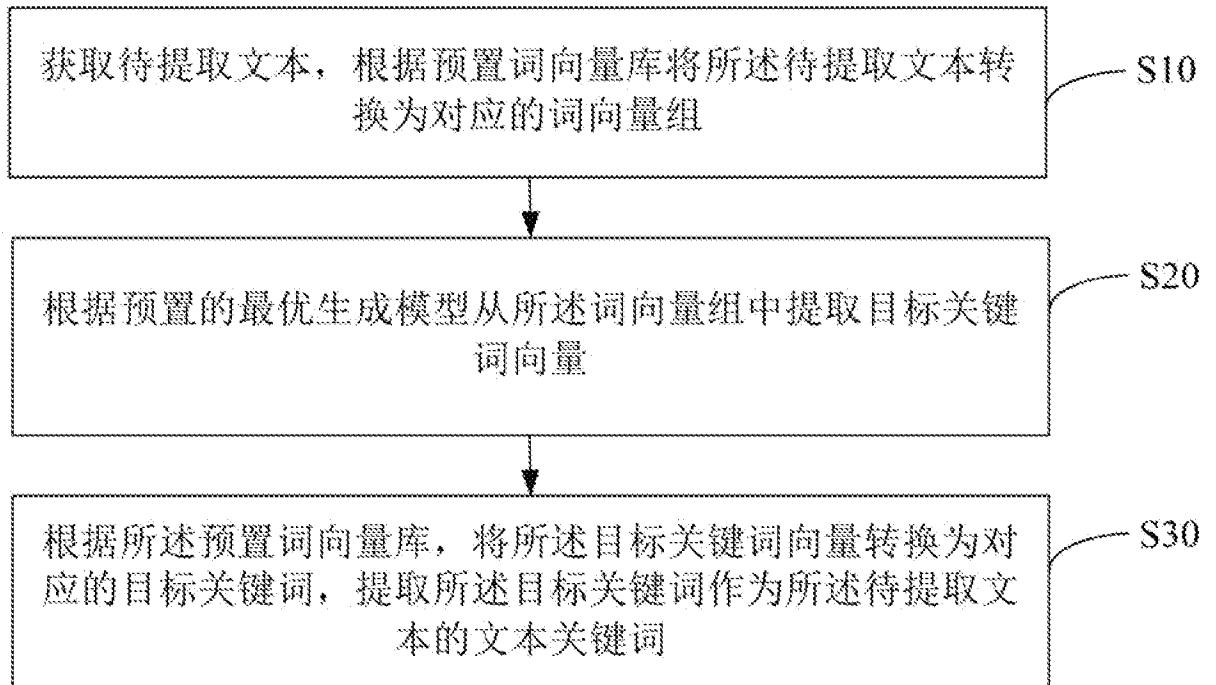


图 2

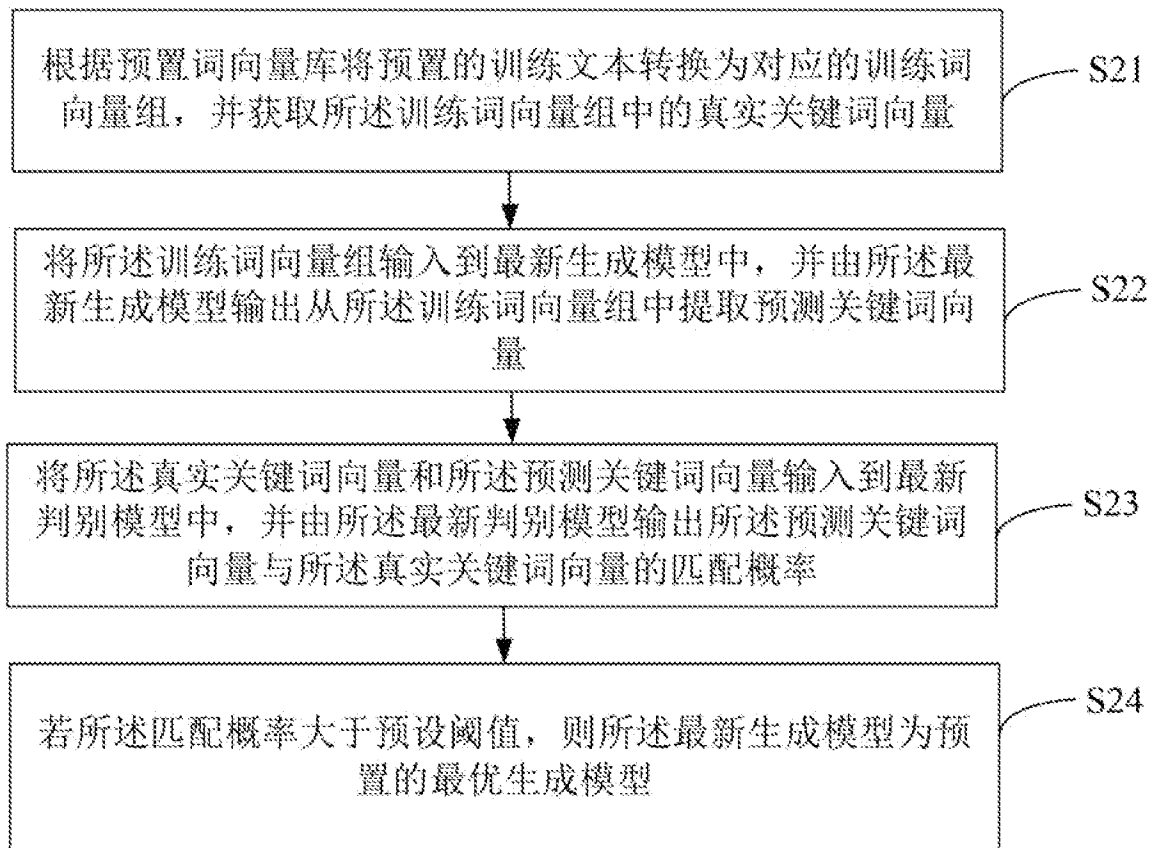


图 3

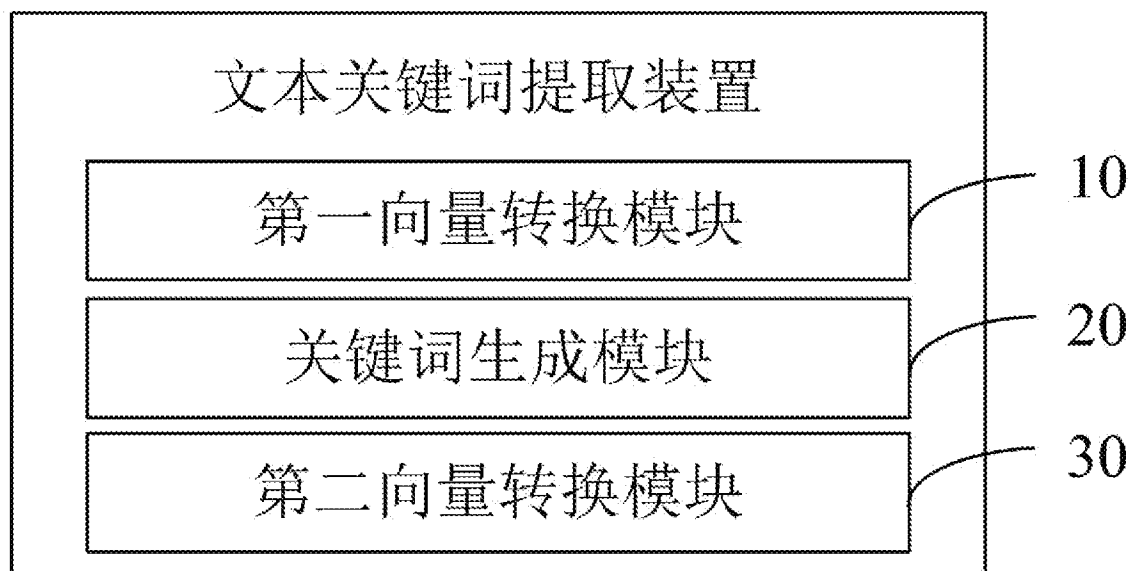


图 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/122813

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)

CNABS; CNTXT; DWPI; VEN; USTXT; EPTXT; WOTXT; CNKI: 文本, 关键词, 提取, 词向量, 训练, 模型, 匹配, 概率, text, keyword, extract, word vector, training, model, match, probability

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 108133045 A (GCI SCIENCE & TECHNOLOGY CO., LTD.) 08 June 2018 (2018-06-08) description, paragraphs [0006]-[0065] and [0192]	1-20
A	CN 108304364 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 20 July 2018 (2018-07-20) entire document	1-20
A	CN 107704503 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 16 February 2018 (2018-02-16) entire document	1-20
A	WO 2018183306 A1 (EBAY INC.) 04 October 2018 (2018-10-04) 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

16 July 2019

Date of mailing of the international search report

29 July 2019

Name and mailing address of the ISA/CN

**China National Intellectual Property Administration (ISA/
CN)**
**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/CN2018/122813

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	108133045	A	08 June 2018	None			
CN	108304364	A	20 July 2018	None			
CN	107704503	A	16 February 2018	WO	2019041521	A1	07 March 2019
				EP	3477495	A1	01 May 2019
				KR	20190038751	A	09 April 2019
				AU	2017408801	A1	14 March 2019
WO	2018183306	A1	04 October 2018	US	2018285928	A1	04 October 2018

国际检索报告

国际申请号

PCT/CN2018/122813

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))

CNABS; CNTXT; DWPI; VEN; USTXT; EPTXT; WOTXT; CNKI: 文本, 关键词, 提取, 词向量, 训练, 模型, 匹配, 概率, text, keyword, extract, word vector, training, model, match, probability

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
X	CN 108133045 A (广州杰赛科技股份有限公司) 2018年 6月 8日 (2018 - 06 - 08) 说明书第6-65、192段	1-20
A	CN 108304364 A (腾讯科技深圳有限公司) 2018年 7月 20日 (2018 - 07 - 20) 全文	1-20
A	CN 107704503 A (平安科技深圳有限公司) 2018年 2月 16日 (2018 - 02 - 16) 全文	1-20
A	W0 2018183306 A1 (EBAY INC) 2018年 10月 4日 (2018 - 10 - 04) 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 16日

国际检索报告邮寄日期

2019年 7月 29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

授权官员

房琦

电话号码 86-(20)-28950714

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/122813

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	108133045	A	2018年 6月 8日	无			
CN	108304364	A	2018年 7月 20日	无			
CN	107704503	A	2018年 2月 16日	W0	2019041521	A1	2019年 3月 7日
				EP	3477495	A1	2019年 5月 1日
				KR	20190038751	A	2019年 4月 9日
				AU	2017408801	A1	2019年 3月 14日
W0	2018183306	A1	2018年 10月 4日	US	2018285928	A1	2018年 10月 4日