

(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 1 月 9 日 (09.01.2020)



(10) 国际公布号
WO 2020/007138 A1

(51) 国际专利分类号:
G06F 16/35 (2019.01)

(21) 国际申请号: PCT/CN2019/089068

(22) 国际申请日: 2019 年 5 月 29 日 (29.05.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201810716380.7 2018 年 7 月 3 日 (03.07.2018) CN

(71) 申请人: 腾讯科技(深圳)有限公司 (TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED) [CN/CN]; 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。

(72) 发明人: 刘树林 (LIU, Shulin); 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。

(74) 代理人: 北京康信知识产权代理有限公司 (KANGXIN PARTNERS, P.C.); 中国北京市海淀区知春路甲 48 号盈都大厦 A 座 16 层, Beijing 100098 (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,

(54) **Title:** METHOD FOR EVENT IDENTIFICATION, METHOD FOR MODEL TRAINING, DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 一种事件识别的方法、模型训练的方法、设备及存储介质

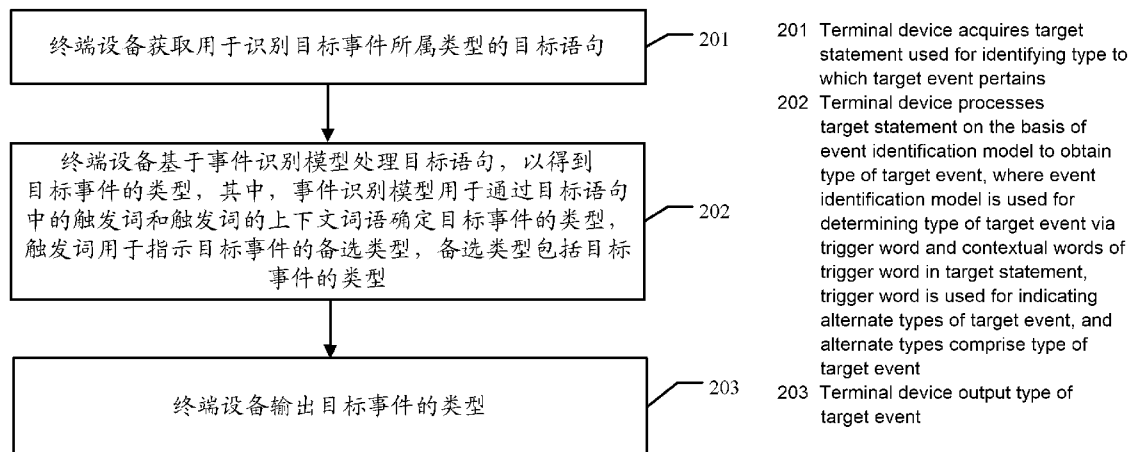


图 5

(57) **Abstract:** A method for event identification, comprising: a terminal device acquires a target statement used for identifying the type to which a target event pertains (201); the terminal device processes the target statement on the basis of an event identification model so as to obtain the type of the target event, where the event identification model is used for determining the type of the target event on the basis of a trigger word and of contextual words of the trigger word in the target statement, the trigger word is used for indicating alternate types of the target event, and the alternate types comprise the type of the target event (202); and the terminal device outputs the type of the target event (203). Because the trigger word and the contextual words of the trigger word are combined for identification in an event identification process, the employment of such method for word categorization for event identification, not only increases the identification efficiency, but, with the contextual words being combined for identification, also increases the accuracy of event identification.

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))。

(57) 摘要: 一种事件识别的方法, 包括: 终端设备获取用于识别目标事件所属类型的目标语句(201); 终端设备基于事件识别模型处理目标语句, 以得到目标事件的类型, 其中, 事件识别模型用于通过目标语句中的触发词和触发词的上下文词语确定目标事件的类型, 触发词用于指示目标事件的备选类型, 备选类型包括目标事件的类型(202); 终端设备输出目标事件的类型(203)。由于在事件识别过程中结合触发词和触发词的上下文词语进行识别, 采用这种词分类的方法进行事件识别, 不仅可以提高识别效率, 而且结合上下文词语进行识别, 可以提高事件识别的准确度。

一种事件识别的方法、模型训练的方法、设备及存储介质

本申请要求于 2018 年 7 月 3 日提交中国专利局、优先权号为 2018107163807、发明名称为“一种事件识别的方法、模型训练的方法、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

本申请涉及语音处理技术领域，具体涉及一种事件识别的方法、事件识别模型训练的方法、设备及计算机可读存储介质。

背景技术

在网络发达的今天，人们对于新闻阅读的需求早已从传统纸媒转移到了网络媒体。在大型网络媒体门户如腾讯、网易、新浪上发布的新闻，都具有发布总量大、更新频率高、阅读用户人数多、读者分布广等特点。对于单个用户而言，都希望能读到自己感兴趣的新闻事件。对此，各个网络媒体门户运营商也各出奇招，例如：通过用户的浏览记录来做后续的新闻推荐。

无论是哪种方式的推荐，都需要后台设备先识别新闻事件的内容才能推荐给相应的用户，当前的新闻事件识别方式主要是通过模式匹配的方式进行识别，模式匹配的识别方式就是给定一段内容，然后从大量数据中去找与该一段内容相同的内容，由此可见，模式匹配的识别方式可移植性差，在每分钟都有可能产生海量新闻事件的情况下，事件识别的效率低下。

发明内容

本申请实施例提供一种事件识别的方法，可以高效准确的识别出各种类型的事件。本申请实施例还提供了相应事件识别模型训练的方法、设备及计算机可读存储介质。

本申请实施例第一方面提供一种事件识别的方法，包括：

终端设备获取用于识别目标事件所属类型的目标语句；

终端设备基于事件识别模型处理上述目标语句，以得到上述目标事件的类型，其中，上述事件识别模型用于通过上述目标语句中的触发词和上述触发词

的上下文词语确定上述目标事件的类型，上述触发词用于指示上述目标事件的备选类型，上述备选类型包括上述目标事件的类型；

终端设备输出上述目标事件的类型。

本申请实施例第二方面提供一种事件识别模型训练的方法，包括：

计算机设备获取用于模型训练的样本集合，上述样本集合中包括多个样本事件各自对应的样本语句和上述多个样本事件各自的类型，每个样本语句中包括触发词和上述触发词的上下文词语；

计算机设备根据上述每个样本语句训练第一事件识别模型，以得到第二事件识别模型，上述第二事件识别模型用于通过目标语句中的触发词和上述目标语句中的触发词的上下文词语确定上述目标事件的类型，上述目标语句为用于识别目标事件所属类型的语句。

本申请实施例第三方面提供一种事件识别的装置，安装在终端设备中，包括：

获取单元，设置为获取用于识别目标事件所属类型的目标语句；

处理单元，设置为基于事件识别模型处理上述获取单元获取的上述目标语句，以得到上述目标事件的类型，其中，上述事件识别模型用于通过上述目标语句中的触发词和上述触发词的上下文词语确定上述目标事件的类型，上述触发词用于指示上述目标事件的备选类型，上述备选类型包括上述目标事件的类型；

输出单元，设置为输出上述处理单元确定的上述目标事件的类型。

本申请实施例第四方面提供一种事件识别模型训练的装置，安装在计算机设备中，包括：

获取单元，设置为获取用于模型训练的样本集合，上述样本集合中包括多个样本事件各自对应的样本语句和上述多个样本事件各自的类型，每个样本语句中包括触发词和上述触发词的上下文词语；

模型训练单元，设置为根据上述获取单元获取的上述每个样本语句训练第一事件识别模型，以得到第二事件识别模型，上述第二事件识别模型用于通过目标语句中的触发词和上述目标语句中的触发词的上下文词语确定上述目标事件的类型，上述目标语句为用于识别目标事件所属类型的语句。

本申请实施例第五方面提供一种终端设备，上述终端设备包括：输入/输出（I/O）接口、处理器和存储器，上述存储器中存储有程序指令；

上述处理器用于执行存储器中存储的程序指令，执行如第一方面上述的方法。

本申请实施例第六方面提供一种计算机设备，上述计算机设备包括：输入/输出（I/O）接口、处理器和存储器，上述存储器中存储有程序指令；

上述处理器用于执行存储器中存储的程序指令，执行如上述第二方面上述的方法。

本申请实施例又一方面提供了一种计算机可读存储介质，上述计算机可读存储介质中存储有指令，当其在计算机上运行时，使得计算机执行上述第一方面上述的方法。

本申请实施例又一方面提供了一种计算机可读存储介质，上述计算机可读存储介质中存储有指令，当其在计算机上运行时，使得计算机执行上述第二方面上述的方法。

本申请实施例又一方面提供了一种包含指令的计算机程序产品，当其在计算机上运行时，使得计算机执行上述第一方面上述的方法。

本申请实施例又一方面提供了一种包含指令的计算机程序产品，当其在计算机上运行时，使得计算机执行上述第二方面上述的方法。

本申请实施例，在事件识别过程中结合触发词和触发词的上下文词语进行识别，采用这种词分类的方法进行事件识别，不仅可以提高识别效率，而且结合上下文词语进行识别，可以提高事件识别的准确度。

附图说明

图1是本申请实施例中事件识别模型训练的系统的一实施例示意图；

图2是本申请实施例中事件识别模型训练的方法的一实施例示意图；

图3是本申请实施例中事件识别模型训练的向量化表示的一示例示意图；

图4是本申请实施例中事件识别应用的一场景示例示意图；

图5是本申请实施例中事件识别的方法的一实施例示意图；

图6是本申请实施例中事件识别的向量化表示的一示例示意图；

图7是本申请实施例中事件识别的装置的一实施例示意图；

图8是本申请实施例中事件识别模型训练的装置的一实施例示意图；

图9是本申请实施例中计算机设备的一实施例示意图。

具体实施方式

下面结合附图，对本申请的实施例进行描述，显然，所描述的实施例仅仅是本申请一部分的实施例，而不是全部的实施例。本领域普通技术人员可知，随着技术的发展和场景的出现，本申请实施例提供的技术方案对于类似的技术问题，同样适用。

本申请实施例提供一种事件识别的方法，可以高效准确的识别出各种类型的事件。本申请实施例还提供了相应事件识别模型训练的方法、设备及计算机可读存储介质。以下分别进行详细说明。

事件识别是基于事件识别模型的，下面先介绍本申请实施例中事件识别模型的训练过程。

图1是本申请实施例中事件识别模型训练的系统的一实施例示意图。

如图1所示，本申请实施例提供的事件识别模型训练的系统包括计算机设备10和数据库20，数据库20中存储有用于事件识别模型训练的样本集合。其中，样本集合中包括多个样本事件各自对应的样本语句和多个样本事件各自的类型。样本事件是指用作样本的事件，事件是发生在某个特定时间或时间段、在某个特定地域范围内，由一个或多个角色参与的，由一个或多个动作构成的事情或状态的改变。当然，没有时间和地点，只是由角色和动作构成的事情或状态的改变也可以称为事件。样本事件的样本语句可以是用于描述该样本事件的核心语句，如样本事件为新闻事件，该样本语句可以为该新闻事件的标题。

每个样本语句中包括触发词和触发词的上下文词语，触发词是样本语句中最能指示该样本事件发生的词。例如：王小五和张小花在深圳结婚了。其中，“结婚”最能指示该这个事件，则“结婚”为触发词。在一起场景中，相同的词可能具有不同的含义，这样，就需要结合上下文词语进行事件识别。如：语句一：“昨天下午，他离开公司后直接去了超市”。语句二：“爷爷永远地离开了我们，但是我们不会忘记他”。语句一中“离开”表示下班，语句二中“离

开”表示去世。所以可见，相同的触发词可能指示了不同的事件类型。所以，需要结合触发词的上下文词语，这样就可以准确的识别出事件的类型。在识别模型训练过程中，针对每个样本事件都会预先标记好该样本事件的类型。这样，计算机设备10就可以使用样本集合中的样本语句和事件的类型进行模型训练，从而训练出用于事件识别的模型。

下面结合图2介绍本申请实施例中事件识别模型训练的方法。

如图2所示，本申请实施例提供的事件识别模型训练的方法的一实施例包括：

101、计算机设备获取用于模型训练的样本集合，样本集合中包括多个样本事件各自对应的样本语句和多个样本事件各自的类型，每个样本语句中包括触发词和触发词的上下文词语。

102、计算机设备根据每个样本语句训练第一事件识别模型，以得到第二事件识别模型，第二事件识别模型用于通过目标语句中的触发词和目标语句中的触发词的上下文词语确定目标事件的类型，目标语句为用于识别目标事件所属类型的语句。

本申请实施例中，在事件识别模型训练过程中，就结合样本事件的样本语句中的触发词和触发词的上下文词语，以及样本事件的类型进行模型训练，从而可以确保训练出能准确进行事件识别的事件识别模型。

其中，可选地，计算机设备根据每个样本语句训练第一事件识别模型，以得到第二事件识别模型，可以包括：

计算机设备对第一样本语句进行预处理，以得到第一样本语句中的每个词语，第一样本语句为每个样本语句中的任意一个样本语句；

计算机设备对每个词语进行向量转换处理，以得到每个词语的输入层向量；

计算机设备根据每个词语的输入层向量，以及第一样本语句所对应的样本事件的类型训练第一事件识别模型，以得到第二事件识别模型。

其中，计算机设备对第一样本语句进行预处理，以得到第一样本语句中的每个词语，可以包括：

计算机设备识别出第一样本语句中的命名实体，命名实体为目标事件中的

参与对象；

计算机设备对命名实体进行匿名化处理；

计算机设备将匿名化处理后的第一样本语句划分为词序列，以得到第一样本语句中的每个词语。

本申请实施例中，预处理的过程包括匿名化处理和对话句的词划分，匿名化是针对命名实体的，命名实体可以是人物，也可以是公司或者其他参与该事件的角色对象。对命名实体进行匿名化可以避免在训练过程中将该命名实体作为强特征将该命名实体与该事件的类型建立必然联系。如：“潘长江参加春晚演出”。该事件的类型是“演出”。如果不匿名化处理，很有可能会建立潘长江与演出之间的必然联系，在模型训练好后，在事件识别的过程中，很可能识别到“潘长江”这个命名实体，就会将相关事件识别是演出事件。但很可能“潘长江”参加的是其他的事件，例如：潘长江的女儿结婚。这是一个结婚事件，该事件的类型为“结婚”，而不是“演出”。所以，在模型训练的过程中对命名实体进行匿名化处理可以使模型学习到正确的特征，并且有利于提高模型的泛化能力。

匿名化就是隐匿具体信息，只保留该信息是命名实体这一类型信息即可，如：潘长江的女儿结婚。匿名化处理后可以为：PER的PER结婚。PER用于表示该位置的词语属于命名实体。

匿名化处理后进行词划分，词划分就是将语句划分为词语，如：“PER的PER结婚”可以划分为“PER”、“的”、“PER”、“结婚”几个词语。

另外，需要说明的是，本申请实施例中，对事件的识别是通过对触发词的识别完成的，在事件识别模型训练阶段，因为样本事件的类型是已知的，所以也可以认为该样本事件的样本语句中的触发词能指示该样本事件的类型。

在模型训练过程中计算机设备可以将每个词语作为候选的触发词，并对其进行分类，这里的目标类别是预先定义好的事件类型，并引入一个NA类别，表示非事件。一条标注数据可以拆分成多条训练样本，上面的例子“潘长江的女儿结婚”可以拆分成如下表1中的四条样本：

表1

候选触发词	所在上下文	类别
潘长江	<u>潘长江</u> 的 女儿 结婚	NA
的	潘长江 <u>的</u> 女儿 结婚	NA
女儿	潘长江 的 <u>女儿</u> 结婚	NA
结婚	潘长江 的 女儿 <u>结婚</u>	结婚

当“结婚”为触发词时，“潘长江”、“的”和“女儿”这三个词为触发词的上下文词语。在样本语句中，确定好触发词和上下文词语后就开始进行向量层面上的处理。

本申请实施例中，无论是第一事件识别模型，还是第二事件识别模型都包括三层全连接神经网络，该三层分别是输入层、隐藏层和输出层，在事件识别过程中，每层都有相应的功能，如：在输入层就会将触发词和上下文词语都转换为输入层向量，然后再根据输入层与隐藏层之间的关系，将输入层向量转换为隐藏层向量。将词语转换为输入层向量可以通过查找向量表的方式实现。

可选地，计算机设备根据每个词语的输入层向量，以及第一样本语句所对应的样本事件的类型训练第一事件识别模型，以得到第二事件识别模型，可以包括：

计算机设备将每个词语的输入层向量转换为隐藏层向量，每个词语包括第一词语，第一词语为每个词语中的触发词；

计算机设备根据每个第二词语的隐藏层向量、第一词语的隐藏层向量以及任意一个第二词语的隐藏层向量，确定任意一个第二词语的关注度权重，第二词语为除第一词语之外的词语；

计算机设备根据每个第二词语的输入层向量和每个第二词语的关注度权重确定第一词语的上下文向量；

计算机设备根据第一词语的上下文向量和第一词语的输入层向量，以及第一样本语句所对应的样本事件的类型，确定第一事件识别模型的参数，以得到第二事件识别模型。

本申请实施例中，如图3所示，将输入层向量转换为隐藏层向量可以通过公式 $\bar{w} = f(w \bullet W)$ 来完成，其中， w 可以为样本语句中每个词语的输入层向量， W 是转换矩阵， $\bar{w}$ 为与 w 对应的隐藏层向量， $f(\cdot)$ 是非线性函数，例如双

曲正切 $\tanh(\cdot)$ 。

因为本申请实施例的方案中涉及到触发词的上下文词语，所以需要确定该触发词的上下文向量，确定该触发词的上下文向量需要先确定该触发词的上下文词语的关注度权重，也就是各上下文词语与触发词相关的权重。

确定上下文词语的关注度权重可以通过公式如下公式计算得到：

$$\alpha_j = \frac{\exp(\bar{w} \cdot \bar{w}_j^T)}{\sum_{i=1}^{n-1} \exp(\bar{w} \cdot \bar{w}_i^T)}$$

其中， n 表示样本语句中总词语的数量， $n-1$ 表示样本语句中上下文词语的数量，也就是除作为触发词的第一词语之外的第二词语的数量， α_j 表示 $n-1$ 个第二词语中第 j 个第二词语的关注度权重， $\bar{w}$ 为第一词语的隐藏层向量， $\bar{w}_j^T$ 表示第 j 个第二词语的隐藏层向量的转置， $\bar{w}_i^T$ 表示第 i 个第二词语的隐藏层向量的转置，其中， i 的取值为从1到 $n-1$ 的整数， j 的取值为从1到 $n-1$ 的整数。

α_j 确定后，可以通过如下公式确定作为触发词的第一词语的上下文向量：

$$c = \frac{1}{n-1} \sum_{j=1}^{n-1} w_j \alpha_j^T$$

其中， c 为第一词语的上下文向量， w_j 为第 j 个第二词语的输入层向量，

α_j^T 为第 j 个第二词语的关注度权重的转置， j 的取值为从1到 $n-1$ 的整数。

在 c 确定后，可以通过 c 和第一词语的输入层向量 w ，以及样本语句所对应的样本事件的类型，确定第一事件识别模型的参数 θ 。

c 和 w 经过隐藏层和输入层两次矩阵转换后，会输入向量 $\mathbf{o}$ ， $\mathbf{o}$ 中的第 i 个元素表示事件识别模型将给定样本语句判断为事件类型 i 的置信度。

为了获得条件概率 $p(i|\mathbf{x}, \theta)$ ，对包含预先配置的事件的 m 个类型置信度的向量 $\mathbf{o}$ 进行softmax归一化操作：

$$p(i|\mathbf{x}, \theta) = \frac{e^{\theta_i}}{\sum_{k=1}^m e^{\theta_k}}$$

k 表示第 k 个事件类型，用于训练的样本集合 $\{(X(i); y(i))\}$ ，若一共有 T 个，则事件识别模型的损失函数定义为：

$$J(\theta) = - \sum_{i=1}^T \log p(y^{(i)} | \mathbf{x}^{(i)}, \theta)$$

可以采用随机梯度下降法，从而计算出 θ 的取值。

以上多个实施例描述了事件识别模型训练的过程，事件识别模型训练好后，可以使用该事件识别模型进行事件识别，事件识别可以应用于新闻推荐中，如图4所示，计算机设备10通过网络30可以与终端设备40通信，该计算机设备10可以用于事件识别。终端设备40的新闻界面上设置不同事件类型的显示位，计算机设备10识别出事件的类型后，将该类型的事件推送给终端设备40，终端设备40会根据新闻事件的类型将该新闻放到相应类型的展示位进行展示。图4中只是列举了几个不同的事件类型，事件类型的数量可以根据需求设置或修改。本申请实施例中事件类型也就是事件的类型。

下面结合图5介绍本申请实施例中事件识别的方法。如图5所示，本申请实施例提供的事件识别的方法的一实施例包括：

201、终端设备获取用于识别目标事件所属类型的目标语句。

若一个与手机发布相关的新闻，如：标题为“iphoneX发布”的新闻事件，则可以将标题“iphoneX发布”确定为目标语句。

202、终端设备基于事件识别模型处理目标语句，以得到目标事件的类型，其中，事件识别模型用于通过目标语句中的触发词和触发词的上下文词语确定目标事件的类型，触发词用于指示目标事件的备选类型，备选类型包括目标事件的类型。

203、终端设备输出目标事件的类型。

本申请实施例，在事件识别过程中结合触发词和触发词的上下文词语进行识别，采用这种词分类的方法进行事件识别，不仅可以提高识别效率，而且结

合上下文词语进行识别，可以提高事件识别的准确度。

可选地，终端设备基于事件识别模型处理目标语句，以得到目标事件的类型，可以包括：

终端设备对目标语句进行预处理，以得到目标语句中的每个词语；

终端设备对每个词语进行向量转换处理，以得到每个词语的输入层向量；

终端设备根据每个词语的输入层向量，确定目标事件的类型。

其中，终端设备对目标语句进行预处理，以得到目标语句中的每个词语，可以包括：

终端设备识别出目标语句中的命名实体，命名实体为目标事件中的参与对象；

终端设备对命名实体进行匿名化处理；

终端设备将匿名化处理后的目标语句划分为词序列，以得到目标语句中的每个词语。

本申请实施例中，预处理的过程包括匿名化处理和对话句的词划分，匿名化是针对命名实体的，命名实体可以是人物，也可以是公司或者其他参与该事件的角色对象。对命名实体进行匿名化可以避免在训练过程中将该命名实体作为强特征将该命名实体与该事件的类型建立必然联系。如：“潘长江参加春晚演出”。该事件的类型是“演出”。如果不匿名化处理，很有可能会建立潘长江与演出之间的必然联系，在事件识别的过程中，很可能识别到“潘长江”这个命名实体，就会将相关事件识别是演出事件。但很可能“潘长江”参加的是其他的事件，例如：潘长江的女儿结婚。这是一个结婚事件，该事件的类型为“结婚”，而不是“演出”。所以在事件识别的过程中会将命名实体匿名化，以避免命名实体对事件识别的干扰，而且也不用具体分析该命名实体，还可以提高识别效率。

匿名化就是隐匿具体信息，只保留该信息是命名实体这一类型信息即可，如：潘长江的女儿结婚。匿名化处理后可以为：PER的PER结婚。PER用于表示该位置的词语属于命名实体。

匿名化处理后进行词划分，词划分就是将语句划分为词语，如：“PER的PER结婚”可以划分为“PER”、“的”、“PER”、“结婚”几个词语。

可选地，终端设备根据每个词语的输入层向量，确定目标事件的类型，可以包括：

终端设备将每个词语的输入层向量转换为隐藏层向量，每个词语包括第一词语，第一词语为每个词语中的任一词语；

终端设备将第一词语作为触发词时，根据每个第二词语的隐藏层向量、第一词语的隐藏层向量以及任意一个第二词语的隐藏层向量，确定任意一个第二词语的关注度权重，第二词语为除第一词语之外的词语；

终端设备根据每个第二词语的输入层向量和每个第二词语的关注度权重确定第一词语的上下文向量；

终端设备根据第一词语的上下文向量和第一词语的输入层向量，确定目标事件的类型。

本申请实施例中，如图6所示，将输入层向量转换为隐藏层向量可以通过公式 $\bar{w} = f(w \bullet W)$ 来完成，其中， w 可以为样本语句中每个词语的输入层向量， W 是转换矩阵， $\bar{w}$ 为与 w 对应的隐藏层向量， $f(\cdot)$ 是非线性函数，例如双曲正切 $\tanh(\cdot)$ 。

因为本申请实施例的方案中涉及到触发词的上下文词语，所以需要确定该触发词的上下文向量，确定该触发词的上下文向量需要先确定该触发词的上下文词语的关注度权重，也就是各上下文词语与触发词相关的权重。

确定上下文词语的关注度权重可以通过公式如下公式计算得到：

$$\alpha_j = \frac{\exp(\bar{w} \bullet \bar{w}_j^T)}{\sum_{i=1}^{n-1} \exp(\bar{w} \bullet \bar{w}_i^T)}$$

其中， n 表示样本语句中总词语的数量， $n-1$ 表示样本语句中上下文词语的数量，也就是除作为触发词的第一词语之外的第二词语的数量， α_j 表示 $n-1$ 个第二词语中第 j 个第二词语的关注度权重， $\bar{w}$ 为第一词语的隐藏层向量， $\bar{w}_j^T$ 表示第 j 个第二词语的隐藏层向量的转置， $\bar{w}_i^T$ 表示第 i 个第二词语的隐藏层向量的转置，其中， i 的取值为从 1 到 $n-1$ 的整数， j 的取值为从 1 到 $n-1$ 的整数。

α_j 确定后,可以通过如下公式确定作为触发词的第一词语的上下文向量:

$$c = \frac{1}{n-1} \sum_{j=1}^{n-1} w_j \alpha_j^T$$

其中, c 为第一词语的上下文向量, w_j 为第 j 个第二词语的输入层向量, α_j^T 为第 j 个第二词语的关注度权重的转置, j 的取值为从 1 到 $n-1$ 的整数。

在 c 确定后,可以通过 c 和第一词语的输入层向量 w , 确定目标事件的类型。

可选地, 终端设备根据第一词语的上下文向量和第一词语的输入层向量, 确定目标事件的类型, 可以包括:

终端设备根据第一词语的上下文向量、第一词语的输入层向量与事件识别模型的参数, 确定第一词语为触发词时, 目标事件属于每个备选类型的置信度;

终端设备根据目标事件属于每个备选类型的置信度, 确定目标事件的类型。

本申请实施例中, 事件识别模型的参数 θ 经过图 3 所示的模型训练过程已经确定取值, 所以可以通过 c 、 w 和 θ 经过隐藏层和输入层两次矩阵转换后得到输入向量 $\mathbf{o}$, $\mathbf{o}$ 中的第 i 个元素表示事件识别模型将给目标样本判断为类型 i 的置信度。

为了获得条件概率 $p(i|\mathbf{x}, \theta)$, 对包含预先配置的事件的 m 个类型置信度的向量 $\mathbf{o}$ 进行 softmax 归一化操作:

$$p(i|\mathbf{x}, \theta) = \frac{e^{o_i}}{\sum_{k=1}^m e^{o_k}}$$

k 表示第 k 个事件类型, 用于训练的样本集合 $\{(X(i); y(i))\}$, 若一共有 T 个, 则事件识别模型的损失函数定义为:

$$J(\theta) = - \sum_{i=1}^T \log p(y^{(i)} | \mathbf{x}^{(i)}, \theta)$$

可以采用随机梯度下降法，从而计算出 $y(i)$ ，也就是目标事件属于各种的类型的概率。在目标事件的类型确定的过程中，会选择最大的 $y(i)$ 的类型作为目标事件的类型。

可选地，本申请实施例中，终端设备基于事件识别模型处理目标语句，以得到目标事件的类型，可以包括：

终端设备对目标语句进行预处理，以得到目标语句中的每个词语；

终端设备确定目标语句中属于触发词的第一词语；

终端设备根据第一词语所对应的备选类型，以及第二词语确定目标事件的类型，第二词语为第一词语的上下文词语。

本申请实施例中，可以预先标记一些能作为触发词的词语与事件的类型的对应关系，即使是有同一个触发词对应不同的事件的类型的情况，也可以缩小计算的范围。其他与向量计算相关的部分可以参阅前述实施例的相应内容进行理解，本处不做重复赘述。

以上介绍了本申请实施例中模型训练和事件识别的过程，下面结合附图介绍本申请实施例中的装置及设备。

如图7所示，本申请实施例提供的事件识别的装置50的一实施例，上述装置50可以安装在终端设备中，包括：

获取单元501，设置为获取用于识别目标事件所属类型的目标语句；

处理单元502，设置为基于事件识别模型处理获取单元501获取的目标语句，以得到目标事件的类型，其中，事件识别模型用于通过目标语句中的触发词和触发词的上下文词语确定目标事件的类型，触发词用于指示目标事件的备选类型，备选类型包括目标事件的类型；

输出单元503，设置为输出处理单元502确定的目标事件的类型。

本申请实施例，在事件识别过程中结合触发词和触发词的上下文词语进行识别，采用这种词分类的方法进行事件识别，不仅可以提高识别效率，而且结合上下文词语进行识别，可以提高事件识别的准确度。

可选地，处理单元502设置为：

对目标语句进行预处理，以得到目标语句中的每个词语；

对每个词语进行向量转换处理，以得到每个词语的输入层向量；

根据每个词语的输入层向量，确定目标事件的类型。

可选地，处理单元502设置为：

将每个词语的输入层向量转换为隐藏层向量，每个词语包括第一词语，第一词语为每个词语中的任一词语；

将第一词语作为触发词时，根据每个第二词语的隐藏层向量、第一词语的隐藏层向量以及任意一个第二词语的隐藏层向量，确定任意一个第二词语的关注度权重，第二词语为除第一词语之外的词语；

根据每个第二词语的输入层向量和每个第二词语的关注度权重确定第一词语的上下文向量；

根据第一词语的上下文向量和第一词语的输入层向量，确定目标事件的类型。

可选地，处理单元502设置为：

根据第一词语的上下文向量、第一词语的输入层向量与事件识别模型的参数，确定第一词语为触发词时，目标事件属于每个备选类型的置信度；

根据目标事件属于每个备选类型的置信度，确定目标事件的类型。

可选地，处理单元502设置为：

识别出目标语句中的命名实体，命名实体为目标事件中的参与对象；

对命名实体进行匿名化处理；

将匿名化处理后的目标语句划分为词序列，以得到目标语句中的每个词语。

可选地，处理单元502设置为：

对目标语句进行预处理，以得到目标语句中的每个词语；

确定目标语句中属于触发词的第一词语；

根据第一词语所对应的备选类型，以及第二词语确定目标事件的类型，第二词语为第一词语的上下文词语。

以上对事件识别的装置50的介绍可以结合上述事件识别的方法部分的相

应内容进行理解，本处不再做重复赘述。

参阅图8，本申请实施例提供的事件识别模型训练的装置60的一实施例，上述装置60可以安装在计算机设备中，包括：

获取单元601，设置为获取用于模型训练的样本集合，样本集合中包括多个样本事件各自对应的样本语句和多个样本事件各自的类型，每个样本语句中包括触发词和触发词的上下文词语；

模型训练单元602，设置为根据获取单元601获取的每个样本语句训练第一事件识别模型，以得到第二事件识别模型，第二事件识别模型用于通过目标语句中的触发词和目标语句中的触发词的上下文词语确定目标事件的类型，目标语句为用于识别目标事件所属类型的语句。

本申请实施例中，在事件识别模型训练过程中，就结合样本事件的样本语句中的触发词和触发词的上下文词语，以及样本事件的类型进行模型训练，从而可以确保训练出能准确进行事件识别的事件识别模型。

可选地，模型训练单元602设置为：

对第一样本语句进行预处理，以得到第一样本语句中的每个词语，第一样本语句为每个样本语句中的任意一个样本语句；

对每个词语进行向量转换处理，以得到每个词语的输入层向量；

根据每个词语的输入层向量，以及第一样本语句所对应的样本事件的类型训练第一事件识别模型，以得到第二事件识别模型。

可选地，模型训练单元602设置为：

将每个词语的输入层向量转换为隐藏层向量，每个词语包括第一词语，第一词语为每个词语中的触发词；

根据每个第二词语的隐藏层向量、第一词语的隐藏层向量以及任意一个第二词语的隐藏层向量，确定任意一个第二词语的关注度权重，第二词语为除第一词语之外的词语；

根据每个第二词语的输入层向量和每个第二词语的关注度权重确定第一词语的上下文向量；

根据第一词语的上下文向量和第一词语的输入层向量，以及第一样本语句

所对应的样本事件的类型，确定第一事件识别模型的参数，以得到第二事件识别模型。

可选地，模型训练单元602设置为：

识别出第一样本语句中的命名实体，命名实体为目标事件中的参与对象；
对命名实体进行匿名化处理；

将匿名化处理后的第一样本语句划分为词序列，以得到第一样本语句中的每个词语。

以上对事件识别模型训练的装置60的介绍可以结合上述事件识别模型训练的方法部分的相应内容进行理解，本处不再做重复赘述。

图9是本申请实施例提供的计算机设备70的结构示意图。计算机设备70包括处理器710、存储器740和输入输出（I/O）接口730，存储器740可以包括只读存储器和随机存取存储器，并向处理器710提供操作指令和数据。存储器740的一部分还可以包括非易失性随机存取存储器（NVRAM）。

在一些实施方式中，存储器740存储了如下的元素，可执行模块或者数据结构，或者他们的子集，或者他们的扩展集：

在本申请实施例中，在事件识别的过程中，通过调用存储器740存储的操作指令（该操作指令可存储在操作系统中），

获取用于识别目标事件所属类型的目标语句；

基于事件识别模型处理目标语句，以得到目标事件的类型，其中，事件识别模型用于通过目标语句中的触发词和触发词的上下文词语确定目标事件的类型，触发词用于指示目标事件的备选类型，备选类型包括目标事件的类型；
输出目标事件的类型。

本申请实施例，在事件识别过程中结合触发词和触发词的上下文词语进行识别，采用这种词分类的方法进行事件识别，不仅可以提高识别效率，而且结合上下文词语进行识别，可以提高事件识别的准确度。

处理器710控制计算机设备70的操作，处理器710还可以称为CPU（Central Processing Unit，中央处理单元）。存储器740可以包括只读存储器和随机存取存储器，并向处理器710提供指令和数据。存储器740的一部分还可以包括非易

失性随机存取存储器（NVRAM）。具体的应用中计算机设备70的各个组件通过总线系统720耦合在一起，其中总线系统720除包括数据总线之外，还可以包括电源总线、控制总线和状态信号总线等。但是为了清楚说明起见，在图中将各种总线都标为总线系统720。

上述本申请实施例揭示的方法可以应用于处理器710中，或者由处理器710实现。处理器710可能是一种集成电路芯片，具有信号的处理能力。在实现过程中，上述方法的各步骤可以通过处理器710中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器710可以是通用处理器、数字信号处理器（DSP）、专用集成电路（ASIC）、现成可编程门阵列（FPGA）或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。可以实现或者执行本申请实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合本申请实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成，或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器，闪存、只读存储器，可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器740，处理器710读取存储器740中的信息，结合其硬件完成上述方法的步骤。

可选地，处理器710用于：

对目标语句进行预处理，以得到目标语句中的每个词语；

对每个词语进行向量转换处理，以得到每个词语的输入层向量；

根据每个词语的输入层向量，确定目标事件的类型。

可选地，处理器710用于：

将每个词语的输入层向量转换为隐藏层向量，每个词语包括第一词语，第一词语为每个词语中的任一词语；

将第一词语作为触发词时，根据每个第二词语的隐藏层向量、第一词语的隐藏层向量以及任意一个第二词语的隐藏层向量，确定任意一个第二词语的关注度权重，第二词语为除第一词语之外的词语；

根据每个第二词语的输入层向量和每个第二词语的关注度权重确定第一

词语的上下文向量;

根据第一词语的上下文向量和第一词语的输入层向量,确定目标事件的类型。

可选地,处理器710用于:

根据第一词语的上下文向量、第一词语的输入层向量与事件识别模型的参数,确定第一词语为触发词时,目标事件属于每个备选类型的置信度;

根据目标事件属于每个备选类型的置信度,确定目标事件的类型。

可选地,处理器710用于:

识别出目标语句中的命名实体,命名实体为目标事件中的参与对象;

对命名实体进行匿名化处理;

将匿名化处理后的目标语句划分为词序列,以得到目标语句中的每个词语。

可选地,处理器710用于:

对目标语句进行预处理,以得到目标语句中的每个词语;

确定目标语句中属于触发词的第一词语;

根据第一词语所对应的备选类型,以及第二词语确定目标事件的类型,第二词语为第一词语的上下文词语。

另外,用于事件识别模型训练的计算机设备也可以参阅图9进行理解。在用于事件识别模型训练过程中:

处理器710用于:

获取用于模型训练的样本集合,样本集合中包括多个样本事件各自对应的样本语句和多个样本事件各自的类型,每个样本语句中包括触发词和触发词的上下文词语;

根据每个样本语句训练第一事件识别模型,以得到第二事件识别模型,第二事件识别模型用于通过目标语句中的触发词和目标语句中的触发词的上下文词语确定目标事件的类型,目标语句为用于识别目标事件所属类型的语句。

可选地,处理器710用于:

对第一样本语句进行预处理,以得到第一样本语句中的每个词语,第一样本语句为每个样本语句中的任意一个样本语句;

对每个词语进行向量转换处理，以得到每个词语的输入层向量；

根据每个词语的输入层向量，以及第一样本语句所对应的样本事件的类型训练第一事件识别模型，以得到第二事件识别模型。

可选地，处理器710用于：

将每个词语的输入层向量转换为隐藏层向量，每个词语包括第一词语，第一词语为每个词语中的触发词；

根据每个第二词语的隐藏层向量、第一词语的隐藏层向量以及任意一个第二词语的隐藏层向量，确定任意一个第二词语的关注度权重，第二词语为除第一词语之外的词语；

根据每个第二词语的输入层向量和每个第二词语的关注度权重确定第一词语的上下文向量；

根据第一词语的上下文向量和第一词语的输入层向量，以及第一样本语句所对应的样本事件的类型，确定第一事件识别模型的参数，以得到第二事件识别模型。

可选地，处理器710用于：

识别出第一样本语句中的命名实体，命名实体为目标事件中的参与对象；

对命名实体进行匿名化处理；

将匿名化处理后的第一样本语句划分为词序列，以得到第一样本语句中的每个词语。

上对计算机设备70的描述可以参阅图1至图6部分的描述进行理解，本处不再重复赘述。

在上述实施例中，可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用软件实现时，可以全部或部分地以计算机程序产品的形式实现。

所述计算机程序产品包括一个或多个计算机指令。在计算机上加载和执行所述计算机程序指令时，全部或部分地产生按照本申请实施例所述的流程或功

能。所述计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。所述计算机指令可以存储在计算机可读存储介质中，或者从一个计算机可读存储介质向另一计算机可读存储介质传输，例如，所述计算机指令可以从一个网站站点、计算机、服务器或数据中心通过有线（例如同轴电缆、光纤、数字用户线（DSL））或无线（例如红外、无线、微波等）方式向另一个网站站点、计算机、服务器或数据中心进行传输。所述计算机可读存储介质可以是计算机能够存储的任何可用介质或者是包含一个或多个可用介质集成的服务器、数据中心等数据存储设备。所述可用介质可以是磁性介质，（例如，软盘、硬盘、磁带）、光介质（例如，DVD）、或者半导体介质（例如固态硬盘Solid State Disk(SSD)）等。

本领域普通技术人员可以理解上述实施例的各种方法中的全部或部分步骤是可以通过程序来指令相关的硬件来完成，该程序可以存储于一计算机可读存储介质中，存储介质可以包括：ROM、RAM、磁盘或光盘等。

以上对本申请实施例所提供的事件识别的方法、事件识别模型训练的方法、装置、设备以及计算机可读存储介质进行了详细介绍，本文中应用了具体个例对本申请的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本申请的方法及其核心思想；同时，对于本领域的一般技术人员，依据本申请的思想，在具体实施方式及应用范围上均会有改变之处，综上所述，本说明书内容不应理解为对本申请的限制。

工业实用性

本申请采用了获取用于识别目标事件所属类型的目标语句，基于事件识别模型处理所述目标语句，以得到所述目标事件的类型，输出所述目标事件的类型的方法，从而可以在事件识别过程中结合触发词和触发词的上下文词语进行识别，采用这种词分类的方法进行事件识别，不仅可以提高识别效率，而且结合上下文词语进行识别，可以提高事件识别的准确度。

权 利 要 求 书

1. 一种事件识别的方法，包括：

终端设备获取用于识别目标事件所属类型的目标语句；

所述终端设备基于事件识别模型处理所述目标语句，以得到所述目标事件的类型，其中，所述事件识别模型用于通过所述目标语句中的触发词和所述触发词的上下文词语确定所述目标事件的类型，所述触发词用于指示所述目标事件的备选类型，所述备选类型包括所述目标事件的类型；

所述终端设备输出所述目标事件的类型。

2. 根据权利要求1所述的方法，其中，所述终端设备基于事件识别模型处理所述目标语句，以得到所述目标事件的类型，包括：

所述终端设备对所述目标语句进行预处理，以得到所述目标语句中的每个词语；

所述终端设备对所述每个词语进行向量转换处理，以得到所述每个词语的输入层向量；

所述终端设备根据所述每个词语的输入层向量，确定所述目标事件的类型。

3. 根据权利要求2所述的方法，其中，所述终端设备根据所述每个词语的输入层向量，确定所述目标事件的类型，包括：

所述终端设备将所述每个词语的输入层向量转换为隐藏层向量，所述每个词语包括第一词语，所述第一词语为所述每个词语中的任一词语；

所述终端设备将所述第一词语作为触发词时，根据每个第二词语的隐藏层向量、所述第一词语的隐藏层向量以及任意一个第二词语的隐藏层向量，确定所述任意一个第二词语的关注度权重，所述第二词语为除所述第一词语之外的词语；

所述终端设备根据每个第二词语的输入层向量和所述每个第二词语的关注度权重确定所述第一词语的上下文向量；

所述终端设备根据所述第一词语的上下文向量和所述第一词语的输入层向量，确定所述目标事件的类型。

4. 根据权利要求3所述的方法，其中，所述终端设备根据所述第一词语的上下文向量和所述第一词语的输入层向量，确定所述目标事件的类型，包括：

所述终端设备根据所述第一词语的上下文向量、所述第一词语的输入层向量与所述事件识别模型的参数，确定所述第一词语为所述触发词时，所述目标事件属于每个备选类型的置信度；

所述终端设备根据所述目标事件属于所述每个备选类型的置信度，确定所述目标事件的类型。

5. 根据权利要求2-4任一所述的方法，其中，所述终端设备对所述目标语句进行预处理，以得到所述目标语句中的每个词语，包括：

所述终端设备识别出所述目标语句中的命名实体，所述命名实体为所述目标事件中的参与对象；

所述终端设备对所述命名实体进行匿名化处理；

所述终端设备将匿名化处理后的所述目标语句划分为词序列，以得到所述目标语句中的每个词语。

6. 根据权利要求1所述的方法，其中，所述终端设备基于事件识别模型处理所述目标语句，以得到所述目标事件的类型，包括：

所述终端设备对所述目标语句进行预处理，以得到所述目标语句中的每个词语；

所述终端设备确定所述目标语句中属于触发词的第一词语；

所述终端设备根据所述第一词语所对应的备选类型，以及第二词语确定所述目标事件的类型，所述第二词语为所述第一词语的上下文词语。

7. 一种事件识别模型训练的方法，包括：

计算机设备获取用于模型训练的样本集合，所述样本集合中包括多个样本事件各自对应的样本语句和所述多个样本事件各自的类型，每个样本语句中包括触发词和所述触发词的上下文词语；

所述计算机设备根据所述每个样本语句训练第一事件识别模型，以得到第二事件识别模型，所述第二事件识别模型用于通过目标语句中的触发词和所述目标语句中的触发词的上下文词语确定目标事件的类型，所述目标语句为用于识别目标事件所属类型的语句。

8. 根据权利要求7所述的方法，其中，所述计算机设备根据所述每个样本语句训练第一事件识别模型，以得到第二事件识别模型，包括：

所述计算机设备对第一样本语句进行预处理,以得到所述第一样本语句中的每个词语,所述第一样本语句为所述每个样本语句中的任意一个样本语句;

所述计算机设备对所述每个词语进行向量转换处理,以得到所述每个词语的输入层向量;

所述计算机设备根据所述每个词语的输入层向量,以及所述第一样本语句所对应的样本事件的类型训练所述第一事件识别模型,以得到所述第二事件识别模型。

9. 根据权利要求8所述的方法,其中,所述计算机设备根据所述每个词语的输入层向量,以及所述第一样本语句所对应的样本事件的类型训练所述第一事件识别模型,以得到所述第二事件识别模型,包括:

所述计算机设备将所述每个词语的输入层向量转换为隐藏层向量,所述每个词语包括第一词语,所述第一词语为所述每个词语中的触发词;

所述计算机设备根据每个第二词语的隐藏层向量、所述第一词语的隐藏层向量以及任意一个第二词语的隐藏层向量,确定所述任意一个第二词语的关注度权重,所述第二词语为除所述第一词语之外的词语;

所述计算机设备根据每个第二词语的输入层向量和所述每个第二词语的关注度权重确定所述第一词语的上下文向量;

所述计算机设备根据所述第一词语的上下文向量和所述第一词语的输入层向量,以及所述第一样本语句所对应的样本事件的类型,确定所述第一事件识别模型的参数,以得到所述第二事件识别模型。

10. 根据权利要求8或9所述的方法,其中,所述计算机设备对第一样本语句进行预处理,以得到所述第一样本语句中的每个词语,包括:

所述计算机设备识别出所述第一样本语句中的命名实体,所述命名实体为所述目标事件中的参与对象;

所述计算机设备对所述命名实体进行匿名化处理;

所述计算机设备将匿名化处理后的所述第一样本语句划分为词序列,以得到所述第一样本语句中的每个词语。

11. 一种事件识别的装置,安装在终端设备中,包括:

获取单元,设置为获取用于识别目标事件所属类型的目标语句;

处理单元, 设置为基于事件识别模型处理所述获取单元获取的所述目标语句, 以得到所述目标事件的类型, 其中, 所述事件识别模型用于通过所述目标语句中的触发词和所述触发词的上下文词语确定所述目标事件的类型, 所述触发词用于指示所述目标事件的备选类型, 所述备选类型包括所述目标事件的类型;

输出单元, 设置为输出所述处理单元确定的所述目标事件的类型。

12. 一种事件识别模型训练的装置, 安装在计算机设备中, 包括:

获取单元, 设置为获取用于模型训练的样本集合, 所述样本集合中包括多个样本事件各自对应的样本语句和所述多个样本事件各自的类型, 每个样本语句中包括触发词和所述触发词的上下文词语;

模型训练单元, 设置为根据所述获取单元获取的所述每个样本语句训练第一事件识别模型, 以得到第二事件识别模型, 所述第二事件识别模型用于通过目标语句中的触发词和所述目标语句中的触发词的上下文词语确定目标事件的类型, 所述目标语句为用于识别目标事件所属类型的语句。

13. 一种终端设备, 所述终端设备包括: 输入/输出 (I/O) 接口、处理器和存储器, 所述存储器中存储有程序指令;

所述处理器用于执行存储器中存储的程序指令, 执行如权利要求1-6任一所述的方法。

14. 一种计算机设备, 所述计算机设备包括: 输入/输出 (I/O) 接口、处理器和存储器, 所述存储器中存储有程序指令;

所述处理器用于执行存储器中存储的程序指令, 执行如权利要求7-10任一所述的方法。

15. 一种计算机可读存储介质, 包括指令, 当所述指令在计算机设备上运行时, 使得所述计算机设备执行如权利要求1-6中任一项所述的方法或权利要求7-10任一所述的方法。

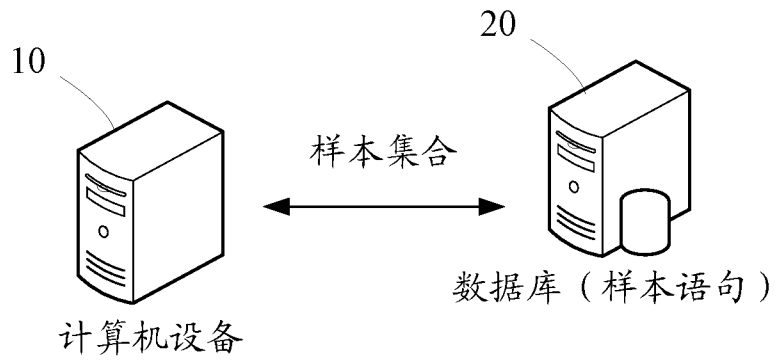


图 1

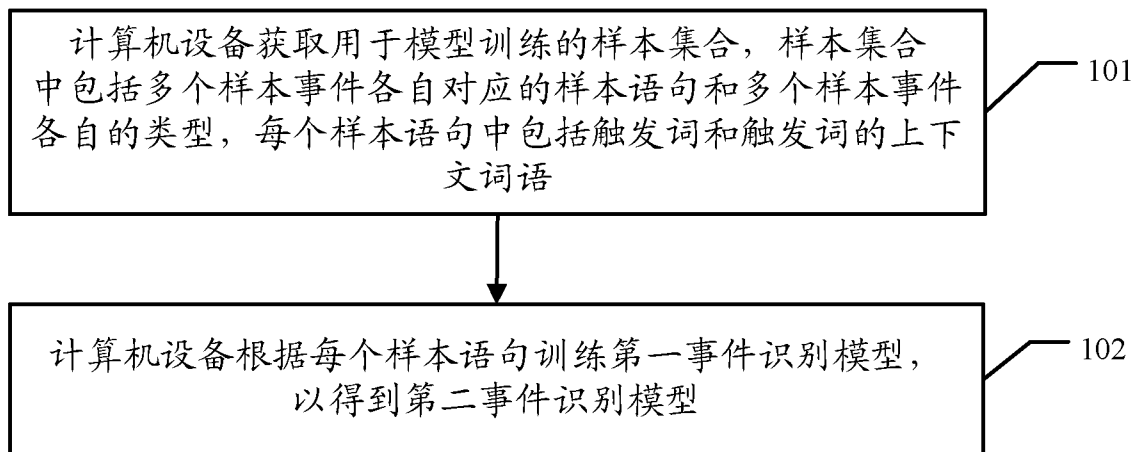


图 2

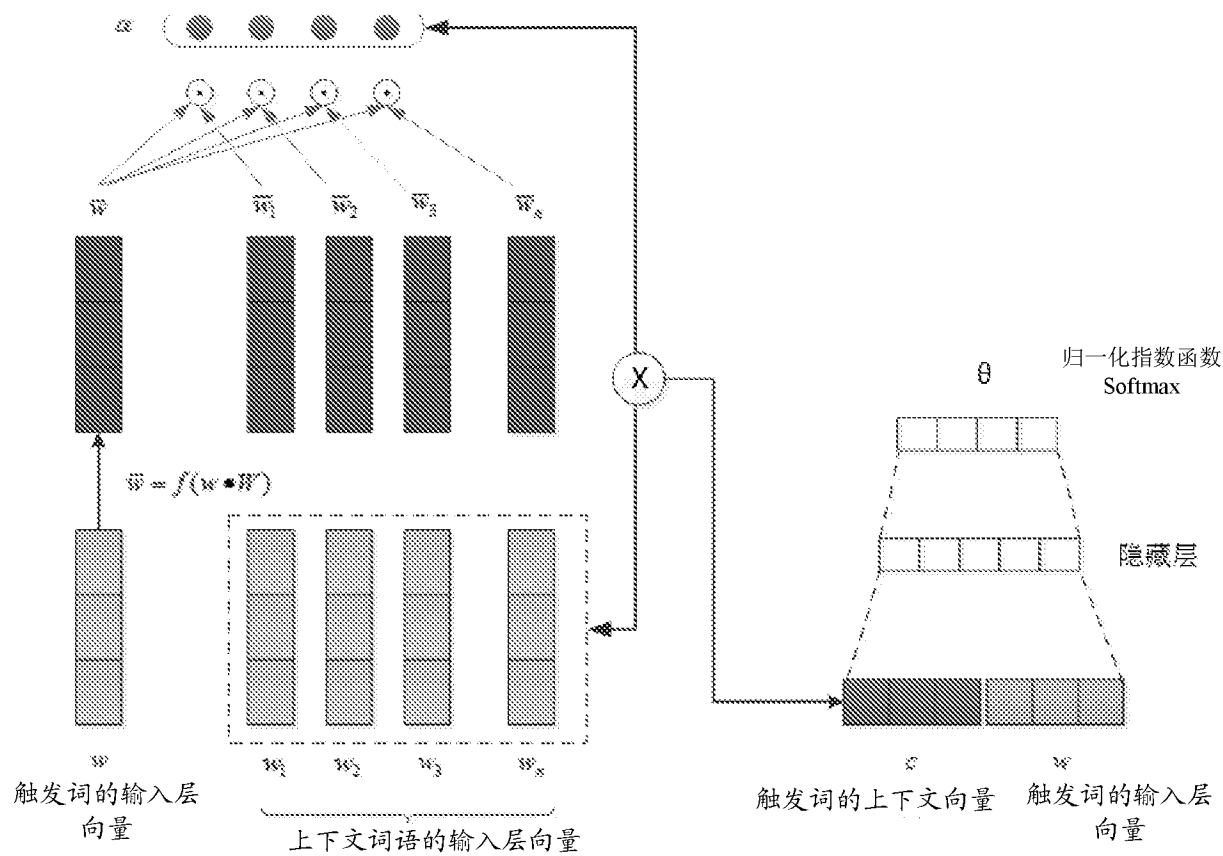


图 3

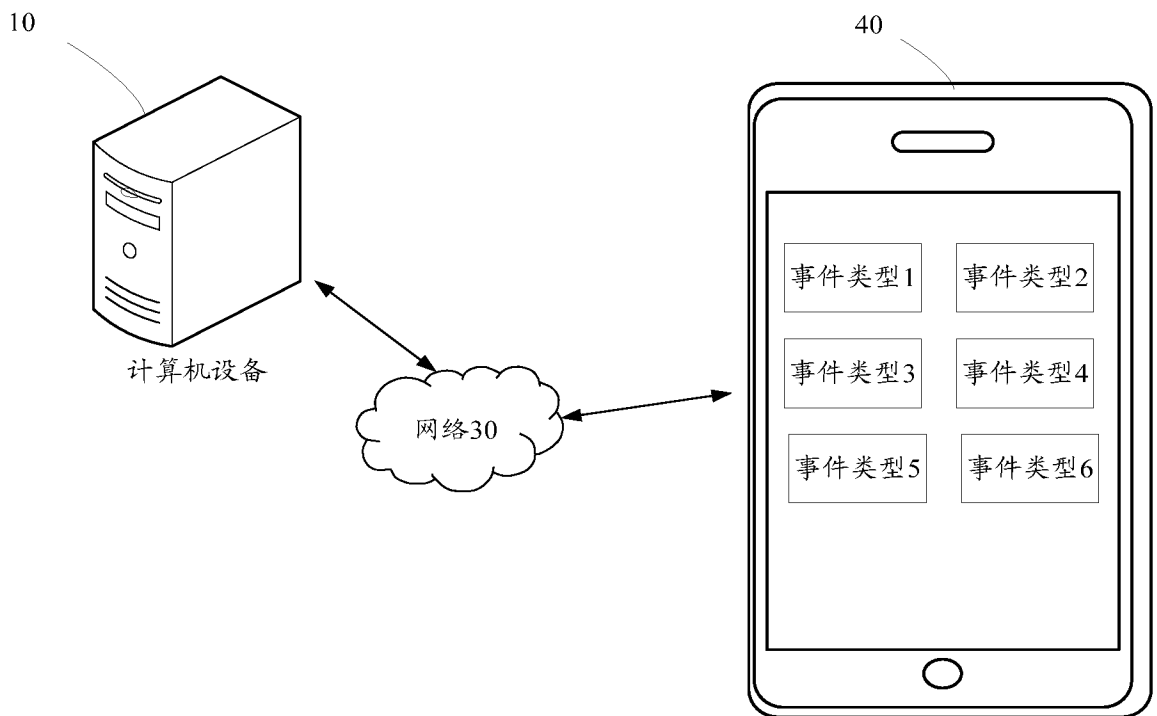


图 4

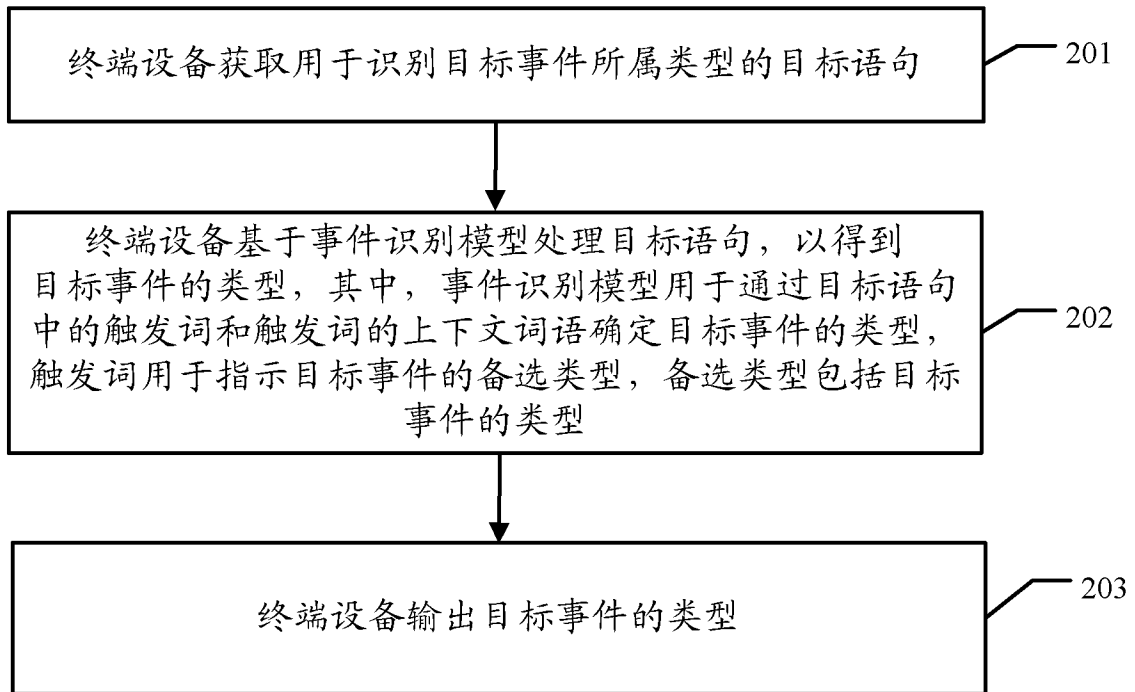


图 5

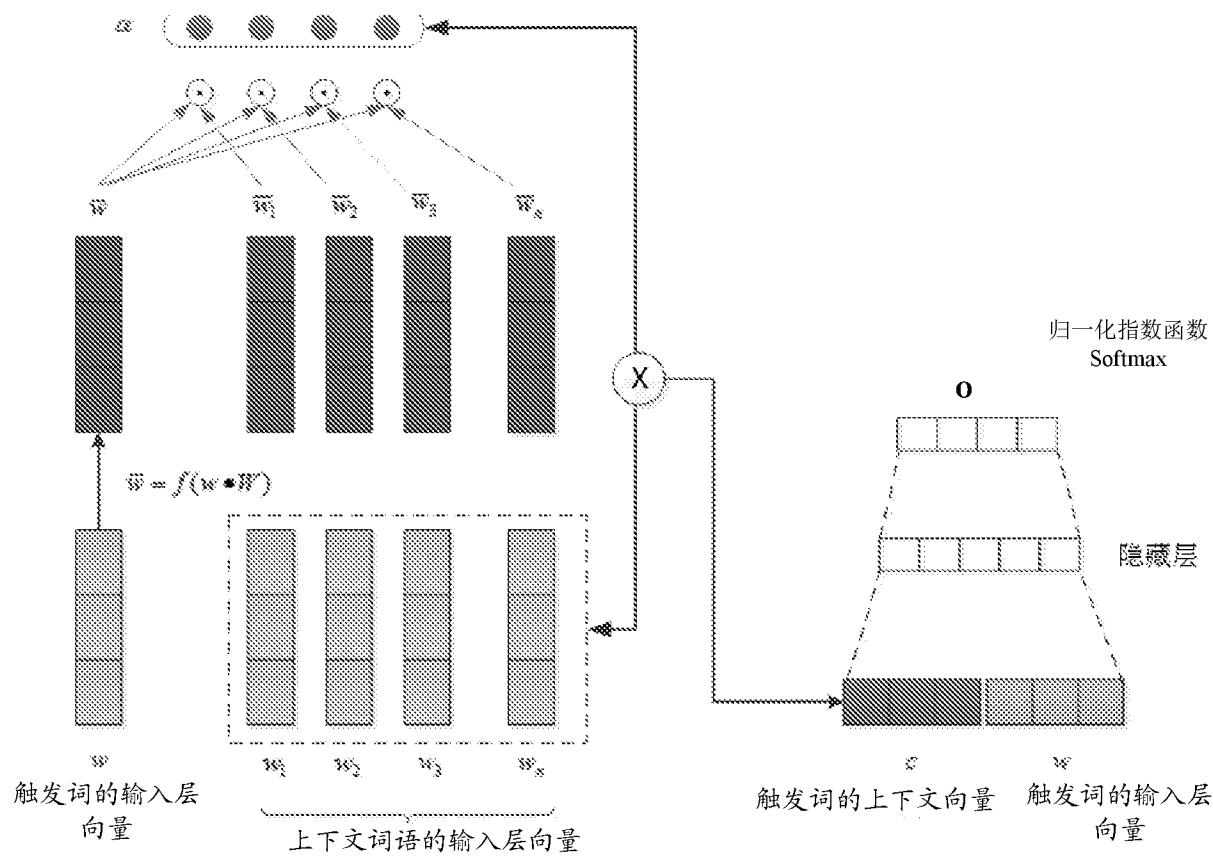


图 6

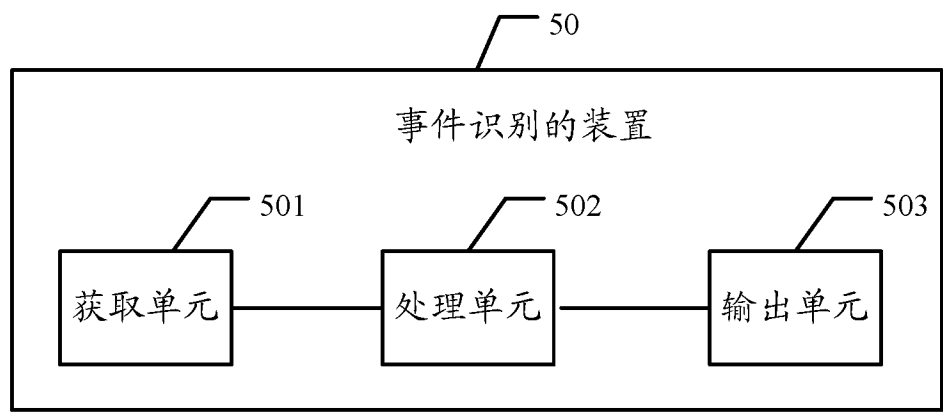


图 7

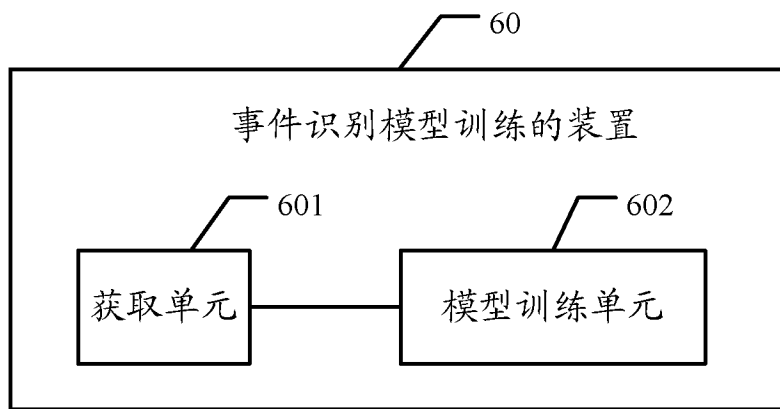


图 8

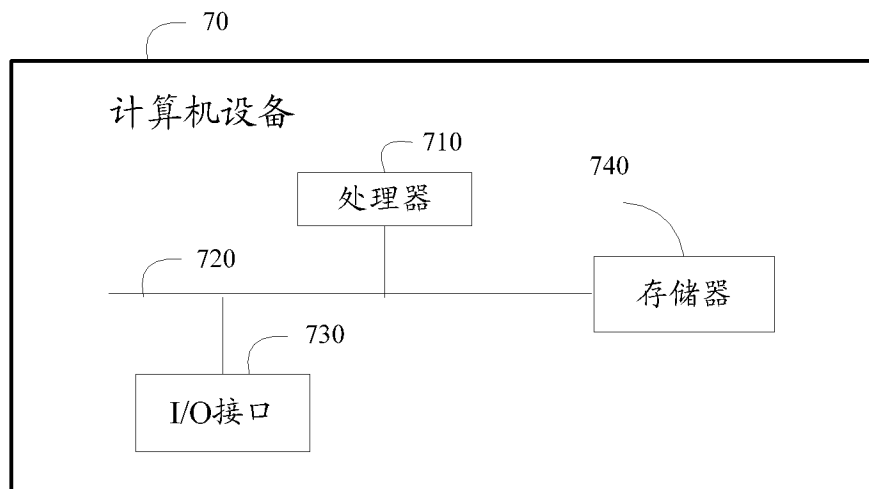


图 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/089068

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/35(2019.01)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)

CNPAT, WPI, EPODOC, CNKI, IEEE: 事件, 识别, 模型, 语句, 触发词, 短语, 上下文, 训练, 类型, event, recognition, model, text, word, trigger, phrase, context, train, type

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 106933800 A (CAPITAL NORMAL UNIVERSITY ET AL.) 07 July 2017 (2017-07-07) description, paragraphs [0036]-[0039], [0060]-[0065] and [0095]	1, 6, 11, 13
Y	CN 104298665 A (SOOCHOW UNIVERSITY) 21 January 2015 (2015-01-21) description, paragraphs [0078] and [0116]-[0140]	1, 6, 11, 13
X	CN 104298665 A (SOOCHOW UNIVERSITY) 21 January 2015 (2015-01-21) description, paragraphs [0078] and [0116]-[0140]	7, 12, 14, 15
A	CN 106445999 A (TIANJIN HYLANDA INFORMATION TECHNOLOGY CO., LTD.) 22 February 2017 (2017-02-22) entire document	1-15
A	CN 102521220 A (CENTRAL CHINA NORMAL UNIVERSITY) 27 June 2012 (2012-06-27) entire document	1-15
A	CN 102298635 A (SOOCHOW UNIVERSITY) 28 December 2011 (2011-12-28) entire document	1-15

☐ 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 August 2019

Date of mailing of the international search report

28 August 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/CN2019/089068

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	106933800	A	07 July 2017	None	
CN	104298665	A	21 January 2015	None	
CN	106445999	A	22 February 2017	None	
CN	102521220	A	27 June 2012	None	
CN	102298635	A	28 December 2011	None	

国际检索报告

国际申请号

PCT/CN2019/089068

A. 主题的分类

G06F 16/35 (2019.01) 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))

CNPAT, WPI, EPDOC, CNKI, IEEE: 事件, 识别, 模型, 语句, 触发词, 短语, 上下文, 训练, 类型, event, recognition, model, text, word, trigger, phrase, context, train, type

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 106933800 A (首都师范大学 等) 2017年 7月 7日 (2017 - 07 - 07) 说明书第36-39、60-65、95段	1、6、11、13
Y	CN 104298665 A (苏州大学) 2015年 1月 21日 (2015 - 01 - 21) 说明书第78、116-140段	1、6、11、13
X	CN 104298665 A (苏州大学) 2015年 1月 21日 (2015 - 01 - 21) 说明书第78、116-140段	7、12、14、15
A	CN 106445999 A (天津海量信息技术股份有限公司) 2017年 2月 22日 (2017 - 02 - 22) 全文	1-15
A	CN 102521220 A (华中师范大学) 2012年 6月 27日 (2012 - 06 - 27) 全文	1-15
A	CN 102298635 A (苏州大学) 2011年 12月 28日 (2011 - 12 - 28) 全文	1-15

☐ 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 8月 13日

国际检索报告邮寄日期

2019年 8月 28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

受权官员

边臻

传真号 (86-10)62019451

电话号码 86-(10)-53961419

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/089068

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	106933800	A	2017年 7月 7日	无	
CN	104298665	A	2015年 1月 21日	无	
CN	106445999	A	2017年 2月 22日	无	
CN	102521220	A	2012年 6月 27日	无	
CN	102298635	A	2011年 12月 28日	无	

表 PCT/ISA/210 (同族专利附件) (2015年1月)