

(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 7 月 23 日 (23.07.2020)



(10) 国际公布号
WO 2020/147609 A1

(51) 国际专利分类号:
G10L 15/22 (2006.01)

(21) 国际申请号: PCT/CN2020/070581

(22) 国际申请日: 2020 年 1 月 7 日 (07.01.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910047340.2 2019年1月18日 (18.01.2019) CN

(71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 开曼群岛
大开曼资本大厦一座四层 847 号 邮箱,
Grand Cayman (KY)。

(72) 发明人: 张帆(ZHANG, Fan); 中国浙江省杭州市
余杭区文一西路969号3号楼5楼阿里巴巴集团
法务部, Zhejiang 311121 (CN)。 郑梓豪(ZHENG,
Zihao); 中国浙江省杭州市余杭区文一西路
969号3号楼5楼阿里巴巴集团法务部, Zhejiang
311121 (CN)。 胡于响(HU, Yuxiang); 中国浙江
省杭州市余杭区文一西路969号3号楼5楼
阿里巴巴集团法务部, Zhejiang 311121 (CN)。
姜飞俊(JIANG, Feijun); 中国浙江省杭州市余
杭区文一西路969号3号楼5楼阿里巴巴集
团法务部, Zhejiang 311121 (CN)。

(74) 代理人: 北京三友知识产权代理
有限公司 (BEIJING SANYOU INTELLECTUAL
PROPERTY AGENCY LTD.); 中国北京市金

(54) Title: SPEECH RECOGNITION METHOD AND APPARATUS

(54) 发明名称: 语音识别方法及装置

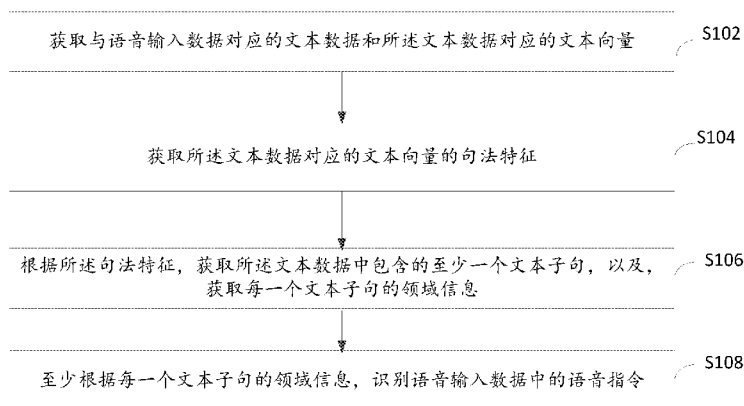


图 1

- S102 Obtain text data corresponding to speech input data and a text vector corresponding to the text data
- S104 Obtain syntax characteristics of the text vector corresponding to the text data
- S106 According to the syntax characteristics, obtain at least one text clause comprised in the text data, and obtain domain information of each text clause
- S108 Recognize a speech instruction in the speech input data at least according to the domain information of each text clause

(57) Abstract: A speech recognition method and apparatus. The method comprises: obtaining text data corresponding to speech input data and a text vector corresponding to the text data (S102); obtaining syntax characteristics of the text vector (S104); according to the syntax characteristics, obtaining at least one text clause comprised in the text data, and obtaining domain information of each text clause (S106); and recognizing a speech instruction in the speech input data at least according to the domain information of each text clause (S108). According to the method, an operating burden of a user is reduced, and the degree for the intelligent processing of a user speech instruction by an intelligent speech device is also improved.



融街 35 号国际企业大厦 A 座 16 层,
Beijing 100033 (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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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: 一种语音识别方法及装置, 该方法包括: 获取与语音输入数据对应的文本数据和文本数据对应的文本向量 (S102); 获取文本向量的句法特征 (S104); 根据句法特征, 获取文本数据中包含的至少一个文本子句, 以及, 获取每一个文本子句的领域信息 (S106); 至少根据每一个文本子句的领域信息, 识别语音输入数据中的语音指令 (S108)。该方法减轻了用户的操作负担, 也提高了智能语音设备对用户语音指令的智能化处理程度。

语音识别方法及装置

本申请要求 2019 年 01 月 18 日递交的申请号为 201910047340.2、发明名称为“语音识别方法及装置”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

本发明实施例涉及计算机技术领域，尤其涉及一种语音识别方法及装置。

背景技术

智能设备是传统电气设备与计算机技术、数据处理技术、控制技术、传感器技术、
10 网络通信技术、电力电子技术等相结合的产物。在各种智能设备中，智能语音设备是其中重要的一个分支。

通过智能语音设备，用户仅需语音即可实现对各种智能设备的控制，包括对智能语音设备自身和智能语音设备所控制的其它智能设备的控制。目前，在用户与智能语音设备进行交互的过程中，每一次对智能语音设备的控制都需要使用唤醒词，然后紧接着说出语音指令以完成用户的意图。例如：“天猫精灵，打开灯”、“天猫精灵，播放音乐”
15 等等，可见，在该交互中，用户每次都需要使用“天猫精灵”这一唤醒词来唤醒智能语音设备，以进行相应的操作和控制。而在“你为什么这么晚回家啊？请打开卧室的灯”这一语句中，“你为什么这么晚回家啊？”是用户之间的交互，而“请打开卧室的灯”则是对智能语音设备的控制指令。对于这类复杂的且无唤醒词的混合指令，目前的智能
20 语音设备则无法进行处理。

但是，这种使用唤醒词唤醒智能语音设备的方式，一方面，用户的每条指令都必须使用唤醒词，增加了用户的操作负担，也使得智能语音设备对用户语音指令的智能化处理程度较低；另一方面，智能语音设备需要重复地对唤醒词进行处理，也增加了智能语音设备的处理负担。

25

发明内容

有鉴于此，本发明实施例提供一种语音识别方案，以解决上述问题。

根据本发明实施例的第一方面，提供了一种语音识别方法，包括：获取与语音输入数据对应的文本数据和所述文本数据对应的文本向量；获取所述文本向量的句法特征；
30 根据所述句法特征，获取所述文本数据中包含的至少一个文本子句，以及，获取每一个

所述文本子句的领域信息；至少根据每一个所述文本子句的领域信息，识别所述语音输入数据中的语音指令。

根据本发明实施例的第二方面，提供了一种语音识别装置，包括：第一获取模块，用于获取与语音输入数据对应的文本数据和所述文本数据对应的文本向量；第二获取模块，用于获取所述文本向量的句法特征；第三获取模块，用于根据所述句法特征，获取所述文本数据中包含的至少一个文本子句，以及，获取每一个所述文本子句的领域信息；识别模块，用于至少根据每一个所述文本子句的领域信息，识别所述语音输入数据中的语音指令。

根据本发明实施例的第三方面，提供了一种智能设备，包括：处理器、存储器、通信接口和通信总线，所述处理器、所述存储器和所述通信接口通过所述通信总线完成相互间的通信；所述存储器用于存放至少一可执行指令，所述可执行指令使所述处理器执行如第一方面所述的语音识别方法对应的操作。

根据本发明实施例的第四方面，提供了一种计算机存储介质，其上存储有计算机程序，该程序被处理器执行时实现如第一方面所述的语音识别方法。

根据本发明实施例提供的语音识别方案，先获取由语音输入数据转换的文本数据和该文本数据对应的文本向量；再通过对文本向量的特征提取获得对应的句法特征；然后，根据句法特征对语音输入数据对应的文本数据进行文本子句的划分及文本子句的领域信息的确定；进而，根据文本子句的领域信息识别语音输入数据中的语音指令。可见，通过本发明实施例的方案，使得智能语音设备更加适用于实际的使用环境，无需用户再使用唤醒词唤醒智能语音设备，不管是用户使用纯语音指令的语音输入数据，还是使用包含语音指令和其它语音数据的混合语音输入数据，都能对语音输入数据进行有效的子句划分，并识别其中包含的语音指令，进而，后续可以通过识别的语音指令对智能语音设备进行操作和控制。

因无需使用唤醒词唤醒智能语音设备，减轻了用户的操作负担，也提高了智能语音设备对用户语音指令的智能化处理程度；并且，智能语音设备也无需再针对唤醒词进行处理，减轻了智能语音设备的数据处理负担。

附图说明

为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术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

发明实施例中记载的一些实施例，对于本领域普通技术人员来讲，还可以根据这些附图获得其他的附图。

图 1 为根据本发明实施例一的一种语音识别方法的步骤流程图；

图 2 为根据本发明实施例二的一种语音识别方法的步骤流程图；

5 图 3 为图 2 所示实施例中的一种神经网络模型的结构示意图；

图 4 为根据本发明实施例三的一种语音识别装置的结构框图；

图 5 为根据本发明实施例四的一种语音识别装置的结构框图；

图 6 为根据本发明实施例五的一种智能设备的结构示意图。

10 具体实施方式

为了使本领域的人员更好地理解本发明实施例中的技术方案，下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅是本发明实施例一部分实施例，而不是全部的实施例。基于本发明实施例中的实施例，本领域普通技术人员所获得的所有其他实施例，都应当属于本发明实施例保护的

15 的范围。

下面结合本发明实施例附图进一步说明本发明实施例具体实现。

实施例一

参照图 1，示出了根据本发明实施例一的一种语音识别方法的步骤流程图。

20 本实施例的语音识别方法包括以下步骤：

步骤 S102：获取与语音输入数据对应的文本数据和所述文本数据对应的文本向量。

在智能语音设备的使用场景中，用户可以通过语音对智能语音设备进行操作和控制；智能语音设备以用户发出的语音作为输入生成相应的语音输入数据，并将该语音输入数据转换为对应的文本数据，进而进行相应的处理。本实施例中，除需要将语音输入数据

25 转换为文本数据外，还会获取该文本数据对应的文本向量，以通过向量形式表征该文本数据，且便于后续的处理。其中，将语音输入数据转换为对应的文本数据，以及，获取该文本数据对应的文本向量的具体实现均可由本领域技术人员根据实际需要采用任意适当的方式实现，本发明实施例对此不作限制。

例如，可以采用卷积神经网络模型或者 BP 神经网络模型或者隐马尔科夫模型 HMM

30 或者多频带谱减法等等方式实现将语音输入数据转换为文本数据；又例如，可以基于深

度学习的方式（如 word2vec 方式）、或者基于图的方式（如 textrank 方式）、或者基于主题模型的方式（如 LDA 方式）、或者基于统计的方式（如 bag of words 方式）等等实现文本数据对应的文本向量的获取。

步骤 S104：获取所述文本数据对应的文本向量的句法特征。

5 本发明实施例中，文本向量的句法特征可以表征文本向量所对应的文本数据中的字词之间的依存关系和语义信息，句法特征可以通过句法特征向量表达。在具体实现中，可以通过卷积神经网络 CNN 模型或者循环神经网络 RNN 模型对所述文本向量进行特征提取，以获取所述文本向量的句法特征。但不限于此，在实际应用中，本领域技术人员也可以采用其它适当方式获取文本向量的句法特征，如文本分类或其它方式等。

10 步骤 S106：根据所述句法特征，获取所述文本数据中包含的至少一个文本子句，以及，获取每一个文本子句的领域信息。

本发明实施例中，语音输入数据对应的文本数据中，包含一个或多个文本子句，当包含一个文本子句时，该文本子句可以是语音指令对应的文本子句，也可以是其它语音数据对应的文本子句；当包含多个文本子句时，该多个文本子句可以均为语音指令对应的文本子句，该多个文本子句也可以均为其它语音数据对应的文本子句，如用户说的与语音指令无关的句子，该多个文本子句还可以为语音指令对应的文本子句与其它语音数据对应的文本子句的混合，如在一个复杂的多人场景中，用户 A 和用户 B 在交流的同时，向智能语音设备发出的语音指令，如“你为什么这么晚回家啊？请打开卧室的灯”，其中，前半句“你为什么这么晚回家啊”会被识别为其它语音数据对应的文本子句，而后
15 20 后半句“请打开卧室的灯”则会被识别为语音指令对应的文本子句。

在实际应用中，在获得了文本向量对应的句法特征后，即可根据句法特征确定所述文本数据中的一个或多个文本子句。其中，获取文本子句的方式可以与获取句法特征的方式相适应，例如，当采用 CNN 模型或 RNN 模型获取所述文本向量的句法特征时，可以依据句法特征对所述文本数据进行序列标注，根据序列标注的结果获得一个或多个文
25 本子句。

此外，本发明实施例中，还会根据文本向量的句法特征，获取每一个文本子句的领域信息。如，通过机器学习算法或者神经网络模型，由文本向量的句法特征获取其对应的文本子句的领域信息，其中，所述领域信息中包括语音指令对应的领域的信息。

需要说明的是，本发明实施例中，若无特别说明，“多个”、“多种”等与“多”
30 有关的数量意指两个及两个以上。

步骤 S108: 至少根据每一个文本子句的领域信息, 识别语音输入数据中的语音指令。

在语音输入数据包括有语音指令的情况下, 其对应的文本数据包含的一个或多个文本子句中, 应当有文本子句的领域信息指示该文本子句对应的那部分语音输入数据为语音指令, 据此, 即可从语音输入数据中识别出语音指令。

5 例如, 在“你为什么这么晚回家啊? 请打开卧室的灯”这一语音输入数据中, 根据对其对应的文本数据及文本数据对应的文本向量的上述处理, 可确定其中的文本子句“请打开卧室的灯”为语音指令。

通过本实施例, 先获取由语音输入数据转换的文本数据和该文本数据对应的文本向量; 再通过对文本向量的特征提取获得对应的句法特征; 然后, 根据句法特征对语音输入数据对应的文本数据进行文本子句的划分及文本子句的领域信息的确定; 进而, 根据文本子句的领域信息识别语音输入数据中的语音指令。可见, 通过本实施例, 使得智能语音设备更加适用于实际的使用环境, 无需用户再使用唤醒词唤醒智能语音设备, 不管是用户使用纯语音指令的语音输入数据, 还是使用包含语音指令和其它语音数据的混合语音输入数据, 都能对语音输入数据进行有效的子句划分, 并识别其中包含的语音指令, 10 进而, 后续可以通过识别的语音指令对智能语音设备进行操作和控制。

因无需使用唤醒词唤醒智能语音设备, 减轻了用户的操作负担, 也提高了智能语音设备对用户语音指令的智能化处理程度; 并且, 智能语音设备也无需再针对唤醒词进行处理, 减轻了智能语音设备的数据处理负担。

本实施例的语音识别方法可以由任意适当的具有数据处理能力的智能语音设备执行, 如, 具有相应功能的各种智能家电等。 20

实施例二

参照图 2, 示出了根据本发明实施例二的一种语音识别方法的步骤流程图。

本实施例的语音识别方法包括以下步骤:

25 步骤 S202: 获取与语音输入数据对应的文本数据和所述文本数据对应的文本向量。

本实施例中, 所述文本数据对应的文本向量包括文本数据中的每一个字词对应的字词向量。其中, 因可能采用的语言的不同, 字词的具体含义也可能不同。例如, 对于像中文、日文、韩文等类似语言体系的文本数据, 一个字词可能是单个的字, 也可能是一个词; 而对于像英文、法文等类似语言体系的文本数据, 一个字词多为一个完整的单词。

30 基于此, 在一种可行方式中, 本步骤可以实现为: 获取语音输入数据, 并生成与语

音输入数据对应的文本数据；生成所述文本数据中的每一个字词对应的字词向量；根据每一个字词对应的字词向量，生成所述文本数据对应的文本向量。其中，根据语音输入数据生成对应的文本数据，以及，生成文本数据中的每一个字词对应的字词向量的具体实现方式均可由本领域技术人员根据实际需求采用任意适当的方式实现，本发明实施例
5 对此不作限制。通过每一个字词对应的字词向量表征文本数据对应的文本向量既可便于对文本数据的处理，也可以有效避免因向量化处理而造成的文本数据的过多信息损失。

步骤 S204：获取所述文本数据对应的文本向量的句法特征。

如实施例一中所述，获取文本向量的句法特征的方式可以有多种，本实施例中采用特征提取的方式，也即，对所述文本数据对应的文本向量进行特征提取，获取所述文本
10 向量的句法特征。

在文本向量包括每一个字词对应的字词向量的情况下，本步骤可以实现为：对所述文本向量中的每一个字词对应的字词向量进行特征提取，获取每一个字词的句法特征。采用特征提取的方式提取到的句法特征，可以更为有效地表征每一个字词向量对应的字词的
15 特性。

步骤 S206：根据所述文本向量的句法特征，获取所述文本数据中包含的至少一个文本子句，以及，获取每一个所述文本子句的领域信息。

其中，在根据所述文本向量的句法特征，获取所述文本数据中包含的至少一个文本子句时，基于前述获取的每一个字词的句法特征，可以根据所述每一个字词的句法特征，获取每一个字词的标签，其中，所述标签包括结束标签；根据每一个字词的标签，获得
20 所述文本数据的序列标注；根据所述序列标注中的结束标签，获取所述文本数据中包含的至少一个文本子句。也即，可以将文本子句的划分问题转换为文本数据的序列标注问题。其中，所述标签的类型可以由本领域技术人员根据实际需求适当设置，但至少包括结束标签。若一个字词被标注为结束标签，则表明从文本数据开头至该字词之间的所有字词组成为一个文本子句，或者，从前一结束标签之后的首个字词至该结束标签对应的
25 字词之间的所有字词组成为一个文本子句。

可选地，所述标签可以包括 B 标签（开始标签，表明当前字词为一个句子的开始）、I 标签（中间标签，表明当前字词在一个句子开始和结尾之间的内部）、E 标签（结束标签，表明当前字词为一个句子的结尾）。若当前文本数据包括多个 E 标签，则表明当前文本数据包括多个子句，可以根据 E 标签对文本子句进行划分；而若当前文本数据仅包
30 括一个 E 标签，则表明当前文本数据仅有一个文本子句，即当前文本数据自身。

通过对字词进行标签标注以形成文本数据的序列标注，进而根据序列标注中的结束标签获得文本子句的方式，可以使得对文本子句的划分更为准确；另外，相对于其它划分文本子句的方式，也简化了划分的操作步骤，降低了划分的实现成本。

而在根据所述文本向量的句法特征，获取每一个文本子句的领域信息时，可以根据
5 所述文本向量的句法特征，获取每一个文本子句对应的领域特征；对每一个文本子句的领域特征，在每个特征维度上进行最大特征值提取，生成每一个文本子句的领域特征向量；根据每一个文本子句的领域特征向量，确定当前文本子句的领域信息。通过这种方式，可以获得每一个文本子句最有效的特征表达，且每一个文本子句的特征表达具有相同的向量长度，以便于后续处理。

10 在一种可行方式中，所述根据所述文本向量的句法特征，获取每一个文本子句对应的领域特征可以包括：根据所述文本向量的句法特征，获取所述文本向量的领域特征；根据每一个文本子句所包含的字词的信息，从所述文本向量的领域特征中获取每一个文本子句对应的领域特征。也即，先根据整个所述文本数据对应的文本向量，获取对应的总领域特征，进而依据文本子句中的字词的信息，从总领域特征中获得每一个文本子句
15 的领域特征。由此，既保证了每一个文本子句的领域特征与总领域特征的一致性，也简化了获取文本子句的领域特征的实现。

步骤 S208：至少根据每一个文本子句的领域信息，识别语音输入数据中的语音指令。

本发明实施例中，属于语音指令的文本子句对应有设定的领域信息，当某一文本子句的领域信息与该设定的领域信息一致时，则可以将该文本子句确定为语音指令对应的
20 文本子句。可选地，还可以设置其它的领域信息，该其它的领域信息可以是指示文本子句为非语音指令的统一的领域信息，也可以再将其其它的领域信息进行细分，以指示出文本子句的具体领域，如，交互领域，等等。

步骤 S210：根据识别出的语音指令，对智能语音设备进行所述语音指令所指示的操作。

25 其中，所述操作可以是任意适当的操作，如，指示智能语音设备开启或关闭相应功能，如打开空调、关闭电灯等，或者，指示智能语音设备进行查询的操作，如，查询并播放某首歌曲、查询并播放某地天气，等等，本发明实施例对语音指令所指示的具体操作不作限制。

30 如前所述，本发明实施例提供的语音识别方案可以通过多种适当的方式实现，在一

种可行的方式中，可以通过神经网络模型实现该语音识别方案中的部分或全部方案。以下，以卷积神经网络 CNN 模型为例，对本实施例的上述过程进行说明。

一种 CNN 模型的结构如图 3 所示，其包括：输入部分 A、特征提取部分 B、句边界探测部分 C、和领域分类部分 D。

5 其中：

输入部分 A 可以为 CNN 的输入层，用于接收输入的文本向量，如语音输入数据对应的文本数据的文本向量。

10 特征提取部分 B 中设置有多个卷积层，本实施例中，设定至少 12 个卷积层，以提高特征提取的精度。可选地，还可以在特征提取部分 B 中设置批规范化层（batch normalization layer）、激活层（activation layer）、卷积层（convolutional layer），其中，还可以对卷积层进行残差（residual）处理。通过设置批规范化层，可以优化 CNN 模型的数据处理速度；所述激活层可采用线性门函数进行非线性化处理，激活层采用线性门函数可以提升对文本向量的非线性化特征转换效果，当然，其它的激活函数也同样适用；通过对卷积层进行残差处理，将文本数据对应的原始的文本向量与当前卷积层输出的句法特征相结合后输出，以优化梯度回传效果，并提升特征提取效果。

句边界探测部分 C 可选地，可依次包括批规范化层、卷积层和输出层，其中，输出层采用 Softmax 函数作为损失函数。通过句边界探测部分 C 可以获得文本向量中每一个字词对应的字词向量的标签，进而获得整个文本数据的序列标注，根据序列标注中的结束标签如 E 标签即可确定文本子句的划分。

20 领域分类部分 D 可选地，可依次包括批规范化层、卷积层、池化层和输出层，其中，池化层采用一维特征池化（1-D RoI pooling），输出层采用 Softmax 函数作为损失函数。领域分类部分 D 根据文本子句的划分结果和每个文本子句的领域信息，即可识别出其中的语音指令对应的文本子句。

需要说明的是，如图 3 所示，本实施例的 CNN 模型中的句边界探测部分 C 和领域分类部分 D 共享特征提取部分 B 提取的句法特征，以提升 CNN 模型数据处理效率，节约 CNN 模型实现成本。

基于图 3 所示的 CNN 模型，以智能语音设备为智能音箱、语音输入数据为“你为什么这么晚回家啊？请打开卧室的灯”为例，则相应的语音识别流程包括：

（1）将语音输入数据转换为文本数据，并获取该文本数据对应的文本向量。

30 该部分是文本向量输入 CNN 模型前，对数据的转换和处理。以用户发出“你为什么

这么晚回家啊？请打开卧室的灯”的语音为例，则在本部分中，需要将“你为什么这么晚回家啊？请打开卧室的灯”这一语音输入数据转换为文本数据，并将其中的每一个字转换为D维的向量，其中，D的具体数值可以由本领域技术人员根据实际需求适当设置。

由此，可以生成N个D维的向量，其中N为字词的数量，本示例中包括17个字，

5 因此，N为17。这N个D维的向量即为文本数据对应的文本向量。

(2) 通过CNN模型的输入部分接收所述文本数据对应的文本向量。

例如，通过CNN模型的输入层接收上述生成的N个D维的向量。

(3) 通过CNN模型的特征提取部分对所述文本向量进行特征提取，获取所述文本向量的句法特征。

10 包括：对输入的向量进行批规范化操作，生成规范化的向量；对所述规范化的向量进行非线性化处理；通过卷积层对非线性处理后的所述向量进行特征提取，获得初始特征；对所述初始特征进行残差分析处理，根据所述残差分析处理结果获得所述向量的句法特征并输出；返回所述对输入的向量进行批规范化操作的步骤继续执行，直至获得所述文本向量的句法特征。可选地，当设置有批规范化层时，通过批规范化层对输入的向量进行批规范化操作，生成规范化的向量。其中，输入首个卷积层部分的所述批规范化层的向量为所述文本数据对应的文本向量；输入非首个卷积层部分的所述批规范化层的向量为前一卷积层部分输出的向量。又可选地，当设置有激活层时，通过激活层对所述规范化的向量进行非线性化处理。

15

也即，先将文本向量输入首个批规范化层，依次通过批规范化层、激活层、卷积层
20 对该文本向量进行批规范化操作、非线性化处理、特征提取和残差处理，获得句法特征；接着，将得到的该句法特征输入邻接的下一个批规范化层、激活层、卷积层等依次进行处理，再次获得处理得到的句法特征；再将该处理得到的句法特征输入下一批规范化层、激活层、卷积层进行处理，依次类推，直至获得最终的文本向量的句法特征。

需要说明的是，输入批规范化层的向量可以是前一卷积层经残差处理后的全部的向量如全部的文本向量或全部的句法特征，也可以是前一卷积层残差处理后的每一个字词的向量如文本向量中的每一个字词的向量或每一个字词对应的句法特征。但不管哪种方式，最终获取到的文本向量的句法特征包括每一个字词的句法特征。

25

具体到“你为什么这么晚回家啊？请打开卧室的灯”这一示例，则通过本步骤可以获得其中的每一个字对应的句法特征。

30 (4) 通过CNN模型的句边界探测部分根据特征提取部分输出的句法特征，获取语音

输入数据对应的文本数据中包含的至少一个文本子句。

包括：对所述句法特征进行批规范化操作（可选地，可通过句边界探测部分的批规范化层对所述句法特征进行批规范化操作），生成规范化的句法特征；通过卷积层对所述规范化的句法特征进行特征提取；通过输出层根据特征提取结果确定所述文本数据中的每个字词的标签，根据每个字词的标签获取所述文本数据中包含的至少一个文本子句。

通过句边界探测部分，实现了文本数据的序列标注。例如，用 **B** 表示对应的字处于一个文本片段的开始（即 **B** 为开始标签），**E** 表示对应的字处于一个文本片段的结束（即 **E** 为结束标签），**I** 表示对应的字处于一个文本片段的中间（即 **I** 为中间标签）。根据每一个字词的句法特征，通过句边界探测部分即可得到文本数据中每一个字词上的 **BIE** 概率分布，对于每一个字词，取其 **BIE** 概率分布的最大值对应的标签。如，“你为什么这么晚回家啊？请打开卧室的灯”这一示例中，若“啊”字的 **B** 标签概率为 0.3，**I** 标签概率为 0.1，**E** 标签概率为 0.8，则可以确定“啊”字的标签应为 **E** 标签。根据每一个字词的标签，可得到整个文本数据的序列标注，进而，根据该序列标注中的结束标签，即可得到每一个文本子句的句子边界，从而得到每一个文本子句的范围。

具体到“你为什么这么晚回家啊？请打开卧室的灯”这一示例，则通过本步骤可以获得其序列标注，如，“**BIIIIIIIEBIIIIIE**”，据此，可以获得两个文本子句，即“你为什么这么晚回家啊？”和“请打开卧室的灯”。

(5)通过 CNN 模型的领域分类部分根据所述文本向量的句法特征和每一个文本子句的信息，获取每一个文本子句的领域信息。

包括：对所述文本向量的句法特征进行批规范化操作（可选地，可通过领域分类部分的批规范化层对所述文本向量的句法特征进行批规范化操作），生成规范化的句法特征；通过卷积层对所述规范化的句法特征进行特征映射，获取所述文本向量的领域特征；通过池化层根据每一个所述文本子句的信息对所述文本向量的领域特征进行池化处理；通过输出层根据所述池化处理的结果，获取每一个所述文本子句的领域信息。

首先，针对文本向量的句法特征，通过领域分类部分的批规范化层和卷积层，可以把该句法特征映射为领域特征 **C**。具体到“你为什么这么晚回家啊？请打开卧室的灯”这一示例，得到的领域特征 **C** 可以为一个 $N \times D$ 的二维矩阵，其中，**N** 为文本数据中包含的字词数，本示例中为 17，**D** 为每一个字词的领域特征向量的维度。

其次，根据句边界探测部分得到的每一个文本子句的范围，可以把领域特征 **C** 转换为 $S=(m_1, m_2, m_3, \dots)$ ，其中，**m** 为每一个文本子句对应的二维领域特征矩阵，**S** 是文本子

句的二维领域特征矩阵的集合，也为一个 $N \times D$ 的二维矩阵。其中，每个 m 为一个 $W \times D$ 的二维矩阵， W 表示当前文本子句中的字词的数量， D 如前所述为每一个字词的二维特征向量的维度。

具体到“你为什么这么晚回家啊？请打开卧室的灯”这一示例，其包括文本子句“你为什么这么晚回家啊”，则其对应的二维领域特征矩阵 m_1 为一个 $10 \times D$ 的矩阵，文本子句“请打开卧室的灯”，其对应的二维领域特征矩阵 m_2 为一个 $7 \times D$ 的矩阵。相对应地， $S=(m_1, m_2)$ 。

接着，针对 S ，在第一维上即 N 维上选取最大值，即在每个文本子句对应的二维特征领域矩阵的第一维进行 $\max$ 操作，得到每一个文本子句对应的二维特征领域矩阵的一维特征 ($1 \times D$)，进而得到所有文本子句的固定长度的特征表达 $T=(u_1, u_2, u_3, \dots)$ ，其中， u 为每个文本子句对应的一维且长度为 D 的领域特征向量。

然后，通过领域分类部分的池化层对 $T=(u_1, u_2, u_3, \dots)$ 进行池化处理后，再通过 Softmax 函数得到每一个文本子句的领域概率分布，根据每一个文本子句的领域概率分布确定该文本子句的领域信息。

(6) 根据每一个文本子句的领域信息，识别语音输入数据中的语音指令。

可见，通过上述过程 (2) - (4)，实现了 CNN 模型对文本子句的划分和文本子句的领域信息确定。通过该 CNN 模型，将语音指令的提取和识别两个任务统一于一个 CNN 模型框架中，有效实现了对用户指令的切分和识别。

进一步地，基于 CNN 输出的结果，即可确定出语音指令。

例如，为语音指令设定领域信息 IOT (Internet Of Things)，则如果某个文本子句的领域信息被分类为 IOT 领域，则可以认为该文本子句对应的那部分语音输入数据即为语音指令。具体到“你为什么这么晚回家啊？请打开卧室的灯”这一示例，“请打开卧室的灯”会被分类到 IOT 领域，因此，确定“请打开卧室的灯”为用于操作和控制智能语音设备的语音指令。

通过本实施例，先获取由语音输入数据转换的文本数据和该文本数据对应的文本向量；再通过对文本向量的特征提取获得对应的句法特征；然后，根据句法特征对语音输入数据对应的文本数据进行文本子句的划分及文本子句的领域信息的确定；进而，根据文本子句的领域信息识别语音输入数据中的语音指令。可见，通过本实施例，使得智能语音设备更加适用于实际的使用环境，无需用户再使用唤醒词唤醒智能语音设备，不管是用户使用纯语音指令的语音输入数据，还是使用包含语音指令和其它语音数据的混合

语音输入数据，都能对语音输入数据进行有效的子句划分，并识别其中包含的语音指令，进而，后续可以通过识别的语音指令对智能语音设备进行操作和控制。

因无需使用唤醒词唤醒智能语音设备，减轻了用户的操作负担，也提高了智能语音设备对用户语音指令的智能化处理程度；并且，智能语音设备也无需再针对唤醒词进行处理，减轻了智能语音设备的数据处理负担。

本实施例的语音识别方法可以由任意适当的具有数据处理能力的智能语音设备执行，如，具有相应功能的各种智能家电等。

实施例三

参照图 4，示出了根据本发明实施例三的一种语音识别装置的结构框图。

本实施例的语音识别装置包括：第一获取模块 302，用于获取与语音输入数据对应的文本数据和所述文本数据对应的文本向量；第二获取模块 304，用于获取所述文本向量的句法特征；第三获取模块 306，用于根据所述句法特征，获取所述文本数据中包含的至少一个文本子句，以及，获取每一个所述文本子句的领域信息；识别模块 308，用于至少根据每一个所述文本子句的领域信息，识别所述语音输入数据中的语音指令。

通过本实施例，先获取由语音输入数据转换的文本数据和该文本数据对应的文本向量；再通过对文本向量的特征提取获得对应的句法特征；然后，根据句法特征对语音输入数据对应的文本数据进行文本子句的划分及文本子句的领域信息的确定；进而，根据文本子句的领域信息识别语音输入数据中的语音指令。可见，通过本实施例，使得智能语音设备更加适用于实际的使用环境，无需用户再使用唤醒词唤醒智能语音设备，不管是用户使用纯语音指令的语音输入数据，还是使用包含语音指令和其它语音数据的混合语音输入数据，都能对语音输入数据进行有效的子句划分，并识别其中包含的语音指令，进而，后续可以通过识别的语音指令对智能语音设备进行操作和控制。

因无需使用唤醒词唤醒智能语音设备，减轻了用户的操作负担，也提高了智能语音设备对用户语音指令的智能化处理程度；并且，智能语音设备也无需再针对唤醒词进行处理，减轻了智能语音设备的数据处理负担。

实施例四

参照图 5，示出了根据本发明实施例四的一种语音识别装置的结构框图。

本实施例的语音识别装置包括：第一获取模块 402，用于获取与语音输入数据对应

的文本数据和所述文本数据对应的文本向量；第二获取模块 404，用于获取所述文本向量的句法特征；第三获取模块 406，用于根据所述句法特征，获取所述文本数据中包含的至少一个文本子句，以及，获取每一个所述文本子句的领域信息；识别模块 408，用于至少根据每一个所述文本子句的领域信息，识别所述语音输入数据中的语音指令。

- 5 可选地，所述第一获取模块 402，用于获取语音输入数据，并生成与所述语音输入数据对应的文本数据；生成所述文本数据中的每一个字词对应的字词向量；根据每一个字词对应的字词向量，生成所述文本数据对应的文本向量。

可选地，所述第二获取模块 404，用于对所述文本向量进行特征提取，获取所述文本向量的句法特征。

- 10 可选地，所述第二获取模块 404，用于对所述文本向量中的每一个字词对应的字词向量进行特征提取，获取每一个字词的句法特征。

可选地，所述第三获取模块 406 包括：子句获取模块 4062，用于根据每一个字词的句法特征，获取每一个字词的标签，其中，所述标签包括结束标签；根据每一个字词的标签，获得所述文本数据的序列标注；根据所述序列标注中的结束标签，获取所述文本数据中包含的至少一个文本子句；领域获取模块 4064，用于根据所述句法特征，获取每一个所述文本子句的领域信息。

- 15 可选地，所述领域获取模块 4064 包括：领域特征模块 40642，用于根据所述文本向量的句法特征，获取每一个所述文本子句对应的领域特征；确定模块 40644，用于对每一个所述文本子句的领域特征，在每个特征维度上进行最大特征值提取，生成每一个所述文本子句的领域特征向量；根据每一个所述文本子句的领域特征向量，确定当前文本子句的领域信息。

可选地，所述领域特征模块 40642，用于根据所述文本向量的句法特征，获取所述文本向量的领域特征；根据每一个所述文本子句所包含的字词的信息，从所述文本向量的领域特征中获取每一个所述文本子句对应的领域特征。

- 25 可选地，所述第二获取模块 404，用于通过卷积神经网络模型的特征提取部分对所述文本向量进行特征提取，获取所述文本向量的句法特征；所述第三获取模块 406，用于通过所述卷积神经网络模型的句边界探测部分根据所述句法特征，获取所述文本数据中包含的至少一个文本子句；通过所述卷积神经网络模型的领域分类部分根据所述句法特征和每一个所述文本子句的信息，获取每一个所述文本子句的领域信息；其中，所述句边界探测部分和所述领域分类部分共享所述特征提取部分提取的句法特征。
- 30

可选地，所述第二获取模块 404，用于对输入的向量进行批规范化操作，生成规范化的向量；对所述规范化的向量进行非线性化处理；通过所述卷积层对非线性处理后的所述向量进行特征提取，获得初始特征；对所述初始特征进行残差分析处理，根据所述残差分析处理结果获得所述向量的句法特征并输出；返回所述对输入的向量进行批规范化操作继续执行，直至获得所述文本向量的句法特征。

可选地，所述特征提取部分至少包括 12 个卷积层，通过线性门函数对所述规范化的向量进行非线性化处理。

可选地，所述第三获取模块 406 在通过所述卷积神经网络模型的句边界探测部分根据所述句法特征，获取所述文本数据中包含的至少一个文本子句时：对所述句法特征进行批规范化操作，生成规范化的句法特征；通过卷积层对所述规范化的句法特征进行特征提取；通过输出层根据特征提取结果确定所述文本数据中的每个字词的标签，根据每个字词的标签获取所述文本数据中包含的至少一个文本子句。

可选地，所述第三获取模块 406 在通过所述卷积神经网络模型的领域分类部分根据所述句法特征和每一个所述文本子句的信息，获取每一个所述文本子句的领域信息时：对所述句法特征进行批规范化操作，生成规范化的句法特征；通过卷积层对所述规范化的句法特征进行特征映射，获取所述文本向量的领域特征；通过池化层根据每一个所述文本子句的信息对所述文本向量的领域特征进行池化处理；通过输出层根据所述池化处理的结果，获取每一个所述文本子句的领域信息。

本实施例的语音识别装置用于实现前述多个方法实施例中相应的语音识别方法，并具有相应的方法实施例的有益效果，在此不再赘述。此外，本实施例的语音识别装置中的各个模块的功能实现均可参照前述方法实施例中的相应部分的描述，在此亦不再赘述。

实施例五

参照图 6，示出了根据本发明实施例六的一种智能设备的结构示意图，本发明具体实施例并不对智能设备的具体实现做限定。

如图 6 所示，该智能设备可以包括：处理器(processor)502、通信接口(Communications Interface)504、存储器(memory)506、以及通信总线 508。

其中：

处理器 502、通信接口 504、以及存储器 506 通过通信总线 508 完成相互间的通信。

通信接口 504，用于与其它电子设备如其它智能设备或服务器进行通信。

处理器 502, 用于执行程序 510, 具体可以执行上述语音识别方法实施例中的相关步骤。

具体地, 程序 510 可以包括程序代码, 该程序代码包括计算机操作指令。

处理器 502 可能是中央处理器 CPU, 或者是特定集成电路 ASIC (Application Specific Integrated Circuit), 或者是被配置成实施本发明实施例的一个或多个集成电路。智能设备包括的一个或多个处理器, 可以是同一类型的处理器, 如一个或多个 CPU; 也可以是不同类型的处理器, 如一个或多个 CPU 以及一个或多个 ASIC。

存储器 506, 用于存放程序 510。存储器 506 可能包含高速 RAM 存储器, 也可能还包括非易失性存储器 (non-volatile memory), 例如至少一个磁盘存储器。

10 程序 510 具体可以用于使得处理器 502 执行以下操作: 获取与语音输入数据对应的文本数据和所述文本数据对应的文本向量; 获取所述文本向量的句法特征; 根据所述句法特征, 获取所述文本数据中包含的至少一个文本子句, 以及, 获取每一个所述文本子句的领域信息; 至少根据每一个所述文本子句的领域信息, 识别所述语音输入数据中的语音指令。

15 在一种可选的实施方式中, 程序 510 还用于使得处理器 502 在获取与语音输入数据对应的文本数据和所述文本数据对应的文本向量时, 获取语音输入数据, 并生成与所述语音输入数据对应的文本数据; 生成所述文本数据中的每一个字词对应的字词向量; 根据每一个字词对应的字词向量, 生成所述文本数据对应的文本向量。

在一种可选的实施方式中, 程序 510 还用于使得处理器 502 在获取所述文本向量的
20 句法特征时, 对所述文本向量进行特征提取, 获取所述文本向量的句法特征。

在一种可选的实施方式中, 程序 510 还用于使得处理器 502 在对所述文本向量进行特征提取, 获取所述文本向量的句法特征时, 对所述文本向量中的每一个字词对应的字词向量进行特征提取, 获取每一个字词的句法特征。

在一种可选的实施方式中, 程序 510 还用于使得处理器 502 在根据所述句法特征,
25 获取所述文本数据中包含的至少一个文本子句时, 根据每一个字词的句法特征, 获取每一个字词的标签, 其中, 所述标签包括结束标签; 根据每一个字词的标签, 获得所述文本数据的序列标注; 根据所述序列标注中的结束标签, 获取所述文本数据中包含的至少一个文本子句。

在一种可选的实施方式中, 程序 510 还用于使得处理器 502 在获取每一个所述文本
30 子句的领域信息时, 根据所述文本向量的句法特征, 获取每一个所述文本子句对应的领

域特征；对每一个所述文本子句的领域特征，在每个特征维度上进行最大特征值提取，生成每一个所述文本子句的领域特征向量；根据每一个所述文本子句的领域特征向量，确定当前文本子句的领域信息。

在一种可选的实施方式中，程序 510 还用于使得处理器 502 在根据所述文本向量的
5 句法特征，获取每一个所述文本子句对应的领域特征时，根据所述文本向量的句法特征，获取所述文本向量的领域特征；根据每一个所述文本子句所包含的字词的信息，从所述文本向量的领域特征中获取每一个所述文本子句对应的领域特征。

在一种可选的实施方式中，程序 510 还用于使得处理器 502 通过卷积神经网络模型的特征提取部分对所述文本向量进行特征提取，获取所述文本向量的句法特征；程序 510
10 还用于使得处理器 502 通过所述卷积神经网络模型的句边界探测部分根据所述句法特征，获取所述文本数据中包含的至少一个文本子句；通过所述卷积神经网络模型的领域分类部分根据所述句法特征和每一个所述文本子句的信息，获取每一个所述文本子句的领域信息；其中，所述句边界探测部分和所述领域分类部分共享所述特征提取部分提取的句法特征。

在一种可选的实施方式中，程序 510 还用于使得处理器 502 在通过卷积神经网络模型的特征提取部分对所述文本向量进行特征提取，获取所述文本向量的句法特征时，对
15 输入的向量进行批规范化操作，生成规范化的向量；对所述规范化的向量进行非线性化处理；通过卷积层对非线性处理后的所述向量进行特征提取，获得初始特征；对所述初始特征进行残差分析处理，根据所述残差分析处理结果获得所述向量的句法特征并输出；
20 返回所述对输入的向量进行批规范化操作的步骤继续执行，直至获得所述文本向量的句法特征。

在一种可选的实施方式中，所述特征提取部分至少包括 12 个卷积层；通过线性门函数对所述规范化的向量进行非线性化处理。

在一种可选的实施方式中，程序 510 还用于使得处理器 502 在通过所述卷积神经网络模型的句边界探测部分根据所述句法特征，获取所述文本数据中包含的至少一个文本
25 子句时，对所述句法特征进行批规范化操作，生成规范化的句法特征；通过卷积层对所述规范化的句法特征进行特征提取；通过输出层根据特征提取结果确定所述文本数据中的每个字词的标签，根据每个字词的标签获取所述文本数据中包含的至少一个文本子句。

在一种可选的实施方式中，程序 510 还用于使得处理器 502 在通过所述卷积神经网络模型的领域分类部分根据所述句法特征和每一个所述文本子句的信息，获取每一个所
30

述文本子句的领域信息时，对所述句法特征进行批规范化操作，生成规范化的句法特征；通过卷积层对所述规范化的句法特征进行特征映射，获取所述文本向量的领域特征；通过池化层根据每一个所述文本子句的信息对所述文本向量的领域特征进行池化处理；通过输出层根据所述池化处理的结果，获取每一个所述文本子句的领域信息。

5 在一种可选的实施方式中，本实施例的智能设备还可以包括麦克风，以接收用户输入的模拟语音信号并转换为数字语音信号即语音输入数据；程序 510 还可以用于使得处理器 502 将所述语音输入数据转换为对应的文本数据。但不限于此，麦克风也可以独立于所述智能设备设置，并通过适当的连接方式与智能设备连接，并将所述语音输入数据发送给处理器。

10 程序 510 中各步骤的具体实现可以参见上述语音识别方法实施例中的相应步骤和单元中对应的描述，在此不赘述。所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的设备和模块的具体工作过程，可以参考前述方法实施例中的对应过程描述，在此不再赘述。

15 通过本实施例的智能设备，先获取由语音输入数据转换的文本数据和该文本数据对应的文本向量；再通过对文本向量的特征提取获得对应的句法特征；然后，根据句法特征对语音输入数据对应的文本数据进行文本子句的划分及文本子句的领域信息的确定；进而，根据文本子句的领域信息识别语音输入数据中的语音指令。可见，通过本实施例，使得智能语音设备更加适用于实际的使用环境，无需用户再使用唤醒词唤醒智能语音设备，不管是用户使用纯语音指令的语音输入数据，还是使用包含语音指令和其它语音数
20 据的混合语音输入数据，都能对语音输入数据进行有效的子句划分，并识别其中包含的语音指令，进而，后续可以通过识别的语音指令对智能语音设备进行操作和控制。

因无需使用唤醒词唤醒智能语音设备，减轻了用户的操作负担，也提高了智能语音设备对用户语音指令的智能化处理程度；并且，智能语音设备也无需再针对唤醒词进行处理，减轻了智能语音设备的数据处理负担。

25

需要指出，根据实施的需要，可将本发明实施例中描述的各个部件/步骤拆分为更多部件/步骤，也可将两个或多个部件/步骤或者部件/步骤的部分操作组合成新的部件/步骤，以实现本发明实施例的目的。

上述根据本发明实施例的方法可在硬件、固件中实现，或者被实现为可存储在记录
30 介质（诸如 CD ROM、RAM、软盘、硬盘或磁光盘）中的软件或计算机代码，或者被实

现通过网络下载的原始存储在远程记录介质或非暂时机器可读介质中并将被存储在本地记录介质中的计算机代码，从而在此描述的方法可被存储在使用通用计算机、专用处理器或者可编程或专用硬件（诸如 ASIC 或 FPGA）的记录介质上的这样的软件处理。可以理解，计算机、处理器、微处理器控制器或可编程硬件包括可存储或接收软件或计算机代码的存储组件（例如，RAM、ROM、闪存等），当所述软件或计算机代码被计算机、处理器或硬件访问且执行时，实现在此描述的语音识别方法。此外，当通用计算机访问用于实现在此示出的语音识别方法的代码时，代码的执行将通用计算机转换为用于执行在此示出的语音识别方法的专用计算机。

本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及方法步骤，能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本发明实施例的范围。

以上实施方式仅用于说明本发明实施例，而并非对本发明实施例的限制，有关技术领域的普通技术人员，在不脱离本发明实施例的精神和范围的情况下，还可以做出各种变化和变型，因此所有等同的技术方案也属于本发明实施例的范畴，本发明实施例的专利保护范围应由权利要求限定。

权利要求书

1.一种语音识别方法，包括：

获取与语音输入数据对应的文本数据和所述文本数据对应的文本向量；

获取所述文本向量的句法特征；

5 根据所述句法特征，获取所述文本数据中包含的至少一个文本子句，以及，获取每一个所述文本子句的领域信息；

至少根据每一个所述文本子句的领域信息，识别所述语音输入数据中的语音指令。

2.根据权利要求1所述的方法，其中，所述获取与语音输入数据对应的文本数据和所述文本数据对应的文本向量，包括：

10 获取语音输入数据，并生成与所述语音输入数据对应的文本数据；

生成所述文本数据中的每一个字词对应的字词向量；

根据每一个字词对应的字词向量，生成所述文本数据对应的文本向量。

3.根据权利要求2所述的方法，其中，所述获取所述文本向量的句法特征，包括：

对所述文本向量进行特征提取，获取所述文本向量的句法特征。

15 4.根据权利要求3所述的方法，其中，所述对所述文本向量进行特征提取，获取所述文本向量的句法特征，包括：

对所述文本向量中的每一个字词对应的字词向量进行特征提取，获取每一个字词的句法特征。

5.根据权利要求4所述的方法，其中，所述根据所述句法特征，获取所述文本数据中包含的至少一个文本子句，包括：

20 根据每一个字词的句法特征，获取每一个字词的标签，其中，所述标签包括结束标签；

根据每一个字词的标签，获得所述文本数据的序列标注；

根据所述序列标注中的结束标签，获取所述文本数据中包含的至少一个文本子句。

25 6.根据权利要求5所述的方法，其中，所述获取每一个所述文本子句的领域信息，包括：

根据所述文本向量的句法特征，获取每一个所述文本子句对应的领域特征；

对每一个所述文本子句的领域特征，在每个特征维度上进行最大特征值提取，生成每一个所述文本子句的领域特征向量；

30 根据每一个所述文本子句的领域特征向量，确定当前文本子句的领域信息。

7.根据权利要求6所述的方法,其中,所述根据所述文本向量的句法特征,获取每一个所述文本子句对应的领域特征,包括:

根据所述文本向量的句法特征,获取所述文本向量的领域特征;

根据每一个所述文本子句所包含的字词的信息,从所述文本向量的领域特征中获取

5 每一个所述文本子句对应的领域特征。

8.根据权利要求1-7任一项所述的方法,其中:

通过卷积神经网络模型的特征提取部分对所述文本向量进行特征提取,获取所述文本向量的句法特征;

10 通过所述卷积神经网络模型的句边界探测部分根据所述句法特征,获取所述文本数据中包含的至少一个文本子句;通过所述卷积神经网络模型的领域分类部分根据所述句法特征和每一个所述文本子句的信息,获取每一个所述文本子句的领域信息;

其中,所述句边界探测部分和所述领域分类部分共享所述特征提取部分提取的句法特征。

15 9.根据权利要求8所述的方法,其中,所述通过卷积神经网络模型的特征提取部分对所述文本向量进行特征提取,获取所述文本向量的句法特征,包括:

对输入的向量进行批规范化操作,生成规范化的向量;

对所述规范化的向量进行非线性化处理;

通过卷积层对非线性处理后的所述向量进行特征提取,获得初始特征;

20 对所述初始特征进行残差分析处理,根据所述残差分析处理结果获得所述向量的句法特征并输出;

返回所述对输入的向量进行批规范化操作的步骤继续执行,直至获得所述文本向量的句法特征。

10.根据权利要求9所述的方法,其中,所述特征提取部分至少包括12个卷积层;通过线性门函数对所述规范化的向量进行非线性化处理。

25 11.根据权利要求8所述的方法,其中,所述通过所述卷积神经网络模型的句边界探测部分根据所述句法特征,获取所述文本数据中包含的至少一个文本子句,包括:

对所述句法特征进行批规范化操作,生成规范化的句法特征;

通过卷积层对所述规范化的句法特征进行特征提取;

30 通过输出层根据特征提取结果确定所述文本数据中的每个字词的标签,根据每个字词的标签获取所述文本数据中包含的至少一个文本子句。

12.根据权利要求 8 所述的方法，其中，所述通过所述卷积神经网络模型的领域分类部分根据所述句法特征和每一个所述文本子句的信息，获取每一个所述文本子句的领域信息包括：

对所述句法特征进行批规范化操作，生成规范化的句法特征；

5 通过卷积层对所述规范化的句法特征进行特征映射，获取所述文本向量的领域特征；

通过池化层根据每一个所述文本子句的信息对所述文本向量的领域特征进行池化处理；

通过输出层根据所述池化处理的结果，获取每一个所述文本子句的领域信息。

10

13.一种语音识别装置，包括：

第一获取模块，用于获取与语音输入数据对应的文本数据和所述文本数据对应的文本向量；

15 第二获取模块，用于获取所述文本向量的句法特征；

第三获取模块，用于根据所述句法特征，获取所述文本数据中包含的至少一个文本子句，以及，获取每一个所述文本子句的领域信息；

识别模块，用于至少根据每一个所述文本子句的领域信息，识别所述语音输入数据中的语音指令。

20

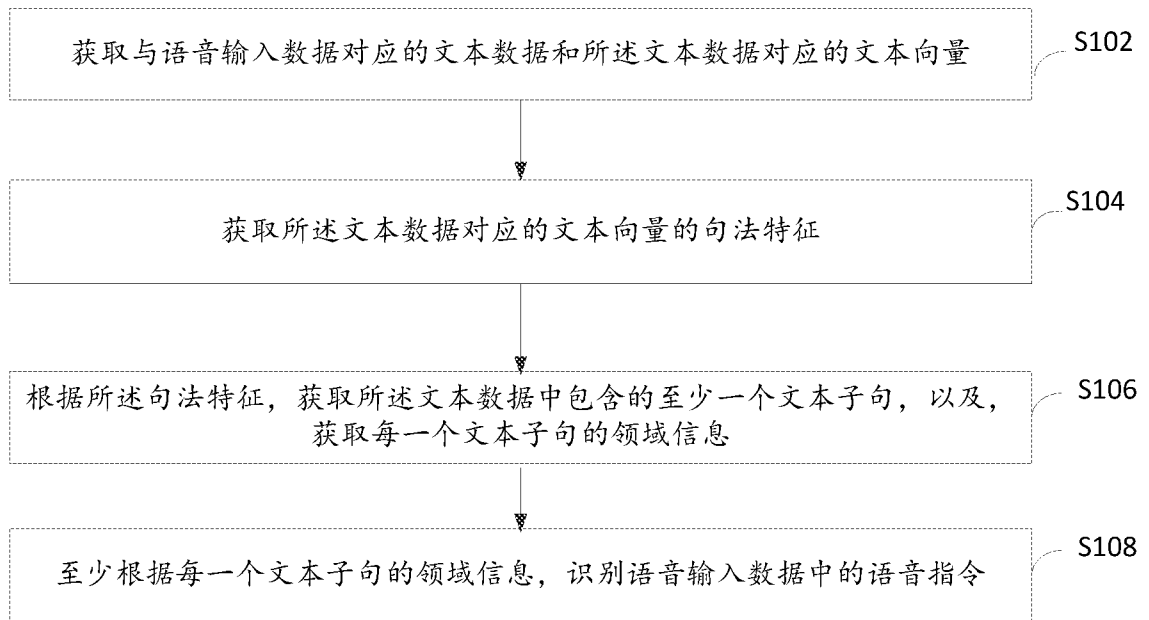


图 1

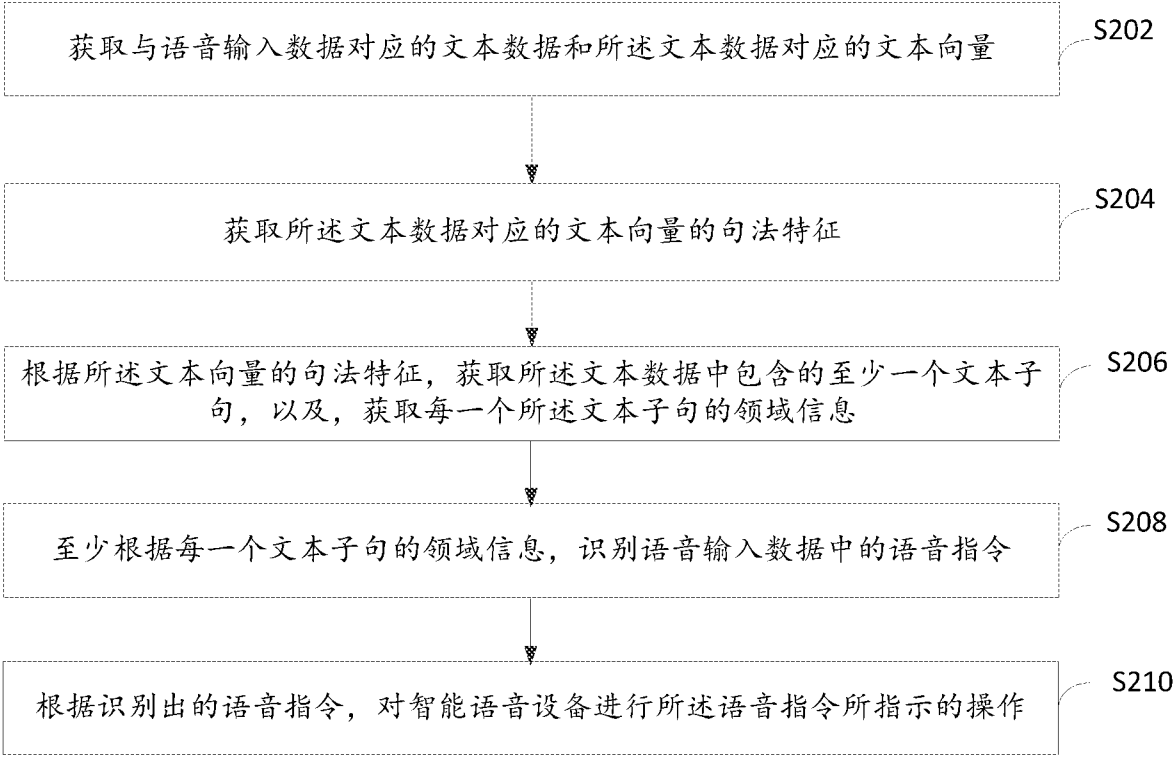


图 2

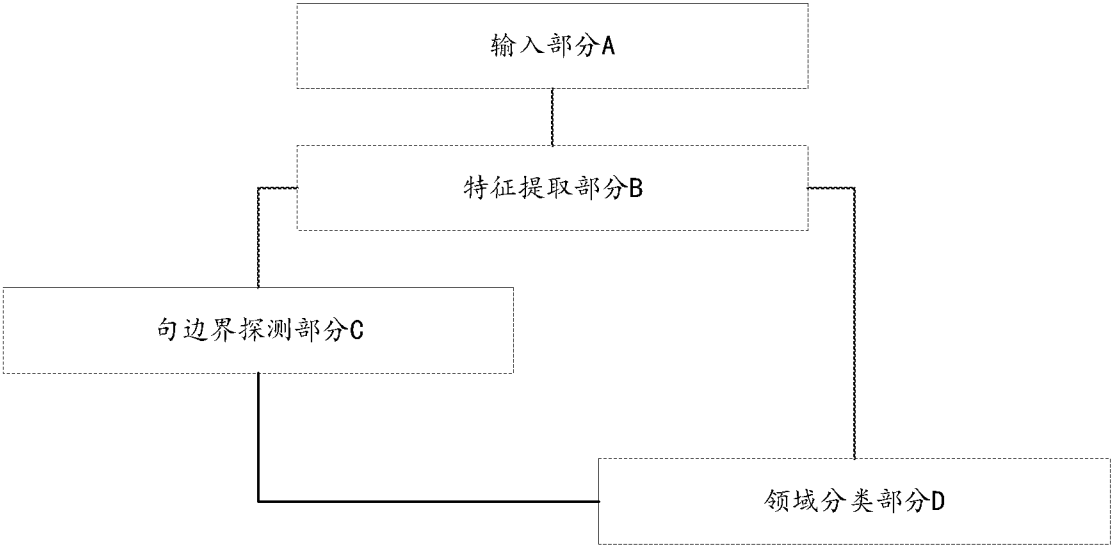


图 3

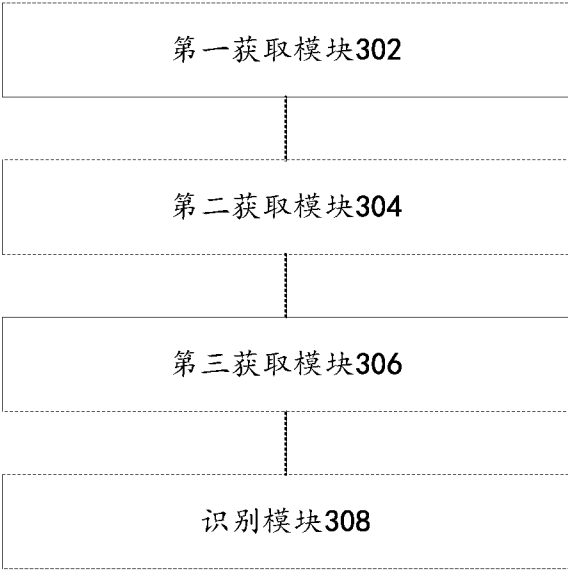


图 4

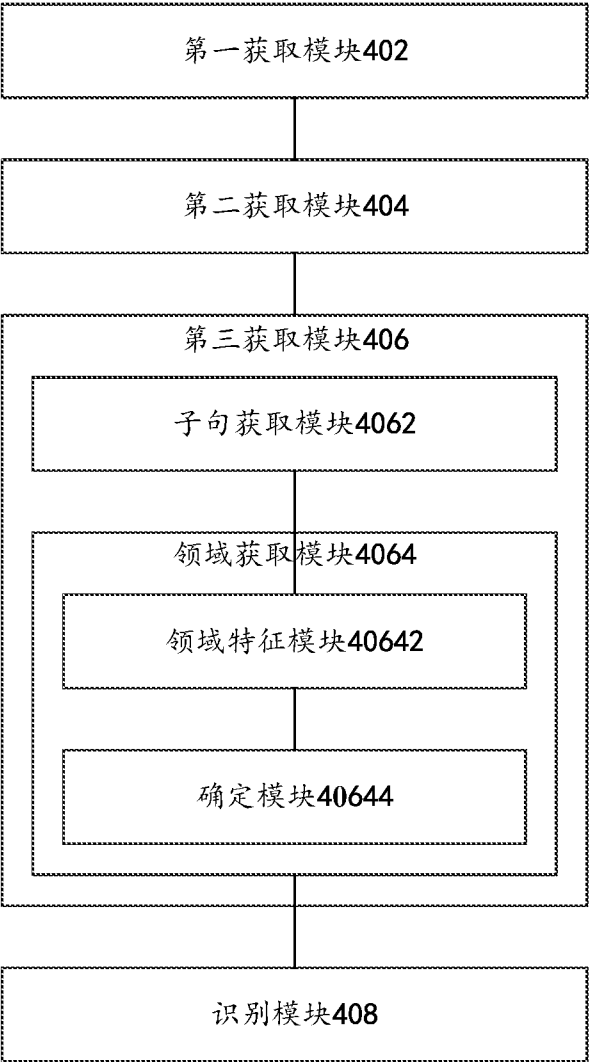


图 5

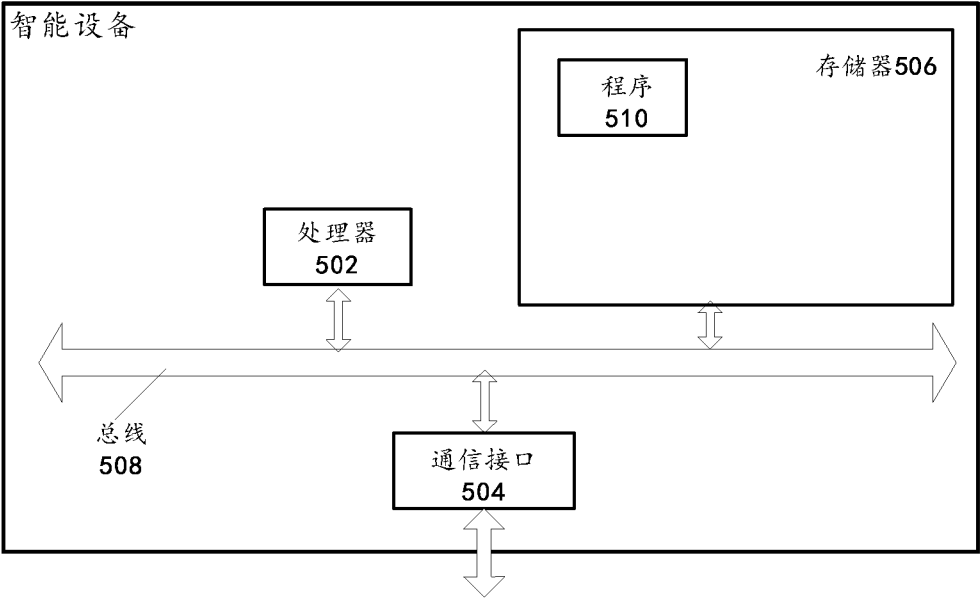


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/070581

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/22(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)

G10L 15

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)

CNTXT; VEN; CNKI; 百度学术: 语音识别, 文本, 语义, 多个, 子句, 短句, 领域, 意图, 分类, 指令, speech recognition, text, semantic, multi+, clause, string, domain, intent, classify, command, instruction

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2018350353 A1 (APPLE INC.) 06 December 2018 (2018-12-06) description, paragraphs [0218]-[0264], figures 8-9	1-7, 13
X	CN 106471570 A (APPLE INC.) 01 March 2017 (2017-03-01) description, paragraphs [0247]-[0265], figures 8-9	1-7, 13
A	CN 107679042 A (BEIJING LINGBAN AIAAS INTELLIGENT TECHNOLOGY CO., LTD.) 09 February 2018 (2018-02-09) entire document	1-13
A	CN 106528522 A (NANJING WECARE SOFTWARE CO., LTD.) 22 March 2017 (2017-03-22) entire document	1-13
A	CN 108563790 A (IFLYTEK CO., LTD.) 21 September 2018 (2018-09-21) entire document	1-13
A	CN 107315737 A (BEIJING QIYI CENTURY SCIENCE AND TECHNOLOGY CO., LTD.) 03 November 2017 (2017-11-03) entire document	1-13



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

06 April 2020

Date of mailing of the international search report

13 April 2020

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/CN2020/070581

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)		Publication date (day/month/year)
US	2018350353	A1	06 December 2018	US	10497365 B2	03 December 2019
CN	106471570	A	01 March 2017	EP	3480811 A1	08 May 2019
				US	9966065 B2	08 May 2018
				AU	2015266863 A1	29 September 2016
				US	2015348551 A1	03 December 2015
				US	2020043482 A1	06 February 2020
				WO	2015184186 A1	03 December 2015
				TW	201612773 A	01 April 2016
				TW	1566107 B	11 January 2017
				EP	3149728 B1	16 January 2019
				CN	106471570 B	01 October 2019
				US	10497365 B2	03 December 2019
				AU	2015266863 B2	15 March 2018
				US	2018350353 A1	06 December 2018
				CN	110797019 A	14 February 2020
				EP	3149728 A1	05 April 2017
CN	107679042	A	09 February 2018	None		
CN	106528522	A	22 March 2017	None		
CN	108563790	A	21 September 2018	None		
CN	107315737	A	03 November 2017	None		

国际检索报告

国际申请号

PCT/CN2020/070581

A. 主题的分类 G10L 15/22 (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) G10L 15 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)) CNTXT; VEN; CNKI; 百度学术: 语音识别, 文本, 语义, 多个, 子句, 短句, 领域, 意图, 分类, 指令, speech recognition, text, semantic, multi+, clause, string, domain, intent, classify, command, instruction		
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	US 2018350353 A1 (APPLE INC.) 2018年 12月 6日 (2018 - 12 - 06) 说明书第[0218]-[0264]段, 附图8-9	1-7, 13
X	CN 106471570 A (苹果公司) 2017年 3月 1日 (2017 - 03 - 01) 说明书第[0247]-[0265]段, 附图8-9	1-7, 13
A	CN 107679042 A (北京灵伴即时智能科技有限公司) 2018年 2月 9日 (2018 - 02 - 09) 全文	1-13
A	CN 106528522 A (南京威卡尔软件有限公司) 2017年 3月 22日 (2017 - 03 - 22) 全文	1-13
A	CN 108563790 A (科大讯飞股份有限公司) 2018年 9月 21日 (2018 - 09 - 21) 全文	1-13
A	CN 107315737 A (北京奇艺世纪科技有限公司) 2017年 11月 3日 (2017 - 11 - 03) 全文	1-13
☐ 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 2020年 4月 6日		国际检索报告邮寄日期 2020年 4月 13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 游晓梅 电话号码 (86-10)62089539

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/070581

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)
US	2018350353	A1	2018年 12月 6日	US 10497365 B2	2019年 12月 3日
CN	106471570	A	2017年 3月 1日	EP 3480811 A1	2019年 5月 8日
				US 9966065 B2	2018年 5月 8日
				AU 2015266863 A1	2016年 9月 29日
				US 2015348551 A1	2015年 12月 3日
				US 2020043482 A1	2020年 2月 6日
				WO 2015184186 A1	2015年 12月 3日
				TW 201612773 A	2016年 4月 1日
				TW 1566107 B	2017年 1月 11日
				EP 3149728 B1	2019年 1月 16日
				CN 106471570 B	2019年 10月 1日
				US 10497365 B2	2019年 12月 3日
				AU 2015266863 B2	2018年 3月 15日
				US 2018350353 A1	2018年 12月 6日
				CN 110797019 A	2020年 2月 14日
				EP 3149728 A1	2017年 4月 5日
CN	107679042	A	2018年 2月 9日	无	
CN	106528522	A	2017年 3月 22日	无	
CN	108563790	A	2018年 9月 21日	无	
CN	107315737	A	2017年 11月 3日	无	