(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 109918680 B
(45)授权公告日 2023.04.07

(21)申请号 201910243952.9
(22)申请日 2019.03.28
(65)同一申请的已公布的文献号
申请公布号 CN 109918680 A
(43)申请公布日 2019.06.21
(73)专利权人 腾讯科技(上海)有限公司
地址 201200 上海市虹梅路1801号C区5层
(72)发明人 杨奇 杨君 吴丹
(74)专利代理机构 深圳市深佳知识产权代理事务所(普通合伙) 44285
专利代理师 王仲凯
(51)Int.Cl.
G06F 40/295(2020.01)
G06F 40/247(2020.01)
G06F 40/242(2020.01)
G06F 16/33(2019.01)
G06N 3/0455(2023.01)

审查员 朱思韦

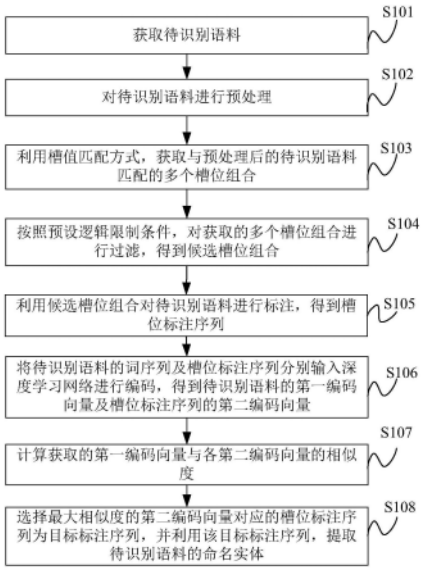
权利要求书2页 说明书11页 附图4页

(54)发明名称

实体识别方法、装置及计算机设备

(57)摘要

本申请提供了一种实体识别方法、装置及计算机设备，在对待识别语料进行召回处理过程中，采用字典匹配方式，召回待识别语料的可能候选槽位组合，并据此待识别语料进行标注，得到相应的槽位标注序列，利用深度学习网络，分别获得待识别语料的编码向量和各槽位标注序列的编码向量，选择与待识别语料的编码向量最相似的槽位标注序列的编码向量，并将其对应的槽位标注序列作为最佳槽位标注序列，据此得到待识别语料的命名实体。可见，本申请利用规则模板和深度学习算法的优点，实现了对各种类型语料的命名实体的快速、简单且准确识别，解决了冷启动问题，且得到的识别结果通常是计算机语言，电子设备能够直接响应识别结果。



CN 109918680 B

1.一种实体识别方法,其特征在于,包括:

获取待识别语料;

获取槽值字典及同义词字典;

将所述待识别语料与所述槽值字典及所述同义词字典进行逐一匹配,得到多个槽位组合;

按照预设逻辑限制条件,对所述多个槽位组合进行过滤,得到候选槽位组合;

利用所述候选槽位组合对所述待识别语料进行标注,得到相应的槽位标注序列;

基于深度学习网络,分别对所述待识别语料的词序列及所述槽位标注序列进行编码,得到所述待识别语料的第一编码向量,以及所述槽位标注序列的第二编码向量;

基于所述第一编码向量与所述第二编码向量的相似度,确定所述槽位标注序列中的目标标注序列;

利用所述目标标注序列,得到所述待识别语料的命名实体。

2.根据权利要求1所述的方法,其特征在于,所述基于所述第一编码向量与所述第二编码向量的相似度,确定所述槽位标注序列中的目标标注序列,包括:

计算所述第一编码向量与所述第二编码向量的相似度;

选择最大相似度的第二编码向量对应的槽位标注序列为目标标注序列。

3.根据权利要求1～2任一项所述的方法,其特征在于,所述基于深度学习网络,分别对所述待识别语料的词序列及所述槽位标注序列进行编码,得到所述待识别语料的第一编码向量,以及所述槽位标注序列的第二编码向量,包括:

获取所述待识别语料的词序列;

将所述词序列输入第一神经网络模型进行编码,得到第一编码向量;

将所述槽位标注序列输入第二神经网络模型进行编码,得到第二编码向量。

4.根据权利要求3所述的方法,其特征在于,所述第一神经网络模型和所述第二神经网络模型是不同类型的双向长短期记忆网络。

5.一种实体识别装置,其特征在于,所述装置包括:

第一获取模块,用于获取待识别语料;

第二获取模块,用于利用槽值匹配方式,获取所述待识别语料的候选槽位组合;

标注模块,用于利用所述候选槽位组合对所述待识别语料进行标注,得到相应的槽位标注序列;

第三获取模块,用于基于深度学习网络,分别对所述待识别语料的词序列及所述槽位标注序列进行编码,得到所述待识别语料的第一编码向量,以及所述槽位标注序列的第二编码向量;

识别模块,用于基于所述第一编码向量与所述第二编码向量的相似度,确定所述槽位标注序列中的目标标注序列,并利用所述目标标注序列,得到所述待识别语料的命名实体;

所述第二获取模块包括:

第一获取单元,用于利用槽值匹配方式,获取所述待识别语料的多个槽位组合;

过滤单元,用于按照预设逻辑限制条件,对所述多个槽位组合进行过滤,得到候选槽位组合;

所述第一获取单元包括:

字典获取单元，用于获取槽值字典及同义词字典；

匹配单元，用于将所述待识别语料与所述槽值字典及所述同义词字典进行逐一匹配，得到多个槽位组合。

6.一种计算机设备，其特征在于，包括：

通信接口；

存储器，用于存储实现如权利要求1～4任一项所述的实体识别方法的程序；

处理器，用于加载并执行所述存储器存储的程序，实现权利要求1～4任一项所述的实体识别方法的各步骤。

7.一种计算机可读存储介质，其特征在于，所述计算机可读存储介质上存储有计算机程序，所述计算机程序在被处理器执行时，实现权利要求1～4任一项所述的实体识别方法的各步骤。

实体识别方法、装置及计算机设备

技术领域

[0001] 本申请涉及人工智能技术领域，具体涉及一种实体识别方法、装置及计算机设备。

背景技术

[0002] 近年来，随着人工智能的发展，人机对话系统被普遍应用到各领域的应用平台。人机对话系统是一种可以与人进行对话的计算机系统，其在获得用户提出的问题后，需要识别问题中的命名实体，以便据此给出用户所需的答案或相应操作，简化人机交互的流程。

[0003] 目前，在命名实体识别(Named Entity Recognition，NER)的应用中，提出利用深度学习来识别语料中的命名实体，即将命名实体作为序列标注，利用大规模语料学习出标注模型，实现对句子的各个位置的标注，由此得到目标语料中的命名实体。但是，这种实体识别方式需要大量高质量的标注语料进行模型训练，可行性较差；只能识别一些简单实体，对于专有名词和充满歧义名词，识别准确度较低，对于更新频率较高的实体，甚至无法识别。

[0004] 对此，技术人员提出了一种利用规则模板识别语料中的命名实体的方法，即将预先构造的模板与目标语料进行匹配，识别目标语料中的命名实体，这种方式虽然能够准确识别出所有复杂的实体，但是，在产品功能域复杂的情况下，需要构造海量模板，工作量极大，且后期维护与可交接性差，也难以复用到不同的场景中。

[0005] 由此可见，如何准确且简便地实现各种场景下的各种命名实体的识别，成为本领域重要研究方向之一。

发明内容

[0006] 有鉴于此，本申请实施例提供一种实体识别方法、装置及计算机设备，利用规则模板和深度学习算法的优点，实现了对各种语料命名实体的快速、简单且准确识别，不仅适用于简单实体识别，也能够适用于复杂实体识别，也解决了冷启动问题，且得到的识别结果通常是计算机语言，电子设备能够直接响应识别结果。

[0007] 为实现上述目的，本申请实施例提供如下技术方案：

[0008] 一种实体识别方法，所述方法包括：

[0009] 获取待识别语料；

[0010] 利用槽值匹配方式，获取所述待识别语料的候选槽位组合；

[0011] 利用所述候选槽位组合对所述待识别语料进行标注，得到相应的槽位标注序列；

[0012] 基于深度学习网络，分别对所述待识别语料的词序列及所述槽位标注序列进行编码，得到所述待识别语料的第一编码向量，以及所述槽位标注序列的第二编码向量；

[0013] 基于所述第一编码向量与所述第二编码向量的相似度，确定所述槽位标注序列中的目标标注序列；

[0014] 利用所述目标标注序列，得到所述待识别语料的命名实体。

[0015] 一种实体识别装置，所述装置包括：

[0016] 第一获取模块，用于获取待识别语料；

[0017] 第二获取模块，用于利用槽值匹配方式，获取所述待识别语料的候选槽位组合；

[0018] 标注模块，用于利用所述候选槽位组合对所述待识别语料进行标注，得到相应的槽位标注序列；

[0019] 第三获取模块，用于基于深度学习网络，分别对所述待识别语料的词序列及所述槽位标注序列进行编码，得到所述待识别语料的第一编码向量，以及所述槽位标注序列的第二编码向量；

[0020] 识别模块，用于基于所述第一编码向量与所述第二编码向量的相似度，确定所述槽位标注序列中的目标标注序列，并利用所述目标标注序列，得到所述待识别语料的命名实体。

[0021] 一种计算机设备，包括：

[0022] 通信接口；

[0023] 存储器，用于存储实现如上所述的实体识别方法的程序；

[0024] 处理器，用于加载并执行所述存储器存储的程序，实现如上所述的实体识别方法的各步骤。

[0025] 一种存储介质，其上存储有程序，所述程序被处理器加载并执行，实现上述实体识别方法的各步骤。

[0026] 由此可见，本申请将待识别语料的槽位抽取问题转化为召回-排序问题，并在对待识别语料进行召回处理过程中，采用字典匹配方式，召回待识别语料的可能候选槽位组合，并据此待识别语料进行标注，得到相应的槽位标注序列，之后，将待识别语料的词序列及槽位标注序列输入深度学习网络，分别获得待识别语料的编码向量，和各槽位标注序列的编码向量，选择与待识别语料的编码向量最相似的槽位标注序列的编码向量，并将其对应的槽位标注序列作为最佳槽位标注序列，据此得到待识别语料的命名实体。可见，本申请利用规则模板和深度学习算法的优点，实现了对各种语料命名实体的快速、简单且准确识别，不仅适用于简单实体识别，也能够适用于复杂、长尾及新出现的实体识别，也解决了冷启动问题，且得到的识别结果通常是计算机语言，电子设备能够直接响应识别结果。

附图说明

[0027] 为了更清楚地说明本申请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申请的实施例，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，还可以根据提供的附图获得其他的附图。

[0028] 图1为本申请实施例提供的一种实体识别系统的结构示意图；

[0029] 图2为本申请实施例提供的一种实体识别方法的流程示意图；

[0030] 图3为本申请实施例提供的另一种实体识别方法的流程示意图；

[0031] 图4为本申请实施例提供的又一种实体识别方法的流出示意图；

[0032] 图5为本申请实施例提供的一种实体识别装置的结构示意图；

[0033] 图6为本申请实施例提供的另一种实体识别装置的结构示意图；

[0034] 图7为本申请实施例提供的又一种实体识别装置的结构示意图；

[0035] 图8为本申请实施例提供的一种计算机设备的硬件结构示意图。

具体实施方式

[0036] 结合背景技术部分的分析，现有技术中利用如BiLSTM-CRF(Bi-directional Long Short-Term Memory-Conditional Random Fields，长短时记忆-条件随机场)算法或其他深度学习算法，实现命名实体识别的方法，具有通用性、代码实现简单，且具有好的维护性和可交接性，能够很好的抽取复杂的语法结构、人类的不同说话习惯中的一些简单实体。而利用规则模板实现命名实体识别的方法，适用于实体复杂，但句式简单的领域，如股票、体育、媒体相关领域。

[0037] 通过上述分析，针对实体复杂、同时句式复杂的领域，如语音助手的音乐领域，本申请需要一种既能识别出所有复杂实体、又能降低“堆模板”算法复杂度的方案，同时还具有可维护性和方法论复用性，也就是将上述两种实体识别方法的优点结合的新方案。

[0038] 基于上述构思，考虑到模板识别文本词汇的优点，以及深度学习处理复杂句式的优点，本申请的发明人引入了推荐系统的Match Ranking(召回-排序)理念，将识别语料中的命名实体问题，即槽位抽取问题转化为Match Ranking问题，也就是先找回后排序的方式。具体的，可以先通过字典匹配和简单规则召回出待识别语料的可能槽位组合，即召回多个槽位组合；再使用每个槽位组合对原始语料进行标注，得到多个序列标注后，建立一个深度学习的模型，对各候选槽位组合(候选序列标注)进行排序，得到最佳的槽位组合，进而得到待识别语料包含的命名实体。

[0039] 下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

[0040] 参照图1，为实现本申请提供的实体识别方法的一种系统结构示意图，该系统可以包括服务器11以及电子设备12，应该理解，图1所呈现的服务器11与电子设备12仅是示例性说明，并不会两者的实现形式做限定。

[0041] 在实际应用中，服务器11与电子设备12之间可以是有线或无线网络连接，具体可以通过GSM、GPRS、LTE等移动网络实现通信连接，或者是通过蓝牙、WIFI、红外线等方式进行通信连接，本申请对服务器11与电子设备12之前的具体通信连接方式不做限定。

[0042] 服务器11可以是为用户提供服务的服务设备，具体可以是独立的应用服务设备，也可以是由多个服务器构成的服务集群，实际应用中，其可以是云服务器、云主机、虚拟中心等，本申请对该服务器的结构及其实现形式不作限定。

[0043] 在人机对话应用场景下，服务器11可以主要负责人机对话过程中的语音识别、语言理解、语言生成、语音合成等功能，并配合电子设备12实现人机对话。可见，本申请提供的实体识别方法可以由服务器11执行，具体实现过程可以参照下文方法实施例的描述。

[0044] 电子设备12可以是面向用户，并能够与用户进行语音交互的终端，如手机、笔记本电脑、iPad、智能音响等，还可以各种语音交互式自助终端，如医院、银行、车站等场所中的自助服务机，此外，电子设备12还可以是支持语音交互的智能机器，如聊天机器人、扫地机器人、点餐服务机器人等。本申请对电子设备的产品类型及其物理形态不做限定，本申请需

要其具有语音交互功能，可以通过安装如语音助手等语音交互类应用程序实现。

[0045] 结合上图1所示的系统结构示意图，参照图2，为本申请实施例提供了一种实体识别方法的流程示意图，可以应用于各种应用场景下的人机对话过程，具体可以由服务器执行，如图2所示，该方法可以包括但并不局限于以下步骤：

[0046] 步骤S101，获取待识别语料；

[0047] 本实施例中，待识别语料可以是用户启动电子设备的某应用程序，且该应用程序具有语音交互功能的情况下输入的数据，其可以是语音数据，也可以是文本数据。需要说明，若待识别语料是文本数据，在后续处理过程中无需进行语音识别，也可以不用进行语音合成。

[0048] 电子设备采集到用户输入的待识别语料后，可以发送至服务器进行处理，在此期间，用户启动的电子设备的应用程序可以等待服务器反馈结果，该反馈结果可以是语音或文本数据，也可以是控制指令，可以依据待识别语料的具体内容确定，本申请对待识别语料的内容及其类型不做限定。

[0049] 步骤S102，对待识别语料进行预处理；

[0050] 本申请对预处理操作内容不做限定，若待识别语料是一个长句，本实施例可以将其截断为多个短句，若待识别语料中包含电子设备的唤醒词，为了避免与唤醒词的功能冲突，本实施例可以剔除待识别语料中的唤醒词等等，本实施例对步骤S102的具体实现方法不做详述。

[0051] 步骤S103，利用槽值匹配方式，获取与预处理后的待识别语料匹配的多个槽位组合；

[0052] 在本领域，语义槽可以是NLU(Natural Language Understanding，自然语言理解)木块从语料中抽取出的特定概念，如命名实体；填槽可以是为了让用户意图转化为用户明确的指令而补全信息的过程；槽填充可以看做是序列标注问题，即对于给定的语料中的每个词分别打上相应的标签。在实际应用中，槽可以由槽位构成，槽值可以是槽可能的取值，如出发地点的槽，可以通过上下文获得、直接获得等槽位获得，具体出发地点内容可以是槽值，一个槽位可以是一种填槽方式，本申请可以将包含多种填槽方式的槽称为槽组，每一个槽位可以对应词槽和接口槽中的一种槽位类型

[0053] 举例说明：以订票场景为例，语义槽可以包括“出发地点”、“出发时间”、“目的地点”等，“出发地点”对应的槽值可以包括“北京”、“上海”、“深圳”等地名，“出发时间”对应的槽值可以包括“上午七点”、“上午十点”、“下午三点”等时间；“目的地点”对应的槽值可以包括“杭州”、“北京”、“海南”等地名。

[0054] 基于上述分析，本实施例可以获取各种槽位对应的槽值字典，由此构成槽位词典并存储，可见，槽位词典可以包括不同槽位对应的各种槽值，本申请对槽值字典及槽位词典的内容，以及获取方式不作限定。

[0055] 之后，可以将待识别语料与槽位词典中的各槽值进行逐一匹配，对槽值匹配到的可能的槽位进行保存，通常情况下，对待识别语料的每一次匹配，槽值匹配到的槽位往往是多个，即一次匹配结果可以是槽位组合，经过与槽位词典的多次匹配，通常会得到多个槽位组合。

[0056] 可选的，由于实际应用中，同一含义的槽值可以有多种说法，如播放这一槽值，可

以由放一首、来一首、播个、点个等多个词汇表示，因此，本申请可以将这类词汇都映射成“播放”。基于此，本申请可以获取用户在人机交互系统中输入的语料，及其对应的能够被计算机设备识别的相应指令内容(即槽值)，得到对应同一指令内容的各种语料，此时，可以认为这些语料属于同义词，并将其映射到该指令内容。按照这种方式，本申请可以得到各种槽值的同义映射，并由此生成同义词词典。可见，该同义词词典可以表征槽值与同义词之间的映射关系，即包括了各种槽值分别对应的同义词。

[0057] 因此，在对待识别语料进行槽值匹配过程中，可以利用槽值字典及同义词词典共同实现，即将待识别语料与槽值字典和同义词词典进行逐一匹配，得到待识别语料对应的槽位组合，具体匹配过程不做限定。

[0058] 综上，本实施例可以利用规则模板的词典机制，将待识别语料中正确的实体准确且完整的识别出来，得到各种可能的实体候选集，也就是得到可能表达待识别语料含义的实体组合。

[0059] 步骤S104，按照预设逻辑限制条件，对获取的多个槽位组合进行过滤，得到候选槽位组合；

[0060] 在实际应用中，由于某些槽位的槽值字典所包含的词汇太多，且部分实体属于常用字，按照上述方式进行槽位匹配后，所得到的槽位组合往往比较多，这将会影响语料命名实体识别效率及准确性，所以，本申请可以针对一些槽值个数、出现次数、槽位之间的关系等逻辑，设定一些过滤条件，记为预设逻辑限制条件。

[0061] 之后，可以按照这些过滤条件，对待识别语料匹配到的多个槽位组合进行过滤，在不降低召回率的情况下，过滤掉一部分匹配度较低的槽位组合，保留最可能的槽位组合，记为候选槽位组合。需要说明，本申请对预设逻辑限制条件的内容不作限定，可以通过本行业的经验或试验结果确定，对于不同类型的待识别语料，所预设的逻辑限制条件内容可以不同。

[0062] 步骤S105，利用候选槽位组合对待识别语料进行标注，得到槽位标注序列；

[0063] 本实施例中，上述得到的每个槽位组合可以包括槽值及其在待识别语料中的位置，以及槽位类型等信息。经过上述过滤处理，得到若干个候选槽位组合后，可以利用每一个候选槽位组合中的槽位类型，对待识别语料中相应槽位值标注槽位类型，得到由多个槽位类型构成的槽位标注序列。

[0064] 之后，可以利用深度学习算法，从得到的槽位标注序列集中，获取与待识别语料的最相似的一个，确定最佳标注语料，以便由此得到待识别语料的命名实体。

[0065] 步骤S106，将待识别语料的词序列及槽位标注序列分别输入深度学习网络进行编码，得到待识别语料的第一编码向量及槽位标注序列的第二编码向量；

[0066] 本申请在Ranking部分，将利用深度学习网络，将初始语料即待识别语料的词序列，及槽位标注序列集转化为相应的编码向量，再用融合层计算这两个编码向量的相似度，以此得到待识别语料对应的最佳槽位标注序列，进而得到待识别语料的命名实体。

[0067] 其中，待识别语料的词序列可以是该待识别语料中的每个汉字、英文字字母组成，也就是说，本实施例可以将待识别语料每一个汉字作为一个序列元素，若存在英文单词，可以将每个英文字母作为一个序列元素，按照这种规则，获取待识别语料的词序列。

[0068] 可选的，上述深度学习网络可以是神经网络，具体的，对待识别语料的词序列进行

编码的神经网络可以是Attention Based on BiLSTM,对于槽位标注序列进行编码的神经网络可以是BiLSTM,但并不局限于本文给出的神经网络。

[0069] 在本实施例的实际应用中,参照图3所示的神经网络对输入序列的处理方法的流程示意图,由于待识别语料的词序列与槽位标注序列本质属于对同一内涵的不同描述,所以说,不能共享深度学习网络架构,即词序列和槽位标注序列不能同时输入一个神经网络,将采用两个的神经网络,分别对这两种序列进行处理。

[0070] 如图3所示,W1～W5可以表示待识别语料(即初始语料)的词序列的各元素,即汉字或英文字母,slot1～slot5表示槽位标注序列中的各元素,即对待识别语料标注后的各槽位类型,可以将一个槽位类型作为一个字符。应该理解,待识别语料包含W的数量并不局限于5个,可以根据实际获取的待识别语料内容,确定W的数量,本实施例仅以5个为例进行说明,同理吗,该对待识别语料标注得到的槽位标注序列包含的元素数量,即slot的数量,也并不局限于图3所示的5个,通常与待识别语料包含的W的数量相同。

[0071] 可选的,在实际应用中,可以通过统计大量语料的词序列组成元素数量,来设定神经网络输入序列的元素个数,即预设输入序列长度,在实际输入待识别语料的词序列和槽位标注序列时,若序列元素个数小于预设序列长度,可以在后面补零。本申请对上述预设输入序列长度的数值不做限定。

[0072] 本实施例中,W1～W5可以是待识别语料字的index输入,经过BiLSTM网络(即预先训练得到的槽位抽取模型)的处理后,可以得到每个时间单元的隐藏层,同时,引入attention机制调整不同时刻的隐藏层的权重,之后,计算调整后的权重与相应隐藏层的输出结果ht(t＝1、2、3、4、5)的点积,对各点积结果进行线性相加,得到待识别语料的编码向量C1,本实施例可以将其记为第一编码向量C1。

[0073] 并且,本实施例可以按照上文描述的利用BiLSTM模型对待识别语料的处理方式,利用BiLSTM网络(即预先训练得到的槽位抽取模型)对槽位标注序列进行相应处理,如图3所示,即将槽位标注序列slot1～slot5输入BiLSTM网络,可以得到该槽位标注序列的编码向量C2,本实施例记为第二编码向量,具体实现过程可以结合这种槽位抽取模型的工作原理确定,本实施例不做详述。

[0074] 其中,BiLSTM表示双向LSTM,其同时考虑了过去的特征(通过向前过程提取)和未来的特征(通过向后过程提取),该向后过程相当于将原始序列(如图3中的W1～W5,或slot1～slot5)逆向输入到LSTM中,因此,双向LSTM相当于两个LSTM,一个正向输入序列,一个反向输入序列,再将两者的输出结合起来作为最终的结果。本申请对BiLSTM网络对输入序列的具体处理过程不做详述,可以结合BiLSTM网络的原理实现。

[0075] 需要说明,对于上述分别对待识别语料和槽位标注序列进行处理的神经网络,可以利用神经网络算法,对大量样本数据训练得到,如从本应用平台或其他应用平台获取大量标注语料,按照上述Match部分的处理方法,针对每条标注语料召回10条(并不局限于10条)槽位标注序列作为样本数据,进而利用BiLSTM算法对召回的槽位标注序列及相应的标注语料进行训练,得到槽位抽取模型,即上述神经网络,本申请对该模型训练的具体实现过程不做详述。

[0076] 另外,对于本申请用来对待识别语料的词序列及槽位标注序列进行编码处理的网络,并不局限于上文给出的神经网络,还可以采用其他深度学习网络,本申请在此不作一一

详述。

[0077] 基于上文对深度学习网络对输入序列的处理过程的描述可知，在在深度学习网络的嵌入层，选取了待识别语料的词序列(也可以称为字向量)作为输入序列，并未引入外部词序列，避免了噪声的引入，提高了输出结果的准确性。而且深度学习层，对于待识别语料的词序列，与槽位标注序列分别采用了更合适的神经网络进行处理，不是共享同一网络架构，进一步提高了输出结果的准确性。在最后的融合层，选用余弦相似进行计算，避免引入其他参数，解决了采用转置点积参数矩阵进行处理，因该矩阵的参数对结果影响较大，容易过拟合，使得网络模型泛化能力下降的问题，也就是说，本申请采用的这种相似度计算方式，避免了过拟合问题，进而提高了获取待识别语料的最匹配槽位组合解的准确性。

[0078] 可选的，由于Ranking部分输入的待识别语料的词序列和槽位标注序列的长度相等，且该槽位标注序列是对待识别语料进行标注得到的，所以，本申请还可以采用区别于本实施例给出的深度学习网络架构，如将待识别语料的词序列和槽位标注序列直接进行卷积或融合等处理，具体实现过程本申请不作详述。

[0079] 步骤S107，计算获取的第一编码向量与各第二编码向量的相似度；

[0080] 步骤S108，选择最大相似度的第二编码向量对应的槽位标注序列为目标标注序列，并利用该目标标注序列，提取待识别语料的命名实体。

[0081] 可选的，本实施例可以采用相似度算法，实现两个编码向量之间的相似度计算，具体可以采用余弦相似算法实现，具体计算过程不做详述，这种相似度计算方式使得融合层不包含任何参数，避免了过拟合问题，且提高了模型泛化能力。

[0082] 结合上文对槽位标注序列的获取过程的描述，其是利用候选槽位组合对待识别语料进行标注得到的，那么，可以将目标标注序列对应的候选槽位组合确定为目标槽位组合，从而利用其包含的槽位值，得到待识别语料的命名实体，即将该槽位值和/或待识别语料包含的同义词作为命名实体，具体实现过程本实施例不做限定。

[0083] 综上所述，参照图4，本实施例将待识别语料的槽位抽取问题转化为召回-排序(Mach-Ranking)问题，并在对待识别语料进行召回处理过程中，采用字典匹配方式，召回待识别语料的可能候选槽位组合，并据此待识别语料进行标注，得到相应的槽位标注序列，之后，将待识别语料的词序列及槽位标注序列输入深度学习网络，分别获得待识别语料的编码向量，和各槽位标注序列的编码向量，选择与待识别语料的编码向量最相似的槽位标注序列的编码向量，并将其对应的槽位标注序列作为最佳槽位标注序列，据此得到待识别语料的命名实体。

[0084] 可见，本申请利用规则模板和深度学习算法的优点，实现了对各种语料命名实体的快速、简单且准确识别，不仅适用于简单实体识别，也能够适用于复杂、长尾及新出现的实体识别，解决了冷启动问题，且得到的识别结果通常是计算机语言，电子设备能够直接响应识别结果。

[0085] 基于上述分析，本申请将结合具体应用场景，来更加清楚地说明本申请实体识别方法，具体以待识别语料为“给我们来首欢快的歌吧”为例进行说明。为了准确提取该待识别语料中的命名实体，可以进行以下操作：

[0086] 本实施例可以先对“给我们来首欢快的歌吧”进行预处理，若其中包含了唤醒词，可以将唤醒词剔除，再将该待识别语料与槽位词典及同义词词典进行逐一匹配，保存匹配

到的可能的候选槽位组合。

[0087] 其中，对于同一槽位值，可能对应不同槽位，同一槽位的槽位值内容多样，比如“给我们”可以是歌曲名称，也可以是专辑名称等，基于此，待识别语料经过与槽位词典及同义词匹配后，往往会得到很多槽位组合，其中包含了一些与待识别语料的真实槽位组合匹配度很低的槽位组合，所以，本实施例还可以对召回的多个槽位组合做进一步过滤，得到若干候选槽位组合。

[0088] 在本实施例的同义词词典中，可以将放一首、放歌、来首等这类词汇映射成“播放”，所以，待识别语料“给我们来首欢快的歌吧”经过上述处理后，可以得到如下四组候选槽位组合：

[0089] {(3,4,‘播放’‘operation’),(0,2,‘给我们’‘songname’),(8,8,‘音乐’‘object’)}；

[0090] {(3,4,‘播放’‘operation’),(5,8,‘欢快的歌’‘songname’)；

[0091] {(3,4,‘播放’‘operation’),(8,8,‘音乐’‘object’),(5,6,‘欢快’‘style’)}；

[0092] {(5,6,‘欢快’‘style’),(8,8,‘音乐’‘object’),(1,5,‘我们’‘albumname’)}；

[0093] 其中，在上述候选槽位组合中，每个小括号中的内容依次表示待识别语料中的位置、经过同义词翻译后的槽位值、槽位类型。由此可见，本申请得到的候选槽位组合可以由槽位类型、槽位值及其在待识别语料中的位置等几部分组成，但并不局限于此。

[0094] 之后，可以依据上述得到四组候选槽位组合，分别对待识别语料进行标注，以得到对应的槽位标注序列，标注方式如下表一所示：

[0095] 表一

[0096]

给	我	们	来	首	欢	快	的	歌	吧
Song name	Song name	Song name	operation	operation	0	0	0	object	0
0	0	0	operation	operation	Song name	Song name	Song name	Song name	0
0	0	0	operation	operation	style	style	0	object	0
0	Album name	Album name	0	0	style	style	0	object	0

[0097] 在上表一种，第一行为待识别语料，后面的每一行表示一个候选槽位组合对该待识别语料的标注序列，本实施例可以将其作为后续Ranking部分的输入，以判断这四组候选槽位组合中最可能正确的一组槽位组合。

[0098] 基于上文举例描述的Mach部分的处理，在真实网络数据(语音助手)的实践中，以音乐域为例，通常其包括12种核心槽位，如歌手名、专辑名、风格类型、声音类型、音乐类型、场景、主题类型等，本实施例不再一一列举，经过上述步骤S102～步骤S105描述的对待识别语料的匹配和筛选处理，对待识别语料的槽位组合的召回率可以达到89.6%左右，平均一句待识别语料能够召回11组槽位候选组合。需要说明，该召回率以及每句召回的候选槽位组合的数量，并不局限于本文列举的示例，可以根据实际情况进行计算。

[0099] 经研究得知，导致槽位组合召回率低的原因，主要是标注语料质量、上游语音识别ASR结果误差、语料本身存在歧义、预设的逻辑限定条件严格等原因，所以，为了提高槽位组合召回率，可以从这几方面进行优化，具体实现过程本实施例不作详述。

[0100] 基于上文利用深度学习网络对待识别语料及槽位标注序列(即标注句子)的处理过程的描述，对于“给我们来首欢快的歌吧”这一待识别语料，将其中的每个汉字作为一个字符，依次记为W1、W2、…、W10，同理将槽位标注序列中的元素依次记为slot1、slot2、…、slot10，根据实际需要，若深度学习网络的输入序列的元素数量预设为14个，那么，对上述得到的词序列和槽位标注序列进行扩展，即在最后四位补0，从而形成符合网络输入要求的序列。

[0101] 之后，可以按照如图3所示的编码方式，得到待识别语料的编码向量，以及各槽位标注序列对应的编码向量，再分别计算待识别语料的编码向量，与各槽位标注序列对应的编码向量的相似度，以确定与待识别语料最匹配的槽位组合。

[0102] 本实施例中，如图4所示，经过融合层的相似度计算后，输出的结果Y可以是一个数值，比如0或1，0可以表示槽位组合序列对应的候选槽位组合，与待识别语料的真实槽位组合不同，即该候选槽位组合与待识别语料的匹配度较低；1可以表示槽位组合序列对应的候选槽位组合，与待识别语料的真实槽位组合相同，即该候选槽位组合与待识别语料的匹配度较高，如上文举例中得到的四个候选槽位组合，第三个候选槽位组合对应得到的Y＝1；其他候选槽位组合对应得到的Y＝0，但并不局限于此。

[0103] 基于此，本实施例可以基于深度学习网络输出的Y的数值，直接确定待识别语料最匹配的槽位组合解，即Y＝1对应的候选槽位组合，进而可以据此得到待识别语料的命名实体，如上述举例，利用第三个候选槽位组合，得到的待识别语料的命名实体可以为“播放”、“音乐”和“欢快”。

[0104] 在本申请中，上述深度学习网络的架构可以预先训练得到，训练所使用的训练语料可以是企业内部某应用中的100w+条标注语料，按照上述mach部分描述的处理方式，平均每条训练语料可以找回10条槽位标注句子(即槽位标注序列)，之后，将其在测试集上测试，以得到满足一定收敛要求的深度学习网络模型，以便在实际应用中，直接调用该深度学习网络模型使用。经过试验，测试集在Ranking部分的网络模型的槽位组合全等率可达到98.4%，输出结果的准确率以及召回率均可达到98%以上，具体数值不做限定。

[0105] 可见，相比于单纯使用规则模板需要堆叠海量模板，单纯深度学习的几乎不可用的测试结果，本申请提出利用规则模板+深度学习实现的实体识别方法，既可以有效借用模板对复杂实体的识别度，在减少召回数量的同时有效提升了召回能力，又可以借用深度学习对通用语法规则的泛化能力。在降低复杂度的同时，有效保证了算法性能指标，使其在复杂领域的槽位抽取上取得了良好的效果。

[0106] 图5所示，为本申请提供的一种实体识别装置的结构示意图，所述装置包括：

[0107] 第一获取模块21，用于获取待识别语料；

[0108] 第二获取模块22，用于利用槽值匹配方式，获取所述待识别语料的候选槽位组合；

[0109] 可选的，如图6所示，所述第二获取模块22可以包括：

[0110] 第一获取单元221，用于利用槽值匹配方式，获取所述待识别语料的多个槽位组合；

[0111] 过滤单元222,用于按照预设逻辑限制条件,对所述多个槽位组合进行过滤,得到候选槽位组合。

[0112] 其中,第一获取单元可以包括:

[0113] 字典获取单元,用于获取槽值字典及同义词字典;

[0114] 匹配单元,用于将所述待识别语料与所述槽值字典及所述同义词字典进行逐一匹配,得到多个槽位组合。

[0115] 标注模块23,用于利用所述候选槽位组合对所述待识别语料进行标注,得到相应的槽位标注序列;

[0116] 第三获取模块24,用于基于深度学习网络,分别对所述待识别语料的词序列及所述槽位标注序列进行编码,得到待识别语料的第一编码向量,以及槽位标注序列的第二编码向量;

[0117] 可选的,如图7所示,该第三获取模块24可以包括:

[0118] 序列获取单元241,用于获取所述待识别语料的词序列;

[0119] 第一编码单元242,用于将所述词序列输入第一神经网络模型进行编码,得到第一编码向量;

[0120] 第二编码单元243,用于将所述槽位标注序列输入第二神经网络模型进行编码,得到第二编码向量。

[0121] 其中,第一神经网络模型和所述第二神经网络模型是不同类型的双向长短期记忆网络BiLSTM,但并不局限于此。

[0122] 识别模块25,用于基于所述第一编码向量与所述第二编码向量的相似度,确定所述槽位标注序列中的目标标注序列,并利用所述目标标注序列,得到所述待识别语料的命名实体。

[0123] 可选的,所述识别模块25可以包括:

[0124] 相似度计算单元,用于计算所述第一编码向量与所述第二编码向量的相似度;

[0125] 选择单元,用于选择最大相似度的第二编码向量对应的槽位标注序列为目标标注序列。

[0126] 本申请实施例还提供了一种存储介质,其上存储有计算机程序,该计算机程序被处理器执行,实现上述实体识别方法的各步骤,具体实现过程可以参照上述方法实施例的描述,本实施例在此不作赘述。

[0127] 如图8所示,本申请实施例还提供了一种计算机设备的硬件结构示意图,该计算机设备可以是上述服务器,可以包括通信接口31、存储器32和处理器33;

[0128] 在本申请实施例中,通信接口31、存储器32和处理器33可以通过通信总线实现相互间的通信,且通信接口31、存储器32和处理器33及通信总线的数量可以为至少一个。

[0129] 可选的,通信接口31可以为通信模块的接口,如GSM模块的接口;

[0130] 处理器33可能是一个中央处理器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申请实施例的一个或多个集成电路。

[0131] 存储器32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。

[0132] 其中，存储器32存储有程序，处理器33调用存储器32所存储的程序，以实现上述实体识别方法的各个步骤。

[0133] 在实际应用中，服务器得到待识别语料的命名实体之后，可以将其发送至相应的客户端，以使客户端按照该命名实体进行后续操作，或者，服务器可以按照得到的命名实体进行数据搜索，再将搜索到的相关数据反馈至客户端输出等，本申请对服务器得到待识别语料的命名实体之后的应用场景不做限定。

[0134] 需要说明，本说明书中各个实施例采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似部分互相参见即可。对于实施例公开的装置、服务器而言，由于其与实施例公开的方法相对应，所以描述的比较简单，相关之处参见方法部分说明即可。

[0135] 专业人员还可以进一步意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、计算机软件或者二者的结合来实现，为了清楚地说明硬件和软件的可互换性，在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。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申请的范围。

[0136] 结合本文中所公开的实施例描述的方法或算法的步骤可以直接用硬件、处理器执行的软件模块，或者二者的结合来实施。软件模块可以置于随机存储器(RAM)、内存、只读存储器(ROM)、电可编程ROM、电可擦除可编程ROM、寄存器、硬盘、可移动磁盘、CD-ROM、或技术领域内所公知的任意其它形式的存储介质中。

[0137] 对所公开的实施例的上述说明，使本领域专业技术人员能够实现或使用本申请。对这些实施例的多种修改对本领域的专业技术人员来说将是显而易见的，本文中所定义的一般原理可以在不脱离本申请的核心思想或范围的情况下，在其它实施例中实现。因此，本申请将不会被限制于本文所示的这些实施例，而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

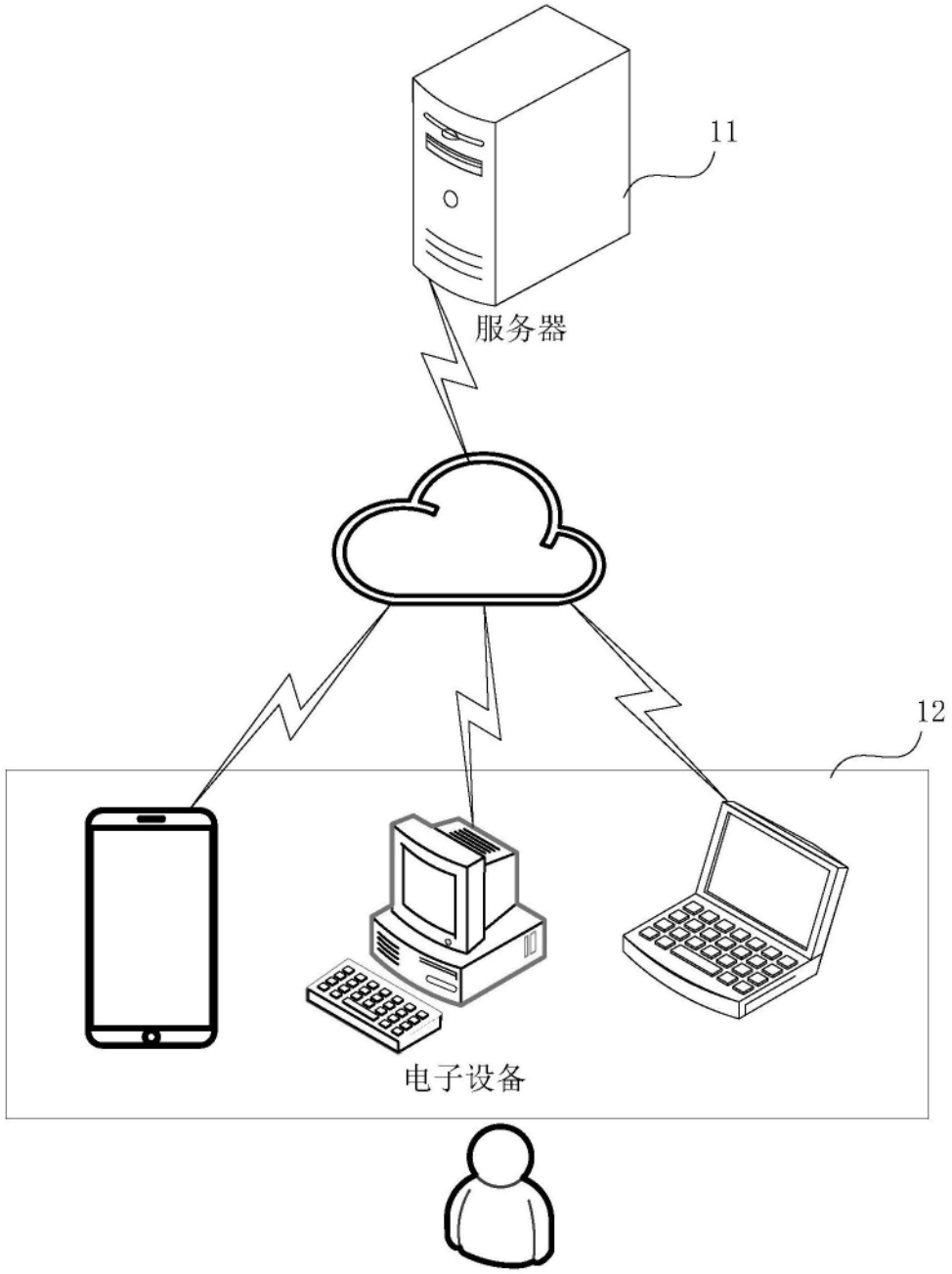


图1

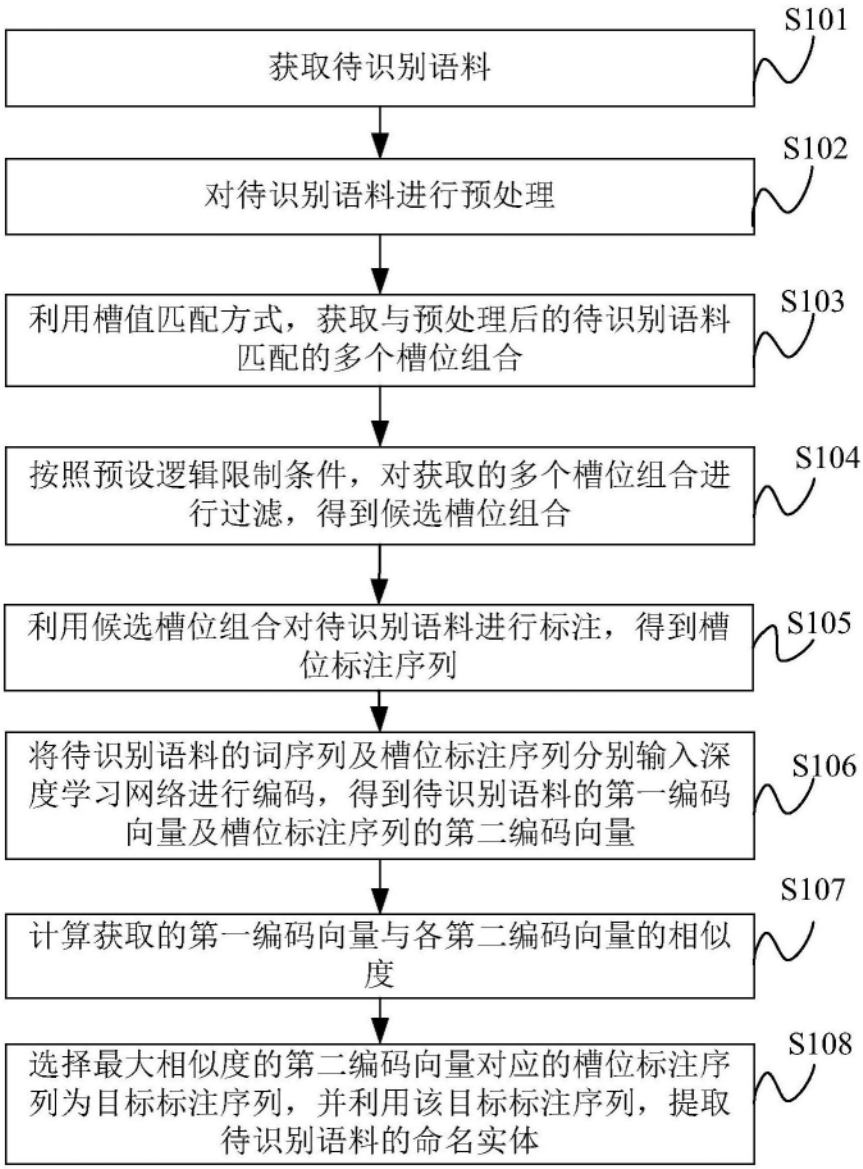

S101
获取待识别语料
S102
对待识别语料进行预处理
S103
利用槽值匹配方式，获取与预处理后的待识别语料匹配的多个槽位组合
S104
按照预设逻辑限制条件，对获取的多个槽位组合进行过滤，得到候选槽位组合
S105
利用候选槽位组合对待识别语料进行标注，得到槽位标注序列
S106
将待识别语料的词序列及槽位标注序列分别输入深度学习网络进行编码，得到待识别语料的第一编码向量及槽位标注序列的第二编码向量
S107
计算获取的第一编码向量与各第二编码向量的相似度
S108
选择最大相似度的第二编码向量对应的槽位标注序列为目标标注序列，并利用该目标标注序列，提取待识别语料的命名实体

图2

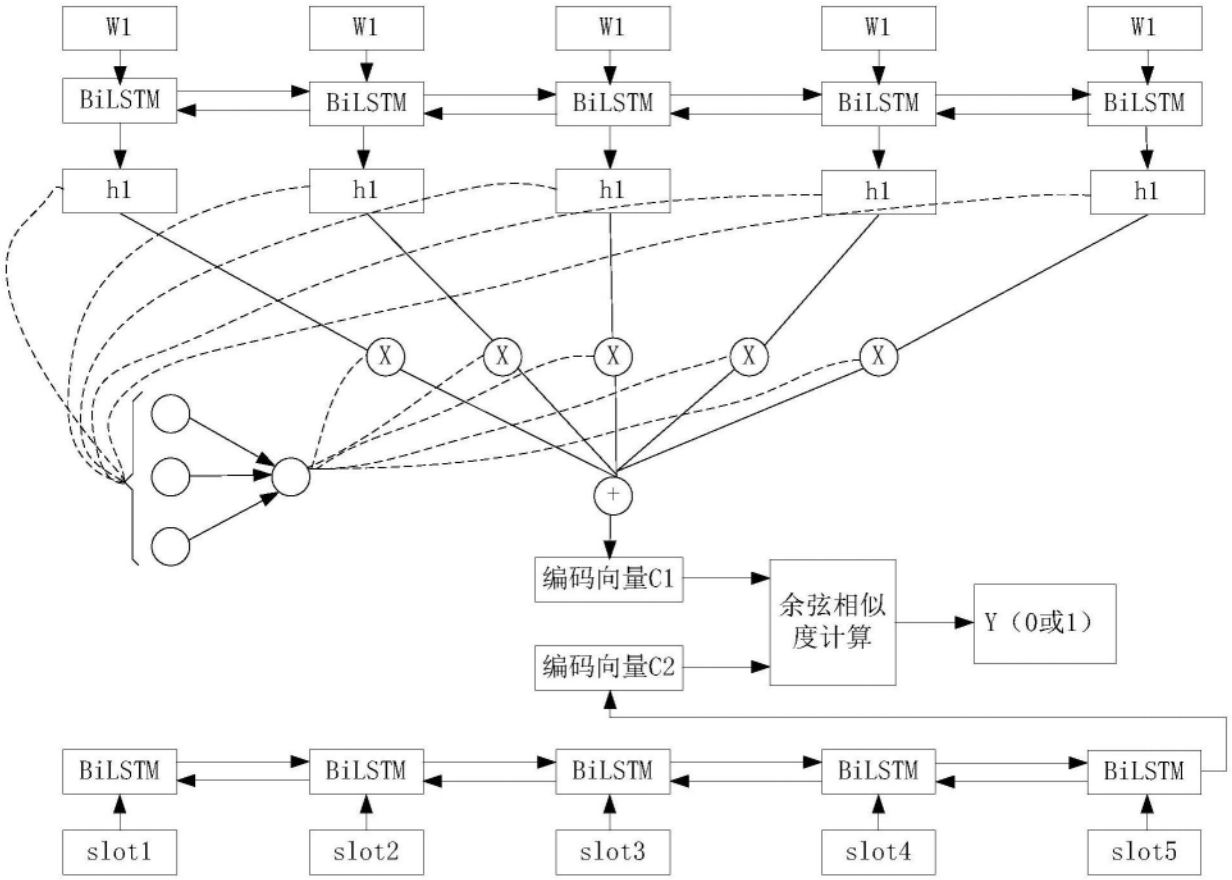

W1
BiLSTM
h1
X
+
编码向量C1
编码向量C2
余弦相似度计算
Y（0或1）
slot1
slot2
slot3
slot4
slot5

图3

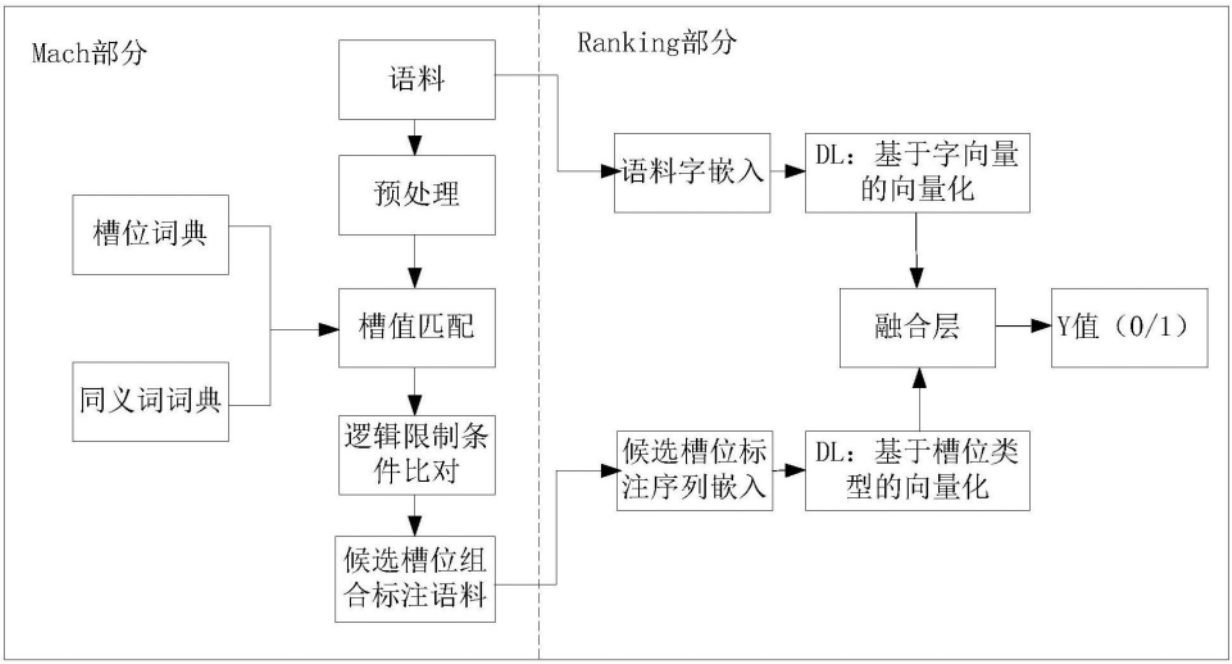

Mach部分
Ranking部分
语料
预处理
槽位词典
同义词词典
槽值匹配
逻辑限制条件比对
候选槽位组合标注语料
语料字嵌入
DL：基于字向量的向量化
融合层
Y值（0/1）
候选槽位标注序列嵌入
DL：基于槽位类型的向量化

图4

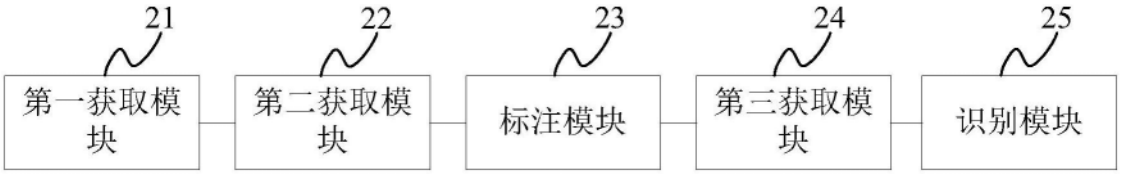

21
22
23
24
25
第一获取模块
第二获取模块
标注模块
第三获取模块
识别模块

图5

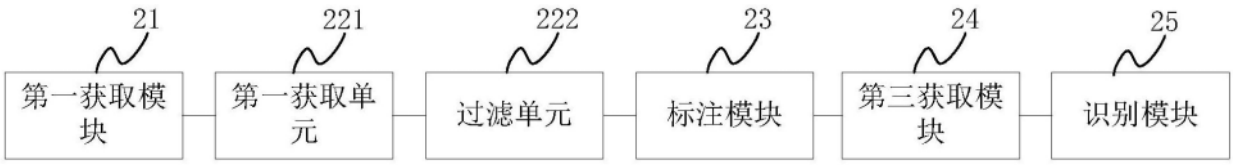

21
221
222
23
24
25
第一获取模块
第一获取单元
过滤单元
标注模块
第三获取模块
识别模块

图6

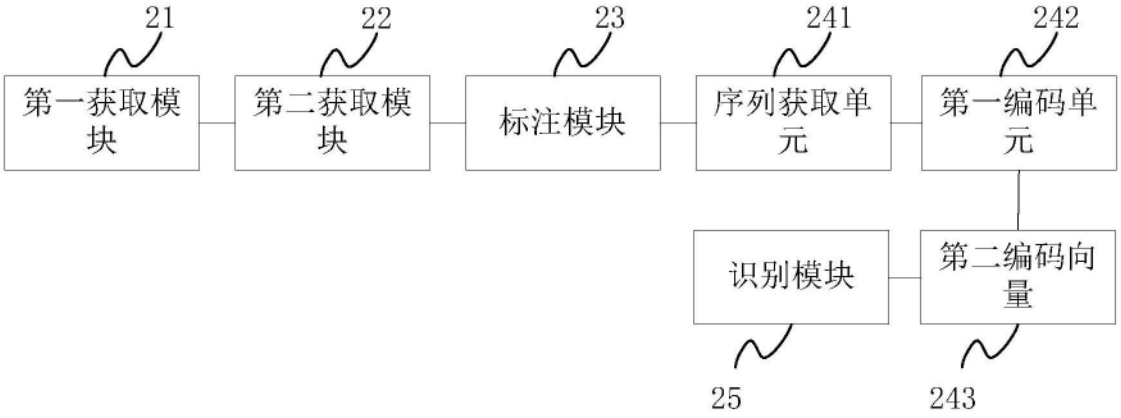

21
22
23
241
242
第一获取模块
第二获取模块
标注模块
序列获取单元
第一编码单元
识别模块
第二编码向量
25
243

图7

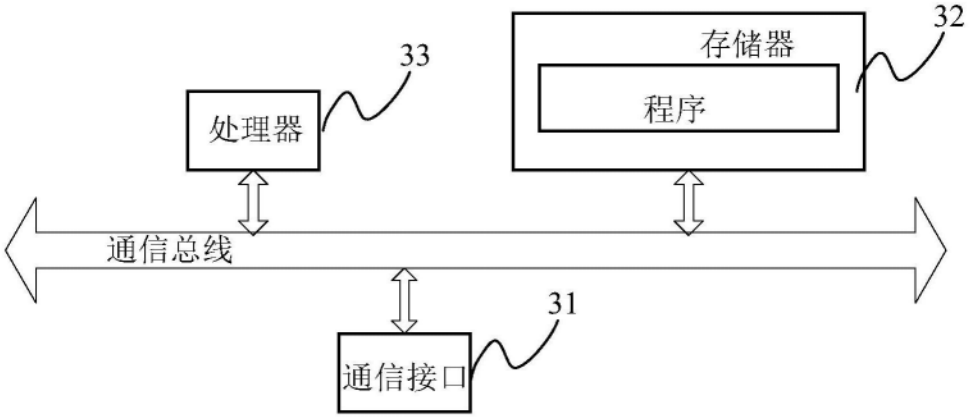

32
存储器
程序
33
处理器
通信总线
31
通信接口

图8