



(12)发明专利

(10)授权公告号 CN 105279552 B

(45)授权公告日 2018.06.22

(21)申请号 201410273269.7

(22)申请日 2014.06.18

(65)同一申请的已公布的文献号
申请公布号 CN 105279552 A

(43)申请公布日 2016.01.27

(73)专利权人 清华大学
地址 100084 北京市海淀区北四环中路229号海泰大厦5层527室
专利权人 北京中科汇联信息技术有限公司

(72)发明人 王东 游世学 刘荣 乔亚飞

(74)专利代理机构 北京润泽恒知识产权代理有限公司 11319
代理人 苏培华

(51)Int.Cl.
G06N 3/02(2006.01)

(56)对比文件

CN 103400577 A,2013.11.20,
CN 103810999 A,2014.05.21,
CN 1677388 A,2005.10.05,
EP 0443976 A2,1991.08.28,
Xiaorong Cheng etal..Application of MPSO-Based Neural Network Model in Chinese Word Segmentation.《Intelligent Computation Technology and Automation, 2009. ICICTA '09. Second International Conference on》.2009,
T Mikolov etal..Distributed Representations of Words and Phrases and their Compositionality.《Advances in Neural Information Processing Systems》.2013,第26卷

审查员 李梦芸

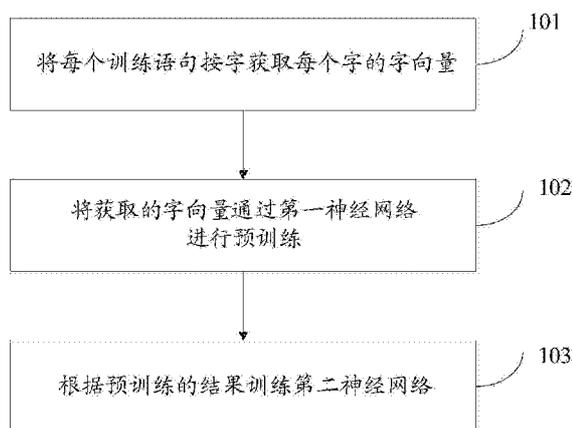
权利要求书2页 说明书8页 附图3页

(54)发明名称

一种基于字的神经网络的训练方法和装置

(57)摘要

本申请提供了一种基于字的神经网络的训练方法和装置,包括:将每个训练语句按字获取每个字的字向量;将所述字向量作为第一神经网络的参数输入到所述第一神经网络中进行预训练,获得预训练结果,其中,所述结果包括字向量的上文特征向量和字向量下文的特征向量;将所述字向量的上文特征向量和字向量下文的特征向量作为第二神经网络的参数输入到所述第二神经网络中,训练所述第二神经网络。因此本申请解决了基于词的神经网络语言模型中训练效率低的问题。



1. 一种基于字的神经网络的训练方法,其特征在于,包括:

将每个训练语句按字获取每个字的字向量;

将所述字向量作为第一神经网络的参数输入到所述第一神经网络中进行预训练,获得预训练结果,其中,所述结果包括字向量的上文特征向量和字向量下文的特征向量;

将所述字向量的上文特征向量和字向量下文的特征向量作为第二神经网络的参数输入到所述第二神经网络中,训练所述第二神经网络;

所述第一神经网络包括输入层、映射层和输出层;

所述将所述字向量作为第一神经网络的参数输入到所述第一神经网络中进行预训练,获得预训练结果的步骤包括:

为训练语句中每个字随机分配字向量;

输入层接收训练语句中每个字随机分配字向量,所述输入层将接收的每个字的字向量投影到映射层,得到投影后的每个字向量;

计算投影后的每个字向量在字的历史子串中出现的概率,将所述概率作为输出层的预训练结果,所述结果包括字向量的上文特征向量和字向量下文的特征向量;

所述第二神经网络包括输入层、映射层、隐藏层和输出层;

所述第二神经网络的输入层接收第一神经网络的输出层输出的字向量的上文特征向量和字向量下文的特征向量;

所述第二神经网络的所述输入层将接收的字向量的上文特征向量和字向量下文的特征向量进行线性变换投影,将投影后的字向量的上文特征向量和字向量下文的特征向量首尾相接,将拼接后的结果作为所述第二神经网络的映射层的输入;

将映射层的输入进行线性变换和第一非线性激活函数处理后得到所述第二神经网络的隐藏层的输入;

将所述第二神经网络的隐藏层的输入进行线性变换和第二非线性激活函数处理后得到所述第二神经网络的输出层的输出结果,所述输出结果包括字向量在字的历史子串中的概率。

2. 根据权利要求1所述的方法,其特征在于,通过以下公式计算投影后的每个字向量在字的历史子串中出现的概率:

$$p(e_{t+j}|e_t) = \frac{\exp(c_{e_{t+j}}c_{e_t})}{\sum_e \exp(c_e c_{e_t})}$$

其中, c_e 为训练语句的字向量, c 为上下文窗长, \sum_e 表示对所有汉字进行求和, e_t 为训练语句的当前选中的字, e_{t+j} 为训练语句的当前选中的字的前面第 j 个字或后面第 j 个字。

3. 一种基于字的神经网络的训练装置,其特征在于,包括:

获取模块,用于将每个训练语句按字获取每个字的字向量;

预训练模块,用于将所述字向量作为第一神经网络的参数输入到所述第一神经网络中进行预训练,获得预训练结果,其中,所述结果包括字向量的上文特征向量和字向量下文的特征向量;

训练模块,用于将所述字向量的上文特征向量和字向量下文的特征向量作为第二神经网络的参数输入到所述第二神经网络中,训练所述第二神经网络;

所述第一神经网络包括输入层、映射层和输出层；

所述预训练模块包括：

分配模块，用于为训练语句中每个字随机分配字向量；

投影模块，用于输入层接收训练语句中每个字随机分配字向量，所述输入层将接收的每个字的字向量投影到映射层，得到投影后的每个字向量；

计算模块，用于计算投影后的每个字向量在字的历史子串中出现的概率，将所述概率作为输出层的预训练结果，所述结果包括字向量的上文特征向量和字向量下文的特征向量；

所述第二神经网络包括输入层、映射层、隐藏层和输出层；

所述第二神经网络的输入层接收第一神经网络的输出层输出的字向量的上文特征向量和字向量下文的特征向量；

所述第二神经网络的所述输入层将接收的字向量的上文特征向量和字向量下文的特征向量进行线性变换投影，将投影后的字向量的上文特征向量和字向量下文的特征向量首尾相接，将拼接后的结果作为所述第二神经网络的映射层的输入；

将映射层的输入进行线性变换和第一非线性激活函数处理后得到所述第二神经网络的隐藏层的输入；

将所述第二神经网络的隐藏层的输入进行线性变换和第二非线性激活函数处理后得到所述第二神经网络的输出层的输出结果，所述输出结果包括字向量在字的历史子串中的概率。

4. 根据权利要求3所述的装置，其特征在于，所述计算模块通过以下公式计算投影后的每个字向量在字的历史子串中出现的概率：

$$p(e_{t+j}|e_t) = \frac{\exp(c_{e_{t+j}}c_{e_t})}{\sum_e \exp(c_e c_{e_t})}$$

其中， c_e 为训练语句的字向量， c 为上下文窗长， \sum_e 表示对所有汉字进行求和， e_t 为训练语句的当前选中的字， e_{t+j} 为训练语句的当前选中的字的前面第 j 个或后面第 j 个字。

一种基于字的神经网络的训练方法和装置

技术领域

[0001] 本申请涉及自然语言处理领域,特别是涉及一种基于字的神经网络的训练方法和装置。

背景技术

[0002] 在自然语言处理和语音识别等领域,语言模型用于描述语言内部的词语搭配关系。一个高质量的语言模型对连续语音识别和机器翻译等都具有重要意义。

[0003] 当前主流的语言模型是基于概率的统计语言模型(n-gram),该统计语言模型显著缺陷是:对于新词或者低频词难以得到有效的概率统计。基于此人们发明了各种平滑算法,如discount,back-off,interpolation等。上述方法在一定程度上改善了n-gram在低频词上的性能,但是由于统计语言模型本身的缺陷,始终无法从根本上解决。

[0004] 为了解决统计语言模型中存在的问题,提出了基于词的神经网络语言模型统计低频词的概率。目前基于词的神经网络语言模型存在以下缺陷:

[0005] 1、现有的汉字可以组成数以万计的词表,基于词的神经网络模型需要对组成的数以万计的词表进行训练,导致神经网络语言模型训练时间长,训练效率低。当词表增大到超过10万以上时,训练时间变得难以接受。

[0006] 2、中文是动态语言,新词不断涌现。目前基于词的神经网络语言模型无法处理这些新词,同时不断增加的新词使得神经网络语言模型的训练越来越复杂,即使要重新训练神经网络语言模型也极为困难。

[0007] 因此,基于词的神经网络语言模型中存在训练时间长,进而导致训练效率低的问题。

发明内容

[0008] 本申请提供一种基于字的神经网络的训练方法和装置,以解决基于词的神经网络语言模型中训练效率低的问题。

[0009] 为了解决上述问题,本申请公开了一种基于字的神经网络的训练方法,包括:将每个训练语句按字获取每个字的字向量;

[0010] 将所述字向量作为第一神经网络的参数输入到所述第一神经网络中进行预训练,获得预训练结果,其中,所述结果包括字向量的上文特征向量和字向量下文的特征向量;

[0011] 将所述字向量的上文特征向量和字向量下文的特征向量作为第二神经网络的参数输入到所述第二神经网络中,训练所述第二神经网络。

[0012] 优选地,所述第一神经网络包括输入层、隐藏层和输出层。

[0013] 优选地,所述将所述字向量作为第一神经网络的参数输入到所述第一神经网络中进行预训练,获得预训练结果的步骤包括:

[0014] 为训练语句中每个字随机分配字向量;

[0015] 输入层接收训练语句中每个字随机分配字向量,所述输入层将接收的每个字的字

向量投影到映射层,得到投影后的每个字向量;

[0016] 计算投影后的每个字向量在字的历史子串中出现的概率,将所述概率作为输出层的预训练结果,所述结果包括字向量的上文特征向量和字向量下文的特征向量。

[0017] 优选地,通过以下公式计算投影后的每个字向量在字的历史子串中出现的概率:

$$[0018] \quad \frac{1}{T} \sum_{t=1}^T \sum_{-c \leq j \leq c, j \neq 0} \log p(e_{t+j}|e_t)$$

[0019] 其中,T为训练语句的长度,c为上下文窗长, $p(e_{t+j}|e_t)$ 为字向量在字的历史子串中出现的概率, $p(e_{t+j}|e_t)$ 通过以下公式获得:

$$[0020] \quad p(e_{t+j}|e_t) = \frac{\exp(c_{e_{t+j}}c_{e_t})}{\sum_e \exp(c_e c_{e_t})}$$

[0021] 其中, c_e 为训练语句的字向量。

[0022] 优选地,所述第二神经网络包括输入层、映射层、隐藏层和输出层;

[0023] 所述第二神经网络的输入层接收第一神经网络的输出层输出的字向量的上文特征向量和字向量下文的特征向量;

[0024] 所述第二神经网络的所述输入层将接收的字向量的上文特征向量和字向量下文的特征向量进行线性变换投影,将投影后的字向量的上文特征向量和字向量下文的特征向量首尾相接,将拼接后的结果作为所述第二神经网络的映射层的输入;

[0025] 将映射层的输入进行线性变换和第一非线性激活函数处理后得到所述第二神经网络的隐藏层的输入;

[0026] 将所述第二神经网络的隐藏层的输入进行线性变换和第二非线性激活函数处理后得到所述第二神经网络的输出层的输出结果,所述输出结果包括字向量在字的历史子串中的概率。

[0027] 为了解决上述问题,本申请还公开了一种基于字的神经网络的训练装置,包括:获取模块,用于将每个训练语句按字获取每个字的字向量;

[0028] 预训练模块,用于将所述字向量作为第一神经网络的参数输入到所述第一神经网络中进行预训练,获得预训练结果,其中,所述结果包括字向量的上文特征向量和字向量下文的特征向量;

[0029] 训练模块,用于将所述字向量的上文特征向量和字向量下文的特征向量作为第二神经网络的参数输入到所述第二神经网络中,训练所述第二神经网络。

[0030] 优选地,所述第一神经网络包括输入层、隐藏层和输出层。

[0031] 优选地,所述预训练模块包括:

[0032] 分配模块,用于为训练语句中每个字随机分配字向量;

[0033] 投影模块,用于输入层接收训练语句中每个字随机分配字向量,所述输入层将接收的每个字的字向量投影到映射层,得到投影后的每个字向量;

[0034] 计算模块,用于计算投影后的每个字向量在字的历史子串中出现的概率,将所述概率作为输出层的预训练结果,所述结果包括字向量的上文特征向量和字向量下文的特征向量。

[0035] 优选地,所述计算模块通过以下公式计算投影后的每个字向量在字的历史子串中出现的概率:

$$[0036] \quad \frac{1}{T} \sum_{t=1}^T \sum_{-c \leq j \leq c, j \neq 0} \log p(e_{t+j}|e_t)$$

[0037] 其中,T为训练语句的长度,c为上下文窗长, $p(e_{t+j}|e_t)$ 为字向量在字的历史子串中出现的概率, $p(e_{t+j}|e_t)$ 通过以下公式获得:

$$[0038] \quad p(e_{t+j}|e_t) = \frac{\exp(c_{e_{t+j}}c_{e_t})}{\sum_e \exp(c_e c_{e_t})}$$

[0039] 其中, c_e 为训练语句的字向量。

[0040] 优选地,所述第二神经网络包括输入层、映射层、隐藏层和输出层;

[0041] 所述第二神经网络的输入层接收第一神经网络的输出层输出的字向量的上文特征向量和字向量下文的特征向量;

[0042] 所述第二神经网络的所述输入层将接收的字向量的上文特征向量和字向量下文的特征向量进行线性变换投影,将投影后的字向量的上文特征向量和字向量下文的特征向量首尾相接,将拼接后的结果作为所述第二神经网络的映射层的输入;

[0043] 将映射层的输入进行线性变换和第一非线性激活函数处理后得到所述第二神经网络的隐藏层的输入;

[0044] 将所述第二神经网络的隐藏层的输入进行线性变换和第二非线性激活函数处理后得到所述第二神经网络的输出层的输出结果,所述输出结果包括字向量在字的历史子串中的概率。

[0045] 与现有技术相比,本申请包括以下优点:

[0046] 首先,本申请将每个训练语句按字获取每个字的字向量,将获取的字向量通过第一神经网络进行预训练,获得预训练结果,其中,预训练结果包括字向量的上文特征向量和字向量下文的特征向量。将所述字向量的上文特征向量和字向量下文的特征向量作为第二神经网络的参数输入,训练第二神经网络。由于汉语中汉字的数量是固定的,常用的汉字数量大约在6千左右,远远小于词语的数量,因此可以降低神经网络的复杂度,提高了语言模型的训练效率。同时,对于不断涌现的新词,可以通过将新词按字拆分成字串,因此无需对出现的新词进行重新训练语言模型,从而提高了语言模型的训练效率。

[0047] 其次,本申请将获取的每个字向量通过第一神经网络进行预训练,获得了每个字向量在上下文中的特征向量,然后将获得的特征向量在通过第二神经网络进行训练,由于获取的字向量通过第一神经网络和第二神经网络的双重处理,有效降低了汉字多义出现的概率,从而解决了汉字本身具有多义性的问题。

附图说明

[0048] 图1是本申请实施例一中的一种基于字的神经网络的训练方法的流程图;

[0049] 图2是本申请实施例二中的一种基于字的神经网络的训练方法的流程图;

[0050] 图3是本申请中的第一神经网络的示意图;

- [0051] 图4是本申请中的第二神经网络的示意图；
- [0052] 图5是本申请中的第一神经网络和第二神经网络结合的示意图；
- [0053] 图6是本申请实施例三中的一种基于字的神经网络的训练装置的结构框图。

具体实施方式

[0054] 为使本申请的上述目的、特征和优点能够更加明显易懂，下面结合附图和具体实施方式对本申请作进一步详细的说明。

[0055] 参照图1，示出了本申请实施例一中的一种基于字的神经网络的训练方法，包括：

[0056] 步骤101：将每个训练语句按字获取每个字的字向量。

[0057] 例如：训练语句为“电脑手机”，将该训练语句按字识别后的训练语句为：电/脑/手/机/，然后获取每个字的字向量。

[0058] 步骤102：将所述字向量作为第一神经网络的参数输入到所述第一神经网络中进行预训练，获得预训练结果，其中，所述结果包括字向量的上文特征向量和字向量下文的特征向量。

[0059] 对字向量进行预处理是为了获得更具有区分性和表达性的字向量。将获取的每个字向量通过第一神经网络进行预训练，获得了每个字向量在上下文中的特征向量，然后将获得的特征向量在通过第二神经网络进行训练，由于获取的字向量通过第一神经网络和第二神经网络的双重处理，从而对字的多义性具有更强的区分能力，有效降低了汉字多义出现的概率。

[0060] 步骤103：将所述字向量的上文特征向量和字向量下文的特征向量作为第二神经网络的参数输入到所述第二神经网络中，训练所述第二神经网络。

[0061] 通过本实施例，首先，本申请将每个训练语句按字获取每个字的字向量，将获取的字向量通过第一神经网络进行预训练，获得预训练结果，其中，预训练结果包括字向量的上文特征向量和字向量下文的特征向量。将所述字向量的上文特征向量和字向量下文的特征向量作为第二神经网络的参数输入，训练第二神经网络。由于汉语中汉字的数量是固定的，常用的汉字数量大约在6千左右，远远小于词语的数量，因此可以降低神经网络的复杂度，提高了语言模型的训练效率。同时，对于不断涌现的新词，可以通过将新词按字拆分成字串，因此无需对出现的新词进行重新训练语言模型，从而提高了语言模型的训练效率。

[0062] 其次，本申请将获取的每个字向量通过第一神经网络进行预训练，获得了每个字向量在上下文中的特征向量，然后将获得的特征向量在通过第二神经网络进行训练，由于获取的字向量通过第一神经网络和第二神经网络的双重处理，有效降低了汉字多义出现的概率，从而解决了汉字本身具有多义性的问题。

[0063] 参照图2，示出了本申请实施例二中的一种基于字的神经网络的训练方法。

[0064] 本实施例中，一种基于字的神经网络的训练方法，包括：

[0065] 步骤201：将每个训练语句按字获取每个字的字向量。

[0066] 步骤202：将所述字向量作为第一神经网络的参数输入到所述第一神经网络中进行预训练，获得预训练结果，其中，所述结果包括字向量的上文特征向量和字向量下文的特征向量。

[0067] 如图3所示为本申请中的第一神经网络的示意图。其中，第一神经网络包括输入

层、映射层和输出层。输入层和映射层的目的是为了训练字向量。

[0068] 将所述字向量作为第一神经网络的参数输入到所述第一神经网络中进行预训练，获得预训练结果的步骤包括：

[0069] 子步骤2021：为训练语句中每个字随机分配字向量。

[0070] 子步骤2022：输入层接收训练语句中每个字随机分配字向量 $e(t)$ ，所述输入层将接收的每个字的字向量投影到映射层，得到投影后的每个字向量。

[0071] 子步骤2023：计算投影后的每个字向量在字的历史子串中出现的概率，将所述概率作为输出层的预训练结果，所述预训练结果为 $e(t-1)$ 、 $e(t+1)$ 、 $e(t-2)$ 和 $e(t+2)$ ，所述结果包括字向量的上文特征向量和字向量下文的特征向量。

[0072] 也即，不仅对历史信息进行优化，而且还对未来信息进行优化。

[0073] 每个字出现的概率依赖于之前出现的所有字，这一子串定义为历史子串。

[0074] 优选地，通过以下公式计算投影后的每个字向量在字的历史子串中出现的概率：

$$[0075] \quad \frac{1}{T} \sum_{t=1}^T \sum_{-c \leq j \leq c, j \neq 0} \log p(e_{t+j} | e_t) \quad (1)$$

[0076] 其中， T 为训练语句的长度， c 为上下文窗长， $p(e_{t+j} | e_t)$ 为字向量在字的历史子串中出现的概率， $p(e_{t+j} | e_t)$ 通过以下公式获得：

$$[0077] \quad p(e_{t+j} | e_t) = \frac{\exp(c_{e_{t+j}} c_{e_t})}{\sum_e \exp(c_e c_{e_t})} \quad (2)$$

[0078] 其中， c_e 为训练语句的字向量，由第一神经网络中的映射层得到。公式(2)中的 Σ_e 表示对所有汉字进行求和。通过映射层的连接权重对第一神经网络进行优化。

[0079] 下面通过一个具体的实例来说明第一神经网络进行预训练的过程。

[0080] 例如：训练语句为“我现在干什么呢”

[0081] 1. 首先初始化字向量：随机的分配每个字的字向量。

[0082] 我：[0.3 0.7 0.3 0.6……]

[0083] 现：[0.6 0.3 0.67 0.45……]

[0084] 2. 训练，通过以下公式计算投影后的每个字向量在字的历史子串中出现的概率。

$$[0085] \quad \frac{1}{T} \sum_{t=1}^T \sum_{-c \leq j \leq c, j \neq 0} \log p(e_{t+j} | e_t)$$

[0086] 其中，以“干”为例计算其在训练语句中的上文和下文中出现的概率。

[0087] 即： $\log p(\text{我} | \text{干}) + \log p(\text{现} | \text{干}) + \log p(\text{在} | \text{干}) + \log p(\text{什} | \text{干}) + \log p(\text{么} | \text{干}) + \log p(\text{呢} | \text{干})$ 取值最大化，通过迭代优化第一神经网络的参数。

$$[0088] \quad p(e_{t+j} | e_t) = \frac{\exp(c_{e_{t+j}} c_{e_t})}{\sum_e \exp(c_e c_{e_t})}$$

[0089] $P(\text{我} | \text{干}) = (\exp(c(\text{我}) * c(\text{干}))) / (\exp(c(\text{我}) * c(\text{干})) + \exp(c(\text{我}) * c(\text{现})) + \exp(c(\text{我}) * c(\text{在})) + \exp(c(\text{我}) * c(\text{什})) + \exp(c(\text{我}) * c(\text{么})) + \exp(c(\text{我}) * c(\text{呢}))$ 。其中， $c(\text{我})$ 为“我”的字向量， $p(\text{我} | \text{干})$ 为从第一神经网络输入“干”来预测“我”出现的概率。

[0090] 依此类推进行计算训练语句中的每个字在历史子串中出现的概率。

[0091] 需要说明的是，c的值越大，第一神经网络所刻画的预测信息就越复杂，得到的字向量性能越好，代价是训练时间越长。

[0092] 步骤203：将所述字向量的上文特征向量和字向量下文的特征向量作为第二神经网络的参数输入到所述第二神经网络中，训练所述第二神经网络。

[0093] 如图4所示为本申请中的第二神经网络的示意图。其中，第二神经网络包括输入层、映射层、隐藏层和输出层。

[0094] 所述第二神经网络的输入层接收第一神经网络的输出层输出的字向量的上文特征向量和字向量下文的特征向量。其中，将每个字向量的上下文特征向量 e_j 表示为一个维度为 $|V|$ 的向量，其中 V 为所有汉字集合， $|V|$ 表示这个集合的大小，即汉字总数。

[0095] 所述第二神经网络的所述输入层将接收的字向量的上文特征向量和字向量下文的特征向量 e_j 对应的 $|V|$ 维向量，进行一个线性变换投影到一个维度为 m 的向量 c_j ， c_j 称为 e_j 的字向量，将投影后的字向量的上文特征向量和字向量下文的特征向量首尾相接，组成一个 $(n-1) \times m$ 的向量 b ，将拼接后的结果（即图4中的映射层）作为所述第二神经网络的映射层的输入。

[0096] 将映射层的输入进行线性变换矩阵 M 和第一非线性激活函数处理后得到所述第二神经网络的隐藏层的输入。其中，第二神经网络的隐藏层的第 i 个节点记为 d_i ，则：

$$[0097] \quad d_i = \tanh(\sum_j b_j M_{i,j})$$

[0098] 其中， $M_{i,j}$ 为线性变换矩阵 M 的第 (i, j) 个元素， $\tanh()$ 为第一非线性激活函数。

[0099] 将所述第二神经网络的隐藏层的输入进行线性变换矩阵 R 和第二非线性激活函数处理后得到所述第二神经网络的输出层的输出结果，所述输出结果包括字向量在字的历史子串中的概率。其中，第二神经网络的输出层由 $|V|$ 个节点组成，其中第 i 个节点的输出值记为 o_i （即图4中 $P(e_{j=1}|h_j) \cdots P(e_{j=N}|h_j)$ ），则：

$$[0100] \quad o_i = \text{softmax}(\sum_j d_j R_{i,j})$$

[0101] 其中 $R_{i,j}$ 为线性变换矩阵 R 的第 (i, j) 个元素， $\text{softmax}()$ 为第二非线性激活函数。

[0102] 上述过程中，第一神经网络和第二神经网络的结合如图5所示，可见，将获取的每个字的字向量通过第一神经网络预训练，获得预训练的结果，将预训练的结果通过第二神经网络的处理，获得了每个字向量在其历史信息（字向量的上文）和未来信息（字向量的下文）中出现了概率，由于通过第一神经网络和第二神经网络对训练语句的双重处理有效降低了汉字多义出现的概率，从而实现了对于训练语句的精准预测。

[0103] 基于上述方法实施例的说明，本申请还提供了相应的一种基于字的神经网络的训练装置的实施例，来实现上述方法实施例所述的内容。

[0104] 通过本实施例，首先，本申请将每个训练语句按字获取每个字的字向量，将获取的字向量通过第一神经网络进行预训练，获得预训练结果，其中，预训练结果包括字向量的上文特征向量和字向量下文的特征向量。将所述字向量的上文特征向量和字向量下文的特征向量作为第二神经网络的参数输入，训练第二神经网络。由于汉语中汉字的数量是固定的，常用的汉字数量大约在6千左右，远远小于词语的数量，因此可以降低神经网络的复杂度，提高了语言模型的训练效率。同时，对于不断涌现的新词，可以通过将新词按字拆分成字串，因此无需对出现的新词进行重新训练语言模型，从而提高了语言模型的训练效率。

[0105] 本申请将获取的每个字向量通过第一神经网络进行预训练，获得了每个字向量在

上下文中的特征向量,然后将获得的特征向量在通过第二神经网络进行训练,由于获取的字向量通过第一神经网络和第二神经网络的双重处理,有效降低了汉字多义出现的概率,从而解决了汉字本身具有多义性的问题。

[0106] 参见图6,示出了本申请实施例三中的一种基于字的神经网络的训练装置的结构框图,具体可以包括:获取模块301,用于将每个训练语句按字获取每个字的字向量。

[0107] 预训练模块302,用于将所述字向量作为第一神经网络的参数输入到所述第一神经网络中进行预训练,获得预训练结果,其中,所述结果包括字向量的上文特征向量和字向量下文的特征向量。

[0108] 训练模块303,用于将所述字向量的上文特征向量和字向量下文的特征向量作为第二神经网络的参数输入到所述第二神经网络中,训练所述第二神经网络。

[0109] 优选地,所述第一神经网络包括输入层、隐藏层和输出层。

[0110] 优选地,所述预训练模块包括:分配模块,用于为训练语句中每个字随机分配字向量。

[0111] 投影模块,用于输入层接收训练语句中每个字随机分配字向量,所述输入层将接收的每个字的字向量投影到映射层,得到投影后的每个字向量。

[0112] 计算模块,用于计算投影后的每个字向量在字的历史子串中出现的概率,将所述概率作为输出层的预训练结果,所述结果包括字向量的上文特征向量和字向量下文的特征向量。

[0113] 优选地,所述计算模块通过以下公式计算投影后的每个字向量在字的历史子串中出现的概率:

$$[0114] \quad \frac{1}{T} \sum_{t=1}^T \sum_{-c \leq j \leq c, j \neq 0} \log p(e_{t+j} | e_t)$$

[0115] 其中,T为训练语句的长度,c为上下文窗长, $p(e_{t+j} | e_t)$ 为字向量在字的历史子串中出现的概率, $p(e_{t+j} | e_t)$ 通过以下公式获得:

$$[0116] \quad p(e_{t+j} | e_t) = \frac{\exp(c_{e_{t+j}} c_{e_t})}{\sum_e \exp(c_e c_{e_t})}$$

[0117] 其中, c_e 为训练语句的字向量。

[0118] 优选地,所述第二神经网络包括输入层、映射层、隐藏层和输出层;

[0119] 所述第二神经网络的输入层接收第一神经网络的输出层输出的字向量的上文特征向量和字向量下文的特征向量。

[0120] 所述第二神经网络的所述输入层将接收的字向量的上文特征向量和字向量下文的特征向量进行线性变换投影,将投影后的字向量的上文特征向量和字向量下文的特征向量首尾相接,将拼接后的结果作为所述第二神经网络的映射层的输入。

[0121] 将映射层的输入进行线性变换和第一非线性激活函数处理后得到所述第二神经网络的隐藏层的输入。

[0122] 将所述第二神经网络的隐藏层的输入进行线性变换和第二非线性激活函数处理后得到所述第二神经网络的输出层的输出结果,所述输出结果包括字向量在字的历史子串

中的概率。

[0123] 综上所述,本申请实施例一种字的神经网络的训练装置主要包括以下优点:

[0124] 首先,本申请将每个训练语句按字获取每个字的字向量,将获取的字向量通过第一神经网络进行预训练,获得预训练结果,其中,预训练结果包括字向量的上文特征向量和字向量下文的特征向量。将所述字向量的上文特征向量和字向量下文的特征向量作为第二神经网络的参数输入,训练第二神经网络。由于汉语中汉字的数量是固定的,常用的汉字数量大约在6千左右,远远小于词语的数量,因此可以降低神经网络的复杂度,提高了语言模型的训练效率。同时,对于不断涌现的新词,可以通过将新词按字拆分成字串,因此无需对出现的新词进行重新训练语言模型,从而提高了语言模型的训练效率。

[0125] 本申请将获取的每个字向量通过第一神经网络进行预训练,获得了每个字向量在上下文中的特征向量,然后将获得的特征向量在通过第二神经网络进行训练,由于获取的字向量通过第一神经网络和第二神经网络的双重处理,有效降低了汉字多义出现的概率,从而解决了汉字本身具有多义性的问题。

[0126] 对于装置实施例而言,由于其与方法实施例基本相似,所以描述的比较简单,相关之处参见方法实施例的部分说明即可。

[0127] 本说明书中的各个实施例均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的部分互相参见即可。

[0128] 以上对本申请所提供的一种基于字的神经网络的训练方法和装置,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。

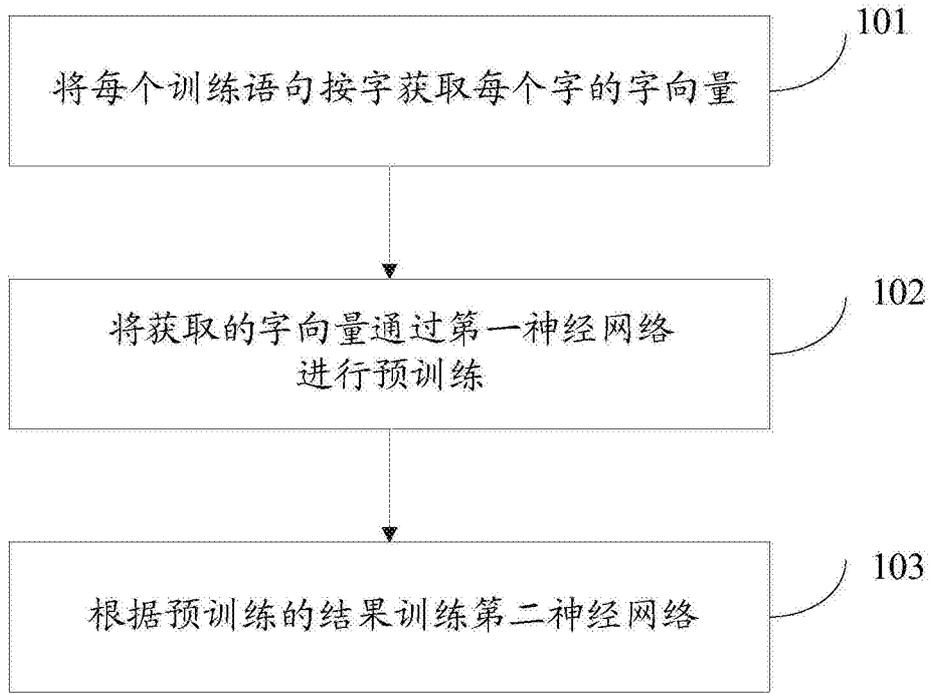


图1

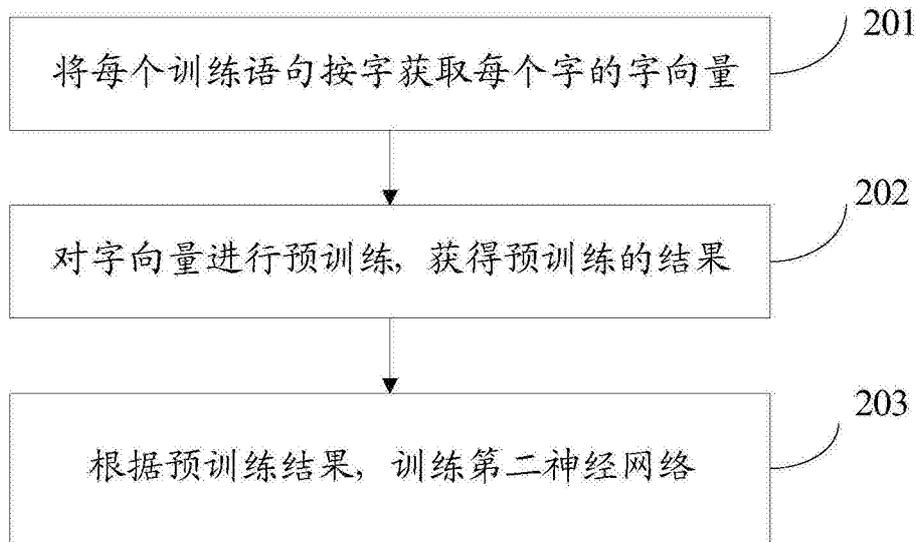


图2

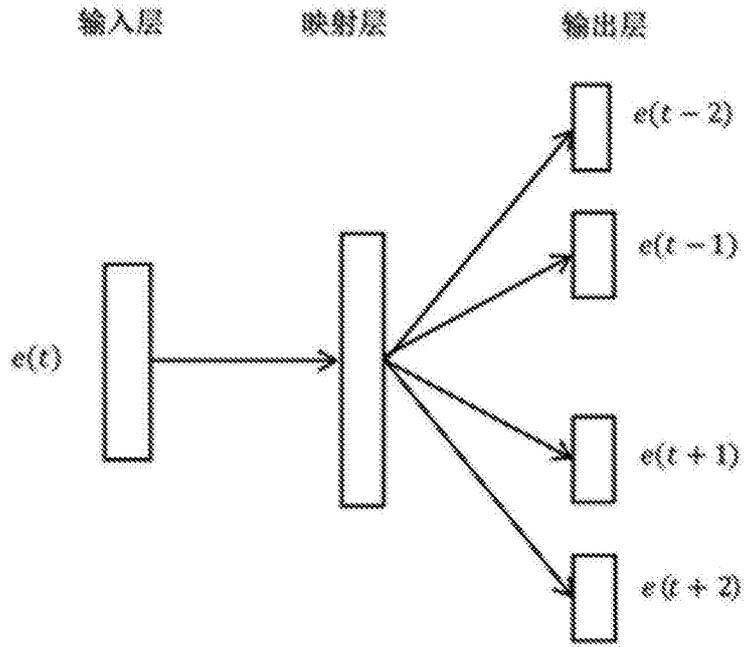


图3

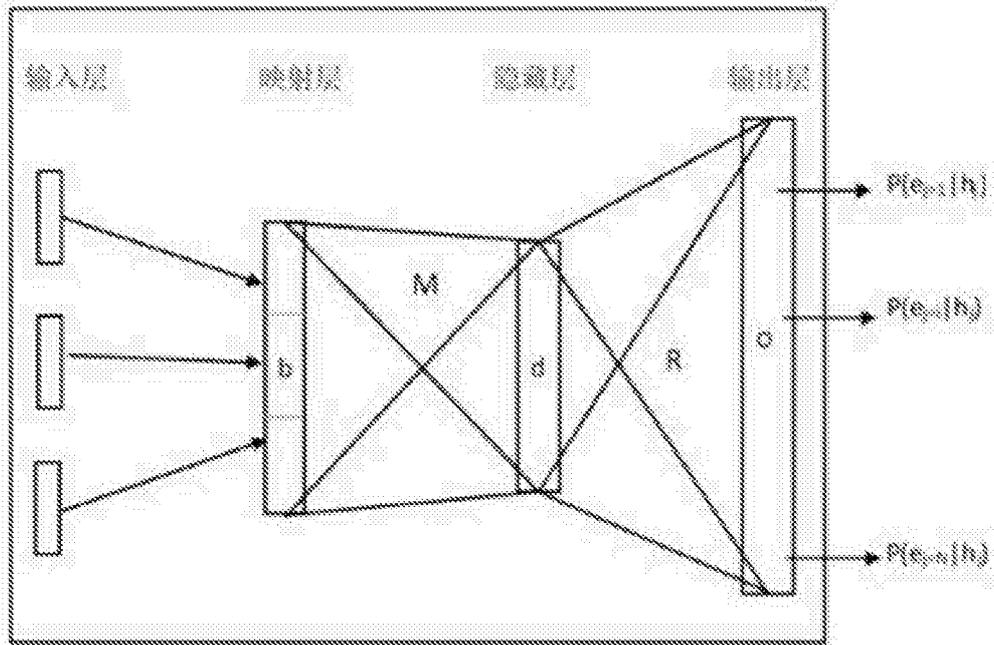


图4

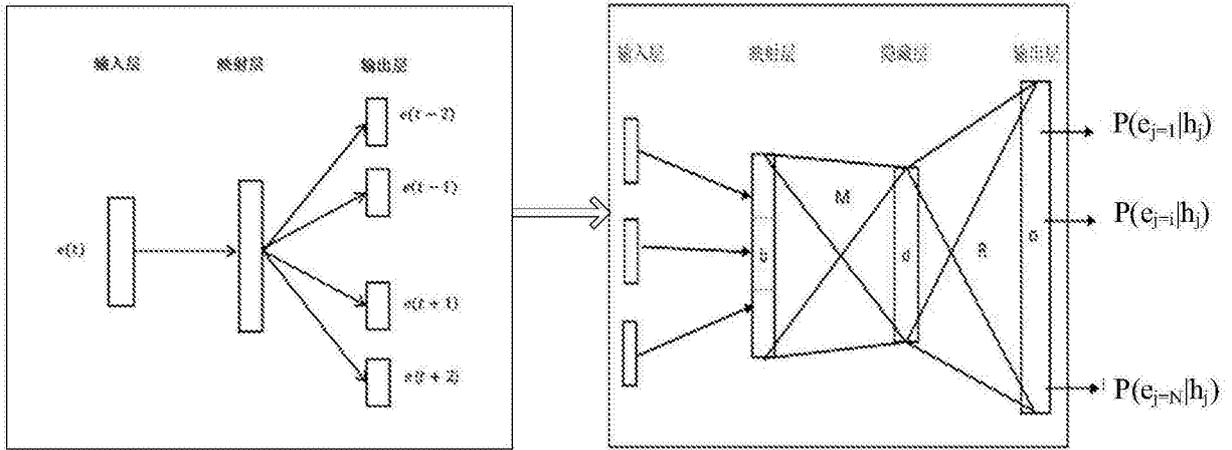


图5

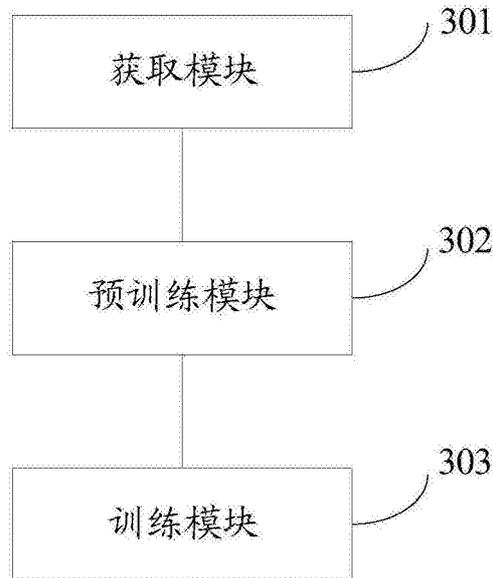


图6