(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 115329888 A
(43)申请公布日 2022.11.11

(21)申请号 202211026746.0

(22)申请日 2022.08.25

(71)申请人 京东科技信息技术有限公司
地址 100176 北京市北京经济技术开发区科创十一街18号院2号楼6层601

(72)发明人 张振羽 陈蒙

(74)专利代理机构 中原信达知识产权代理有限责任公司 11219
专利代理师 李阳 王安娜

(51)Int.Cl.
G06K 9/62(2022.01)
G06N 3/04(2006.01)
G06F 40/289(2020.01)
G06F 40/30(2020.01)

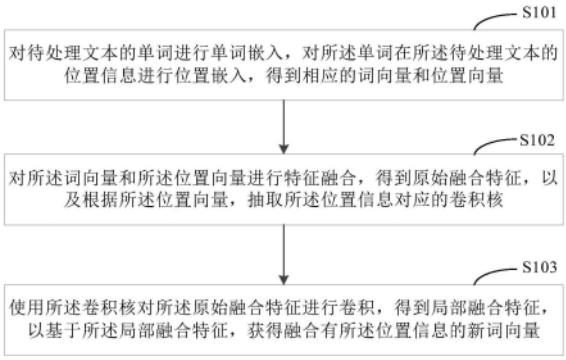
权利要求书2页 说明书10页 附图5页

(54)发明名称

一种信息融合方法和装置

(57)摘要

本发明公开了一种信息融合方法和装置，涉及计算机技术领域。该方法的一具体实施方式包括：对待处理文本的单词进行单词嵌入，对单词在待处理文本的位置信息进行位置嵌入，得到相应的词向量和位置向量；对词向量和位置向量进行特征融合，得到原始融合特征，以及根据位置向量，抽取位置信息对应的卷积核；使用卷积核对原始融合特征进行卷积，得到局部融合特征，以基于局部融合特征，获得融合有位置信息的新词向量。该实施方式通过提取局部信息，实现局部感知以及卷积核共享，进而利用局部信息获取融合有位置信息的新词向量，能够高效、充分的融合位置信息和单词信息。



CN 115329888 A

1.一种信息融合方法,其特征在于,包括:

对待处理文本的单词进行单词嵌入,对所述单词在所述待处理文本的位置信息进行位置嵌入,得到相应的词向量和位置向量;

对所述词向量和所述位置向量进行特征融合,得到原始融合特征,以及根据所述位置向量,抽取所述位置信息对应的卷积核;

使用所述卷积核对所述原始融合特征进行卷积,得到局部融合特征,以基于所述局部融合特征,获得融合有所述位置信息的新词向量。

2.根据权利要求1所述的方法,其特征在于,所述根据所述位置向量,抽取所述位置信息对应的卷积核,包括:

对所述位置向量进行线性变换,得到第一变换结果;

采用第一激活函数,对所述第一变换结果进行非线性变换,得到第二变换结果;

采用第二激活函数,将所述第二变换结果映射到设定区间,获得所述位置向量的相应位置信息所对应的卷积核。

3.根据权利要求1所述的方法,其特征在于,所述使用所述卷积核对所述原始融合特征进行卷积,得到局部融合特征,包括:

采用深度可分离卷积,对一个或者多个输入通道对应的原始融合特征分别进行卷积,得到所述输入通道对应的局部融合特征;其中,所述多个输入通道共享所述卷积核。

4.根据权利要求3所述的方法,其特征在于,所述基于所述局部融合特征,获得融合有所述位置信息的新词向量,包括:

对所述待处理文本的单词对应的局部融合特征集进行线性变换,得到融合有所述位置信息的新词向量;其中,所述局部融合特征集包括所述一个或者多个输入通道对应的局部融合特征。

5.根据权利要求1所述的方法,其特征在于,所述对所述词向量和所述位置向量进行特征融合,得到原始融合特征,包括:

拼接所述词向量和所述位置向量,得到拼接向量;

采用Pointwise卷积,对所述拼接向量进行浅层特征融合,得到原始融合特征。

6.根据权利要求2所述的方法,其特征在于,所述第一激活函数为不饱和非线性激活函数,所述第二激活函数为softmax函数。

7.根据权利要求1至6的任一项所述的方法,其特征在于,所述方法还包括:

将所述新词向量输入至预训练语言模型,通过所述预训练语言模型中的Transformer结构,得到所述待处理文本中单词的语义表示。

8.一种信息融合装置,其特征在于,包括:

嵌入模块,用于对待处理文本的单词进行单词嵌入,对所述单词在所述待处理文本的位置信息进行位置嵌入,得到相应的词向量和位置向量;

抽取模块,用于对所述词向量和所述位置向量进行特征融合,得到原始融合特征,以及根据所述位置向量,抽取所述位置信息对应的卷积核;

融合模块,用于使用所述卷积核对所述原始融合特征进行卷积,得到局部融合特征,以基于所述局部融合特征,获得融合有所述位置信息的新词向量。

9.一种电子设备,其特征在于,包括:

一个或多个处理器；

存储装置，用于存储一个或多个程序，

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如权利要求1-7中任一所述的方法。

10.一种计算机可读介质，其上存储有计算机程序，其特征在于，所述程序被处理器执行时实现如权利要求1-7中任一所述的方法。

一种信息融合方法和装置

技术领域

[0001] 本发明涉及计算机技术领域，尤其涉及一种信息融合方法和装置。

背景技术

[0002] Transformer是谷歌团队提出的一种基于自注意力(Self-Attention)的神经网络，是解决自然语言处理(Natural Language Process，NLP)问题的核心模块。目前比较成熟的预训练语言模型，例如BERT、Roberta等，都是基于Transformer设计的。

[0003] 基于Transformer对文本序列建模时，单词在文本中出现的位置信息会参与自注意力计算，但是Transformer不具备位置信息的建模能力，导致在自注意力计算中位置信息和单词信息无法充分融合。为了解决该问题，现有技术主要通过绝对位置编码、相对位置编码和分离位置编码三种方式进行文本序列建模。

[0004] 在实现本发明过程中，现有技术中至少存在如下问题：

[0005] 在自注意力计算中，上述三种方式对位置信息和单词信息的融合不充分，且融合效率低。

发明内容

[0006] 有鉴于此，本发明实施例提供一种信息融合方法和装置，该方法先利用待处理文本对应的位置向量抽取位置相关的卷积核，之后使用该卷积核对原始融合特征进行卷积，提取局部信息，实现局部感知以及卷积核共享，进而利用局部信息获取融合有位置信息的新词向量，能够高效、充分的融合位置信息和单词信息。

[0007] 为实现上述目的，根据本发明实施例的一个方面，提供了一种信息融合方法。

[0008] 本发明实施例的一种信息融合方法，包括：对待处理文本的单词进行单词嵌入，对所述单词在所述待处理文本的位置信息进行位置嵌入，得到相应的词向量和位置向量；对所述词向量和所述位置向量进行特征融合，得到原始融合特征，以及根据所述位置向量，抽取所述位置信息对应的卷积核；使用所述卷积核对所述原始融合特征进行卷积，得到局部融合特征，以基于所述局部融合特征，获得融合有所述位置信息的新词向量。

[0009] 可选地，所述根据所述位置向量，抽取所述位置信息对应的卷积核，包括：对所述位置向量进行线性变换，得到第一变换结果；采用第一激活函数，对所述第一变换结果进行非线性变换，得到第二变换结果；采用第二激活函数，将所述第二变换结果映射到设定区间，获得所述位置向量的相应位置信息所对应的卷积核。

[0010] 可选地，所述使用所述卷积核对所述原始融合特征进行卷积，得到局部融合特征，包括：采用深度可分离卷积，对一个或者多个输入通道对应的原始融合特征分别进行卷积，得到所述输入通道对应的局部融合特征；其中，所述多个输入通道共享所述卷积核。

[0011] 可选地，所述基于所述局部融合特征，获得融合有所述位置信息的新词向量，包括：对所述待处理文本的单词对应的局部融合特征集进行线性变换，得到融合有所述位置信息的新词向量；其中，所述局部融合特征集包括所述一个或者多个输入通道对应的局部

融合特征。

[0012] 可选地，所述对所述词向量和所述位置向量进行特征融合，得到原始融合特征，包括：拼接所述词向量和所述位置向量，得到拼接向量；采用Pointwise卷积，对所述拼接向量进行浅层特征融合，得到原始融合特征。

[0013] 可选地，所述第一激活函数为不饱和非线性激活函数，所述第二激活函数为softmax函数。

[0014] 可选地，所述方法还包括：将所述新词向量输入至预训练语言模型，通过所述预训练语言模型中的Transformer结构，得到所述待处理文本中单词的语义表示。

[0015] 为实现上述目的，根据本发明实施例的另一方面，提供了一种信息融合装置。

[0016] 本发明实施例的一种信息融合装置，包括：嵌入模块，用于对待处理文本的单词进行单词嵌入，对所述单词在所述待处理文本的位置信息进行位置嵌入，得到相应的词向量和位置向量；抽取模块，用于对所述词向量和所述位置向量进行特征融合，得到原始融合特征，以及根据所述位置向量，抽取所述位置信息对应的卷积核；融合模块，用于使用所述卷积核对所述原始融合特征进行卷积，得到局部融合特征，以基于所述局部融合特征，获得融合有所述位置信息的新词向量。

[0017] 可选地，所述抽取模块，还用于对所述位置向量进行线性变换，得到第一变换结果；采用第一激活函数，对所述第一变换结果进行非线性变换，得到第二变换结果；采用第二激活函数，将所述第二变换结果映射到设定区间，获得所述位置向量的相应位置信息所对应的卷积核。

[0018] 可选地，所述融合模块，还用于采用深度可分离卷积，对一个或者多个输入通道对应的原始融合特征分别进行卷积，得到所述输入通道对应的局部融合特征；其中，所述多个输入通道共享所述卷积核。

[0019] 可选地，所述融合模块，还用于对所述待处理文本的单词对应的局部融合特征集进行线性变换，得到融合有所述位置信息的新词向量；其中，所述局部融合特征集包括所述一个或者多个输入通道对应的局部融合特征。

[0020] 可选地，所述抽取模块，还用于拼接所述词向量和所述位置向量，得到拼接向量；采用Pointwise卷积，对所述拼接向量进行浅层特征融合，得到原始融合特征。

[0021] 可选地，所述第一激活函数为不饱和非线性激活函数，所述第二激活函数为softmax函数。

[0022] 可选地，所述装置还包括：输入模块，用于将所述新词向量输入至预训练语言模型，通过所述预训练语言模型中的Transformer结构，得到所述待处理文本中单词的语义表示。

[0023] 为实现上述目的，根据本发明实施例的再一方面，提供了一种电子设备。

[0024] 本发明实施例的一种电子设备，包括：一个或多个处理器；存储装置，用于存储一个或多个程序，当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现本发明实施例的一种信息融合方法。

[0025] 为实现上述目的，根据本发明实施例的再一方面，提供了一种计算机可读介质。

[0026] 本发明实施例的一种计算机可读介质，其上存储有计算机程序，所述程序被处理器执行时实现本发明实施例的一种信息融合方法。

[0027] 上述发明中的一个实施例具有如下优点或有益效果:先利用待处理文本对应的位置向量抽取位置相关的卷积核,之后使用该卷积核对原始融合特征进行卷积,提取局部信息,实现局部感知以及卷积核共享,进而利用局部信息获取融合有位置信息的新词向量,能够高效、充分的融合位置信息和单词信息。

[0028] 对于每一个位置向量,先采用非线性变换,再对变换结果进行区间映射,即可利用位置向量得到位置相关的卷积核,便于后续提取局部信息,降低计算量。采用深度可分离卷积对每个输入通道进行卷积,保证能够充分融合词向量和位置向量。通过对局部融合特征集进行线性变换,进一步保证能够得到充分融合的具有位置信息的新词向量。

[0029] 通过Pointwise卷积,对词向量和位置向量的拼接向量进行浅层特征融合,在保证融合效果的前提下,进一步减少计算量。通过不饱和非线性激活函数、softmax函数把输入数据映射到输出端,便于后续处理。将获得的新词向量输入至预训练语言模型,用于其内部的自注意力计算,实现了结合位置信息的文本序列建模。

[0030] 上述的非惯用的可选方式所具有的进一步效果将在下文中结合具体实施方式加以说明。

附图说明

[0031] 附图用于更好地理解本发明,不构成对本发明的不当限定。其中:

[0032] 图1是根据本发明实施例的信息融合方法的主要步骤的示意图;

[0033] 图2是根据本发明实施例的信息融合方法的实现原理示意图;

[0034] 图3是根据本发明实施例的信息融合方法的主要流程示意图;

[0035] 图4是根据本发明实施例的信息融合装置的主要模块的示意图;

[0036] 图5是本发明实施例可以应用于其中的示例性系统架构图;

[0037] 图6是适用于来实现本发明实施例的电子设备的计算机装置的结构示意图。

具体实施方式

[0038] 以下结合附图对本发明的示范性实施例做出说明,其中包括本发明实施例的各种细节以助于理解,应当将它们认为仅仅是示范性的。因此,本领域普通技术人员应当认识到,可以对这里描述的实施例做出各种改变和修改,而不会背离本发明的范围和精神。同样,为了清楚和简明,以下的描述中省略了对公知功能和结构的描述。

[0039] 结合背景技术,Transformer的核心结构是自注意力,其计算过程涉及查询(Query,Q),键(Key,K)和值(Value,V)三项元素。输入的表征向量首先会通过一个矩阵变换获得对应的Q,K和V。目前基于Transformer的语言模型在对文本序列建模时,主要通过绝对位置编码、相对位置编码或者分离位置编码的方式实现。

[0040] 但是,绝对位置编码方式中,位置向量和词向量共用计算参数,导致自注意力计算得到的注意力分数无法高效融合位置信息和单词信息。相对位置编码方式中,在自注意力计算时,只是简单的采用了相对位置信息来增强键值的表征,依旧难以高效获得位置向量和词向量之间的深度融合信息。分离位置编码方式中,虽然不共享计算参数,但计算参数完全随机化,与输入信息无关,不利于建模位置信息和单词信息的关联关系。

[0041] 为了解决上述三种方式中位置信息和单词信息融合不充分、融合效率低的缺陷,

本发明实施例提供了一种新的信息融合方法，该方法能够高效、充分的融合位置信息和单词信息，具体实现过程如下。

[0042] 图1是根据本发明实施例的信息融合方法的主要步骤的示意图。

[0043] 如图1所示，本发明实施例的信息融合方法，主要包括如下步骤：

[0044] 步骤S101：对待处理文本的单词进行单词嵌入，对所述单词在所述待处理文本的位置信息进行位置嵌入，得到相应的词向量和位置向量。单词嵌入(Word Embedding)是将单词映射到高维向量，比如可以用Word2Vec(是一个将单词转换成向量形式的工具)、FastText(是一个词向量计算和文本分类工具)等方式实现。位置嵌入(Position Embedding)是将位置信息映射到高维向量。

[0045] 实施例中，先对待处理文本进行分词处理，得到单词序列，再对单词序列的每个单词进行单词嵌入，得到各单词对应的词向量，以及对每个单词在待处理文本的位置信息进行位置嵌入，得到各位置信息对应的位置向量。

[0046] 步骤S102：对所述词向量和所述位置向量进行特征融合，得到原始融合特征，以及根据所述位置向量，抽取所述位置信息对应的卷积核。本步骤中，采用卷积方式，对词向量和位置向量进行特征融合，得到原始融合特征。并且对于每个位置向量，采用非线性变换和区间映射的方式，获得位置相关的卷积核。

[0047] 步骤S103：使用所述卷积核对所述原始融合特征进行卷积，得到局部融合特征，以基于所述局部融合特征，获得融合有所述位置信息的新词向量。使用步骤S102获得的卷积核，对至少一个输入通道进行卷积，得到融合有局部单词信息和位置信息的局部融合特征；后续对该局部融合特征进行线性变换，即可获得具有位置信息的新词向量。

[0048] 本实施例实现了局部感知以及卷积核共享，能够高效、充分的融合位置信息和单词信息。

[0049] 图2是根据本发明实施例的信息融合方法的实现原理示意图。如图2所示，本实施例的待处理文本包含n个单词，首先对这n个单词分别进行单词嵌入，得到相应的词向量；以及对n个单词在待处理文本的位置分别进行位置嵌入，得到相应的位置向量；之后基于位置向量，抽取位置相关的卷积核；最后使用该卷积核，对词向量和位置向量对应的原始融合特征进行深度可分离卷积，以获得n个单词的、具有位置信息的新词向量。

[0050] 为了便于说明，可将n个单词表示为$X=\{x_1,x_2,\cdots\cdots,x_n\}$；n个单词对应的词向量可表示为$W=\{w_1,w_2,\cdots\cdots,w_n\}$；n个单词对应的位置向量可表示为$P=\{p_1,p_2,\cdots\cdots,p_n\}$；n个单词对应的新词向量可表示为$W'=\{w'_1,w'_2,\cdots\cdots,w'_n]$。

[0051] 图3是根据本发明实施例的信息融合方法的主要流程示意图。如图3所示，本发明实施例的信息融合方法，主要包括如下步骤：

[0052] 步骤S301：对待处理文本的单词进行单词嵌入，得到各单词对应的词向量。对待处理文本进行分词处理，得到单词序列，再对单词序列的每个单词进行单词嵌入，得到各单词对应的词向量。假设待处理文本为：这个手机多少钱，则分词所得的单词序列可以为：“这”“个”“手机”“多少”“钱”。

[0053] 假设待处理文本包含n个单词，将其表示为$X=[x_1,x_2,\cdots\cdots,x_n\}$，其中，$X\in R^n$，即X为n维向量，n为整数，向量的每个维度数值为实数。对n个单词分别进行单词嵌入，获得对应的词向量$W=\{w_1,w_2,\cdots\cdots,w_n\}$，其中，$w_i\in R^d$，$1\leqslant i\leqslant n$，d为嵌入维度。

[0054] 步骤S302:对单词在待处理文本的位置信息进行位置嵌入,得到各位置信息对应的位置向量。对n个单词在待处理文本出现的n个位置的位置信息进行位置编码,获得对应的位置向量$P=\{p_1,p_2,\cdots\cdots,p_n\}$,其中,$p_i\in R^d$,d为嵌入维度,其取值与步骤S301中进行单词嵌入的嵌入维度相同。

[0055] 步骤S303:将对应相同单词的词向量和位置向量进行特征融合,得到原始融合特征。将对应相同单词的词向量和位置向量进行拼接,得到各单词对应的拼接向量,之后采用Pointwise卷积,对拼接向量进行浅层特征融合,得到各单词对应的原始融合特征。其中,浅层特征为深度学习网络的嵌入层,或者第一、二层的网络层的输出向量表示。

[0056] Pointwise(PW)卷积的卷积核大小为1×1×M,M为上一层的通道数。该卷积运算会将上一步的特征图在深度方向上进行加权组合,生成新的特征图。有几个卷积核就有几个输出特征图。该卷积具有可学习参数W^{bottle}。

[0057] 其中,n个单词的其中一个单词对应的原始融合特征的表达式如下:

[0058] $$h_i=\sum_{c=1}^{2d} W_c^{bottle}\, g[w_i;p_i]$$

公式1

[0059] 式中,$h_i\in R^{2d}$,为待处理文本的第i个单词的原始融合特征;2d为输入通道的数量,c为当前输入通道;$W_c^{bottle}\in R^{2d}$,为Pointwise卷积的可学习参数;$g[w_i;p_i]$为待处理文本的第i个单词对应的拼接向量。

[0060] 基于公式1,n个单词对应的原始融合特征可以表示为$H=\{h_1,h_2,\cdots\cdots,h_n\}\in R^{n\times 2d}$。

[0061] 步骤S304:根据各位置信息的位置向量,抽取各位置信息对应的卷积核。本步骤中,为了抽取位置相关(position-specific)的卷积核,先对各位置信息的位置向量进行线性变换,得到第一变换结果;之后采用第一激活函数,对第一变换结果进行非线性变换,得到第二变换结果;最后采用第二激活函数,将第二变换结果映射到设定区间,获得各位置信息所对应的卷积核。

[0062] 实施例中,第一激活函数为不饱和非线性激活函数,比如Relu函数、GELU函数;第二激活函数为softmax函数。在将第二变换结果映射到设定区间时,需沿着嵌入维度d对第二变换结果进行softmax,即可将第二变换结果映射到(0,1)区间。此方式获得的卷积核在每个位置都不同,与位置信息紧密相关。

[0063] 其中,第二变换结果的表达式如下:

[0064] $$Kernel_i^K=W_{p2}Relu\left(W_{p1}p_i+b_{p1}\right)+b_{p2}$$

公式2

[0065] 式中,$Kernel_i^K$为待处理文本的第i个单词对应的第二变换结果;W_{p1}、b_{p1}、W_{p2}和b_{p2}均为可学习参数,用于对特征表示结果进行高维空间的投影变换等操作。其中,$W_{p1}\in R^{d_1\times d}$,$W_{p2}\in R^{k\times d_1}$,$b_{p1}\in R^{d_1}$,$b_{p2}\in R^k$;$d_1=\gamma\times d$;$\gamma\in R$,$\gamma$大于1,优选为4。

[0066] 卷积核的表达式如下:

[0067] $$W_i^K=softmax\left(Kernel_i^K\right)$$

公式 3

[0068] 式中，$W_i^K \in R^k$，为待处理文本的第i个单词对应的卷积核，且卷积核大小为k。

[0069] 步骤S305：使用对应相同单词的卷积核对原始融合特征进行卷积，得到各单词对应的局部融合特征集。基于上述获取的卷积核，采用深度可分离卷积，沿着单词序列的维度n，对一个或者多个输入通道对应的原始融合特征分别进行卷积，得到输入通道对应的局部融合特征。多个输入通道共享卷积核。

[0070] 深度可分离卷积(Depthwise Separable Convolution)，由Depthwise(DW)卷积和Pointwise(PW)卷积两部分结合起来，可用来提取特征，相比常规的卷积操作，其参数数量和运算成本较低。

[0071] 实施例中，对输入通道1～2d共享卷积核$W_{i,j}^K$，$1 \le j \le k$。最终对于每个输入通道，可以获得一个融合局部单词信息和位置信息的局部融合特征$0_{i,c}$。其中，$0_{i,c}$可用如下公式表示：

[0072] $$O_{i,c}=\sum_{j=1}^{j=k} W_{i,j}^K \cdot H_{i+j-\left\lfloor\frac{k+1}{2}\right\rfloor,c}$$

公式 4

[0073] 式中，$0_{i,c}$为待处理文本的第i个单词、第c个输入通道的局部融合特征；k为卷积核大小；$W_{i,j}^K$为待处理文本的第i个单词、第j个位置的卷积核(也即长度为k的卷积核W_i^K的第j个位置的取值)；$H_{i+j-\left\lfloor\frac{k+1}{2}\right\rfloor,c}$为待处理文本的第i个单词、第c个输入通道的原始融合特征；$1 \leqslant c \leqslant 2d$。

[0074] 公式4表示，对于待处理文本的第i个单词，2d个特征维度共享卷积核$W_{i,j}^K$，在计算第i个单词对应的输出时，需要在i左右选取大小为k的窗口，即[i-(k+1)/2,i+(k+1)/2]，并使用卷积核$W_{i,j}^K$计算局部感受野输出。

[0075] 基于上述公式，待处理文本的第i个单词的全部输入通道对应的局部融合特征构成局部融合特征集，故待处理文本的第i个单词的对应的局部融合特征集可用0_i表示，其取值为$0_i=\{0_{i,1},0_{i,2},\cdots\cdots,0_{i,2d}\} \in R^{2d}$。待处理文本的n个单词对应的局部融合特征集可用0表示，其取值为$0=\{0_1,0_2,\cdots\cdots,0_n\} \in R^{n\times 2d}$。

[0076] 步骤S306：对各单词对应的局部融合特征集进行线性变换，得到融合有位置信息的新词向量。待处理文本的n个单词的全部输入通道对应的局部融合特征构成局部融合特征集。对该局部融合特征集进行线性变换，即可获得具有位置信息的新词向量。

[0077] 其中，待处理文本的第i个单词对应的新词向量w'_i的表达式如下：

[0078] $$w'_i=W_0 0_i+b_0$$

[0079] 公式5

[0080] 式中，$w'_i \in R^d$，$1 \leqslant i \leqslant n$；$W_0$和$b_0$是线性变换的可学习参数，$W_0 \in R^{d\times 2d}$，$b_0$为偏移量。

[0081] 需要说明的是，本实施例并不限定上述步骤的执行顺序。这些步骤中，步骤S301和步骤S302可以先后执行，也可以并列执行；步骤S303和步骤S304同样可以先后执行，也可以并列执行；且步骤S303需在步骤S301和步骤S302执行结束后再执行，步骤S304需在步骤S302执行结束后再执行。

[0082] 在一优选的实施例中，在获得新词向量之后，还可以将新词向量输入至预训练语言模型，通过预训练语言模型中的Transformer结构，得到待处理文本中单词的语义表示。

[0083] 通过本发明实施例的信息融合方法可以看出，本实施例先利用待处理文本对应的位置向量抽取位置相关的卷积核，之后采用深度可分离卷积，将单词信息和位置信息进行融合编码，提取了局部信息，实现局部感知以及输入通道的卷积核共享，进而利用局部信息获取融合有位置信息的新词向量，在保证融合效果的前提下，减少了计算量，能够高效、充分的融合位置信息和单词信息。

[0084] 图4是根据本发明实施例的信息融合装置的主要模块的示意图。

[0085] 如图4所示，本发明实施例的信息融合装置400，主要包括：

[0086] 嵌入模块401，用于对待处理文本的单词进行单词嵌入，对所述单词在所述待处理文本的位置信息进行位置嵌入，得到相应的词向量和位置向量。对待处理文本进行分词处理，得到单词序列，再对单词序列的每个单词进行单词嵌入，得到各单词对应的词向量，以及对每个单词在待处理文本的位置信息进行位置嵌入，得到各位置信息对应的位置向量。

[0087] 抽取模块402，用于对所述词向量和所述位置向量进行特征融合，得到原始融合特征，以及根据所述位置向量，抽取所述位置信息对应的卷积核。采用卷积方式，对词向量和位置向量进行特征融合，得到原始融合特征。并且对于每个位置向量，采用非线性变换和区间映射的方式，获得位置相关的卷积核。

[0088] 融合模块403，用于使用所述卷积核对所述原始融合特征进行卷积，得到局部融合特征，以基于所述局部融合特征，获得融合有所述位置信息的新词向量。使用抽取模块402获得的卷积核，对至少一个输入通道进行卷积，得到融合有局部单词信息和位置信息的局部融合特征；后续对该局部融合特征进行线性变换，即可获得具有位置信息的新词向量。

[0089] 另外，本发明实施例的信息融合装置400还可以包括：输入模块(图4中未示出)，该模块用于将所述新词向量输入至预训练语言模型，通过所述预训练语言模型中的Transformer结构，得到所述待处理文本中单词的语义表示。

[0090] 从以上描述可以看出，先利用待处理文本对应的位置向量抽取位置相关的卷积核，之后使用该卷积核对原始融合特征进行卷积，提取局部信息，实现局部感知以及卷积核共享，进而利用局部信息获取融合有位置信息的新词向量，能够高效、充分的融合位置信息和单词信息。

[0091] 图5示出了可以应用本发明实施例的信息融合方法或信息融合装置的示例性系统架构500。

[0092] 如图5所示，系统架构500可以包括终端设备501、502、503，网络504和服务器505。网络504用以在终端设备501、502、503和服务器505之间提供通信链路的介质。网络504可以包括各种连接类型，例如有线、无线通信链路或者光纤电缆等等。

[0093] 用户可以使用终端设备501、502、503通过网络504与服务器505交互，以接收或发送消息等。终端设备501、502、503可以是具有显示屏并且支持网页浏览的各种电子设备，包

括但不限于智能手机、平板电脑、膝上型便携计算机和台式计算机等等。

[0094] 服务器505可以是提供各种服务的服务器，例如对管理员利用终端设备501、502、503发送的待处理文本进行处理的后台管理服务器。后台管理服务器可以对待处理文本进行单词嵌入、位置嵌入、特征融合、抽取卷积核、卷积操作等处理，并将处理结果(例如得到的新词向量)反馈给终端设备。

[0095] 需要说明的是，本申请实施例所提供的信息融合方法一般由服务器505执行，相应地，信息融合装置一般设置于服务器505中。

[0096] 应该理解，图5中的终端设备、网络和服务器的数目仅仅是示意性的。根据实现需要，可以具有任意数目的终端设备、网络和服务器。

[0097] 根据本发明的实施例，本发明还提供了一种电子设备和一种计算机可读介质。

[0098] 本发明的电子设备包括：一个或多个处理器；存储装置，用于存储一个或多个程序，当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现本发明实施例的一种信息融合方法。

[0099] 本发明的计算机可读介质，其上存储有计算机程序，所述程序被处理器执行时实现本发明实施例的一种信息融合方法。

[0100] 下面参考图6，其示出了适用于来实现本发明实施例的电子设备的计算机系统600的结构示意图。图6示出的电子设备仅仅是一个示例，不应对本发明实施例的功能和使用范围带来任何限制。

[0101] 如图6所示，计算机系统600包括中央处理单元(CPU)601，其可以根据存储在只读存储器(ROM)602中的程序或者从存储部分608加载到随机访问存储器(RAM)603中的程序而执行各种适当的动作和处理。在RAM 603中，还存储有计算机系统600操作所需的各种程序和数据。CPU 601、ROM 602以及RAM 603通过总线604彼此相连。输入/输出(I/O)接口605也连接至总线604。

[0102] 以下部件连接至I/O接口605：包括键盘、鼠标等的输入部分606；包括诸如阴极射线管(CRT)、液晶显示器(LCD)等以及扬声器等的输出部分607；包括硬盘等的存储部分608；以及包括诸如LAN卡、调制解调器等的网络接口卡的通信部分609。通信部分609经由诸如因特网的网络执行通信处理。驱动器610也根据需要连接至I/O接口605。可拆卸介质611，诸如磁盘、光盘、磁光盘、半导体存储器等等，根据需要安装在驱动器610上，以便于从其上读出的计算机程序根据需要被安装入存储部分608。

[0103] 特别地，根据本发明公开的实施例，上文主要步骤图描述的过程可以被实现为计算机软件程序。例如，本公开的实施例包括一种计算机程序产品，其包括承载在计算机可读介质上的计算机程序，该计算机程序包含用于执行主要步骤图所示的方法的程序代码。在这样的实施例中，该计算机程序可以通过通信部分609从网络上被下载和安装，和/或从可拆卸介质611被安装。在该计算机程序被中央处理单元(CPU)601执行时，执行本发明的系统中限定的上述功能。

[0104] 需要说明的是，本发明所示的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件，或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于：具有一个或多个导线的电连接、便

携式计算机磁盘、硬盘、随机访问存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本发明中,计算机可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本发明中,计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号,其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式,包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质,该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输,包括但不限于:无线、电线、光缆、RF等等,或者上述的任意合适的组合。

[0105] 附图中的流程图和框图,图示了按照本发明各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分,上述模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个接连地表示的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意的是,框图或流程图中的每个方框、以及框图或流程图中的方框的组合,可以用执行规定的功能或操作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0106] 描述于本发明实施例中所涉及到的模块可以通过软件的方式实现,也可以通过硬件的方式来实现。所描述的模块也可以设置在处理器中,例如,可以描述为:一种处理器包括嵌入模块、抽取模块和融合模块。其中,这些模块的名称在某种情况下并不构成对该模块本身的限定,例如,嵌入模块还可以被描述为“对待处理文本的单词进行单词嵌入,对所述单词在所述待处理文本的位置信息进行位置嵌入,得到相应的词向量和位置向量的模块”。

[0107] 作为另一方面,本发明还提供了一种计算机可读介质,该计算机可读介质可以是上述实施例中描述的设备中所包含的;也可以是单独存在,而未装配入该设备中。上述计算机可读介质承载有一个或者多个程序,当上述一个或者多个程序被一个该设备执行时,使得该设备包括:对待处理文本的单词进行单词嵌入,对所述单词在所述待处理文本的位置信息进行位置嵌入,得到相应的词向量和位置向量;对所述词向量和所述位置向量进行特征融合,得到原始融合特征,以及根据所述位置向量,抽取所述位置信息对应的卷积核;使用所述卷积核对所述原始融合特征进行卷积,得到局部融合特征,以基于所述局部融合特征,获得融合有所述位置信息的新词向量。

[0108] 根据本发明实施例的技术方案,先利用待处理文本对应的位置向量抽取位置相关的卷积核,之后使用该卷积核对原始融合特征进行卷积,提取局部信息,实现局部感知以及卷积核共享,进而利用局部信息获取融合有位置信息的新词向量,能够高效、充分的融合位置信息和单词信息。

[0109] 上述产品可执行本发明实施例所提供的方法,具备执行方法相应的功能模块和有益效果。未在本实施例中详尽描述的技术细节,可参见本发明实施例所提供的方法。

[0110] 上述具体实施方式，并不构成对本发明保护范围的限制。本领域技术人员应该明白的是，取决于设计要求和其他因素，可以发生各种各样的修改、组合、子组合和替代。任何在本发明的精神和原则之内所作的修改、等同替换和改进等，均应包含在本发明保护范围之内。

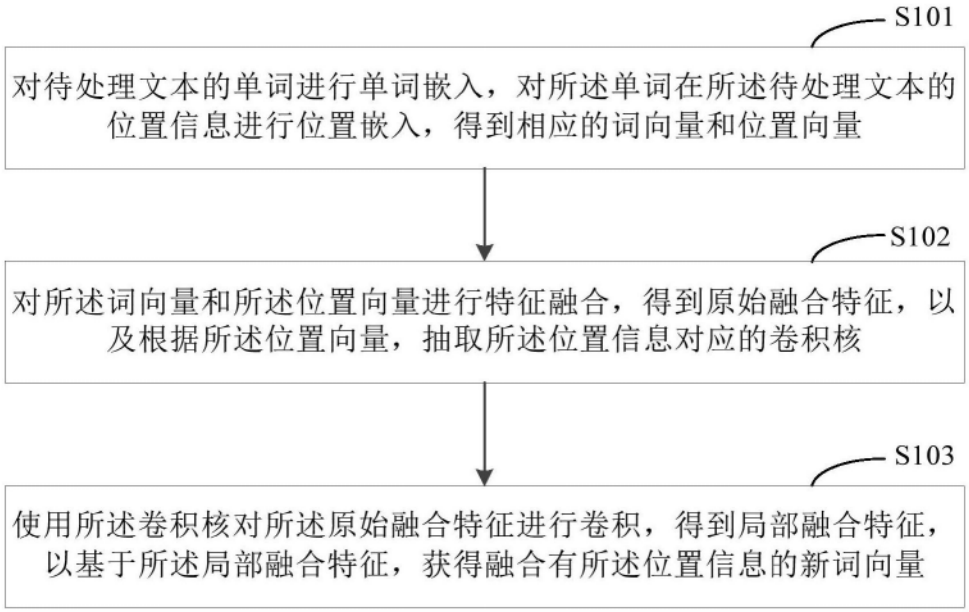

S101
对待处理文本的单词进行单词嵌入，对所述单词在所述待处理文本的位置信息进行位置嵌入，得到相应的词向量和位置向量
S102
对所述词向量和所述位置向量进行特征融合，得到原始融合特征，以及根据所述位置向量，抽取所述位置信息对应的卷积核
S103
使用所述卷积核对所述原始融合特征进行卷积，得到局部融合特征，以基于所述局部融合特征，获得融合有所述位置信息的新词向量

图1

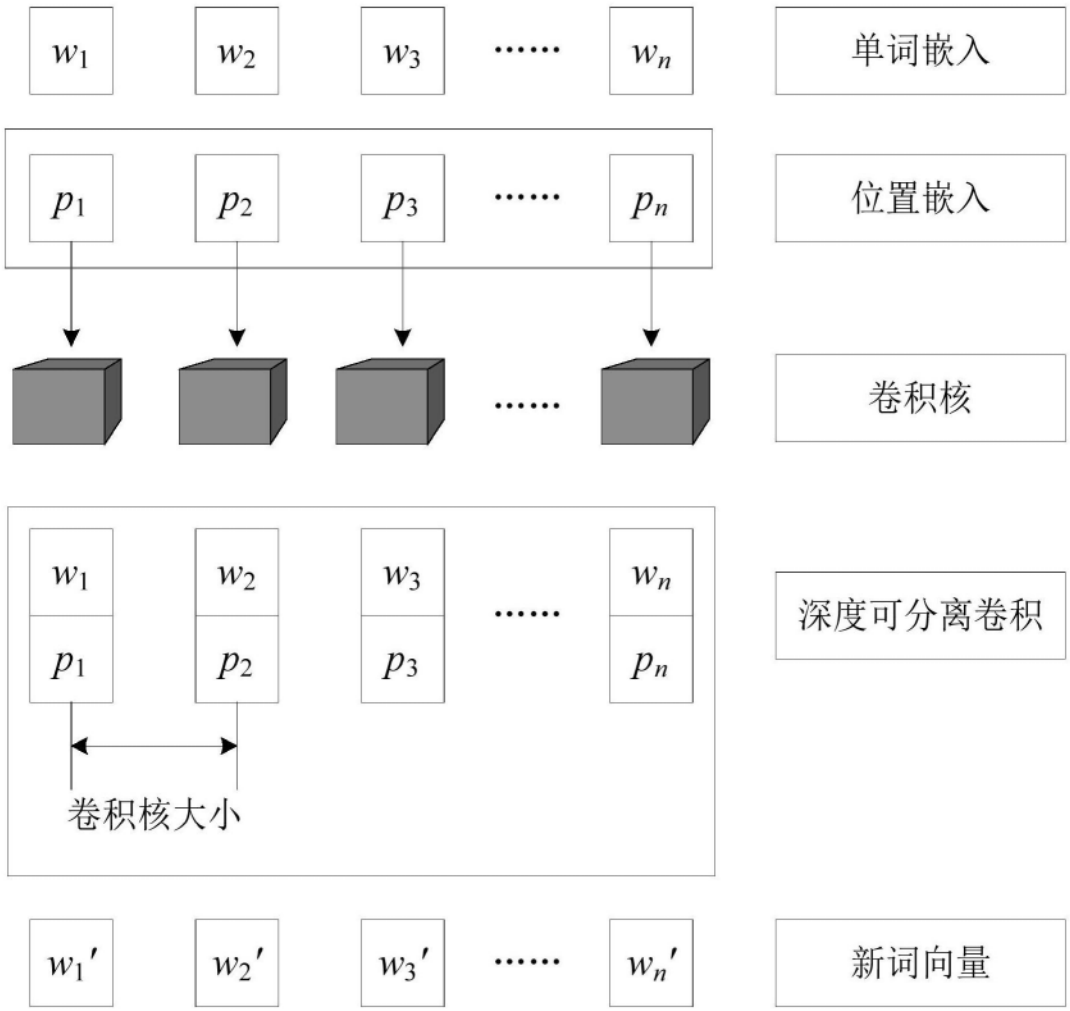

w_1
w_2
w_3
……
w_n
单词嵌入
p_1
p_2
p_3
……
p_n
位置嵌入
……
卷积核
深度可分离卷积
卷积核大小
w_1'
w_2'
w_3'
……
w_n'
新词向量

图2

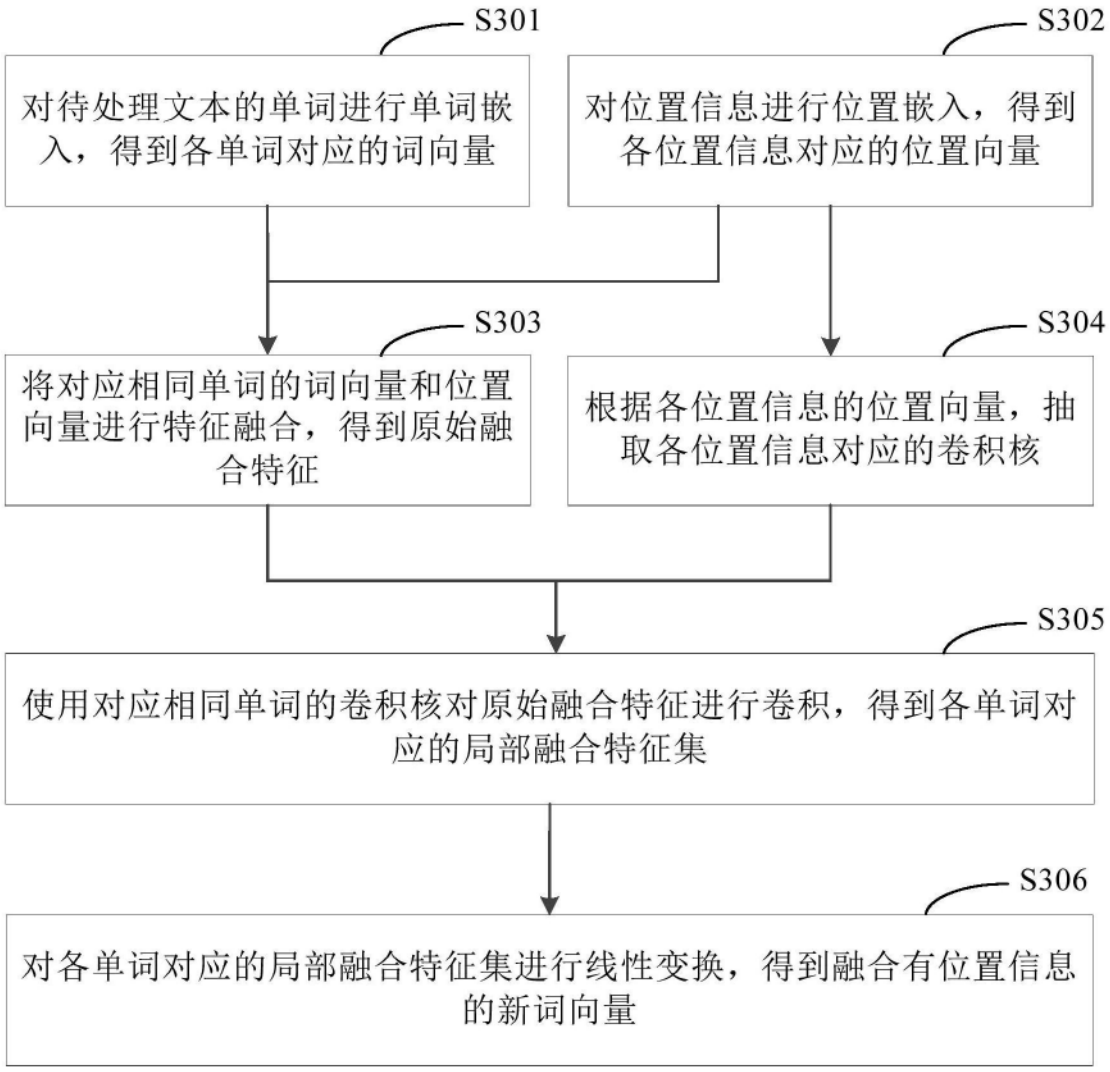

S301
对待处理文本的单词进行单词嵌入，得到各单词对应的词向量
S302
对位置信息进行位置嵌入，得到各位置信息对应的位置向量
S303
将对应相同单词的词向量和位置向量进行特征融合，得到原始融合特征
S304
根据各位置信息的位置向量，抽取各位置信息对应的卷积核
S305
使用对应相同单词的卷积核对原始融合特征进行卷积，得到各单词对应的局部融合特征集
S306
对各单词对应的局部融合特征集进行线性变换，得到融合有位置信息的新词向量

图3

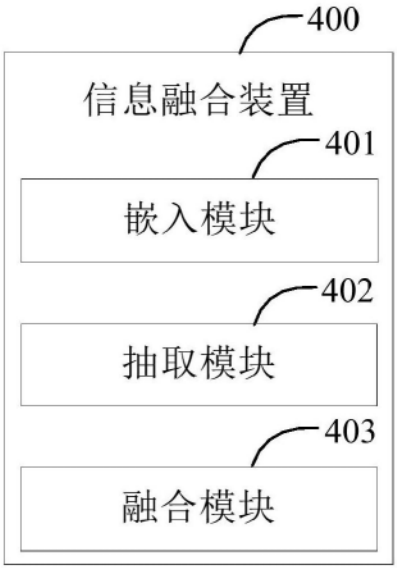


图4

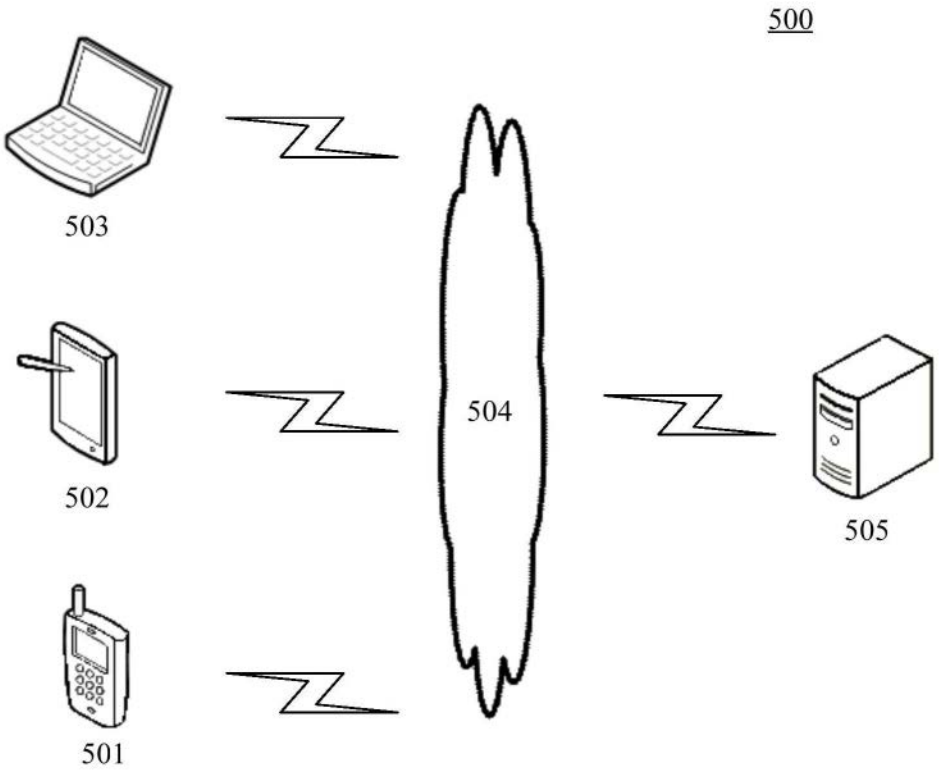


图5

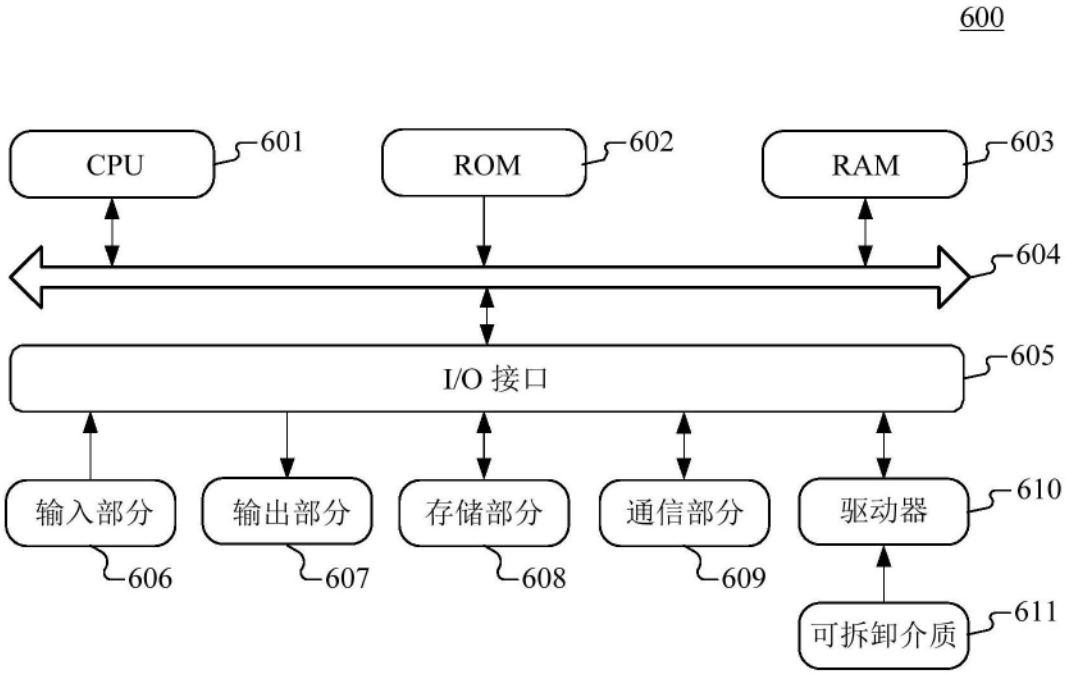

600
CPU
601
ROM
602
RAM
603
604
I/O 接口
605
输入部分
606
输出部分
607
存储部分
608
通信部分
609
驱动器
610
可拆卸介质
611

图6