



(12) 发明专利申请

(10) 申请公布号 CN 117892739 A

(43) 申请公布日 2024. 04. 16

(21) 申请号 202410059666.8

(22) 申请日 2024.01.15

(71) 申请人 京东城市(北京)数字科技有限公司

地址 100086 北京市海淀区知春路76号(写字楼)1号楼9层1-7-5号

申请人 京东科技信息技术有限公司

(72) 发明人 段超群 王培英 陈蒙 何晓冬

(74) 专利代理机构 北京世辉律师事务所 16093

专利代理师 罗利娜

(51) Int. Cl.

G06F 40/58 (2020.01)

G06F 40/289 (2020.01)

G06N 3/0464 (2023.01)

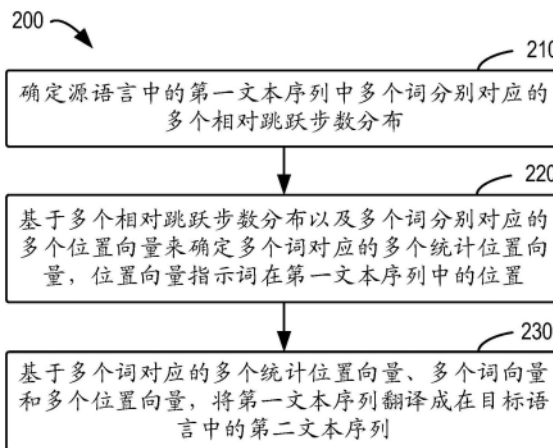
权利要求书2页 说明书13页 附图3页

(54) 发明名称

用于语言翻译的方法、装置、设备和可读介质

(57) 摘要

本公开的实施例提供用于语言翻译的方法、装置、设备和可读介质。该方法包括：确定源语言中的第一文本序列中多个词分别对应的多个相对跳跃步数分布，每个词对应的相对跳跃步数分布指示在将第一文本序列翻译成目标语言后，词在目标语言中的对应词在目标语言中的文本序列中的位置分布；基于多个相对跳跃步数分布以及多个词分别对应的多个位置向量来确定多个词对应的多个统计位置向量，位置向量指示词在第一文本序列中的位置；以及基于多个词对应的多个统计位置向量、多个词向量和多个位置向量，将第一文本序列翻译成在目标语言中的第二文本序列。以此方式，可以提高翻译结果的准确性。



1. 一种用于语言翻译的方法,包括:

确定源语言中的第一文本序列中多个词分别对应的多个相对跳跃步数分布,每个词对应的相对跳跃步数分布指示在将所述第一文本序列翻译成目标语言后,所述词在所述目标语言中的对应词在所述目标语言中的文本序列中的位置分布;

基于所述多个相对跳跃步数分布以及所述多个词分别对应的多个位置向量来确定所述多个词对应的多个统计位置向量,所述位置向量指示所述词在所述第一文本序列中的位置;以及

基于所述多个词对应的所述多个统计位置向量、多个词向量和多个位置向量,将所述第一文本序列翻译成在所述目标语言中的第二文本序列。

2. 根据权利要求1所述的方法,其中确定源语言中的第一文本序列中多个词分别对应的多个相对跳跃步数分布包括:对于所述多个词中的每个词,

确定所述词的多个相对跳跃步数,每个相对跳跃步数指示所述词在所述第一文本序列的位置相对于所述词在所述目标语言中的文本序列中的对应位置的差异;以及

基于所述多个相对跳跃步数确定所述词对应的相对跳跃步数分布。

3. 根据权利要求2所述的方法,其中对于所述多个词中的每个词,确定所述词的多个相对跳跃步数包括:

从语料库获取多个源语言-目标语言文本序列对,每个源语言-目标语言文本序列对中的源语言文本序列至少包括所述词,并且每个源语言-目标语言文本序列对中的源语言文本序列被标注有所述词在所述目标语言中的对应词;以及

从所述多个源语言-目标语言文本序列对中分别确定所述词的多个相对跳跃步数。

4. 根据权利要求3所述的方法,其中所述多个相对跳跃步数是利用经训练的词对齐模型所生成的,所述词对齐模型至少基于所述语料库来训练。

5. 根据权利要求2所述的方法,其中对于所述多个词中的每个词,基于所述多个相对跳跃步数确定所述词对应的相对跳跃步数分布包括:

基于所述多个相对跳跃步数中不同的相对跳跃步数各自对应的数目,以及所述多个相对跳跃步数对应的总数目,来确定所述词对应的相对跳跃步数分布。

6. 根据权利要求1所述的方法,其中确定所述多个词对应的多个统计位置向量包括:

基于所述多个相对跳跃步数分布中每个相对跳跃步数分布以及与该相对跳跃步数分布对应的词的位置向量,确定与所述相对跳跃步数分布相关联的期望结果;以及

将与所述多个相对跳跃步数分布相关联的期望结果确定为所述多个统计位置向量。

7. 根据权利要求1所述的方法,其中将所述第一文本序列翻译成所述目标语言中的第二文本序列包括:

针对所述多个词中的每个词,

基于所述词对应的统计位置向量以及词向量确定第一特征表示,

基于所述词对应的词向量以及位置向量确定第二特征表示,

通过对所述第一特征表示和所述第二特征表示进行加权融合来生成所述词在所述目标语言中的对应词;以及

基于所述多个词中每个词在所述目标语言中的对应词,生成所述第一文本序列在所述目标语言中的所述第二文本序列。

8.一种用于语言翻译的装置,包括:

分布确定模块,被配置为确定源语言中的第一文本序列中多个词分别对应的多个相对跳跃步数分布,每个词对应的相对跳跃步数分布指示在将所述第一文本序列翻译成目标语言后,所述词在所述目标语言中的对应词在所述目标语言中的文本序列中的位置分布;

向量确定模块,被配置为基于所述多个相对跳跃步数分布以及所述多个词分别对应的多个位置向量来确定所述多个词对应的多个统计位置向量,所述位置向量指示所述词在所述第一文本序列中的位置;以及

文本翻译模块,被配置为基于所述多个词对应的所述多个统计位置向量、多个词向量和多个位置向量,将所述第一文本序列翻译成在所述目标语言中的第二文本序列。

9.一种电子设备,包括:

至少一个处理单元;以及

至少一个存储器,所述至少一个存储器被耦合到所述至少一个处理单元并且存储用于由所述至少一个处理单元执行的指令,所述指令在由所述至少一个处理单元执行时使所述电子设备执行根据权利要求1至7任一项所述的方法。

10.一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序可由处理器执行以实现根据权利要求1至7任一项所述的方法。

用于语言翻译的方法、装置、设备和可读介质

技术领域

[0001] 本公开的示例实施例总体涉及计算机技术领域,并且更具体地,涉及用于语言翻译的方法、装置、设备和计算机可读存储介质。

背景技术

[0002] 随着互联网技术的发展,越来越多的应用或平台等均提供自然语言处理功能,给广大用户带来了诸多便利。具有自然语言处理功能的应用或平台可以基于Transformer(转换神经网络)模型向用户提供自然语言处理服务。语言翻译任务是自然语言处理任务中的重要任务。人们期望能够提高利用Transformer模型实现语言翻译的准确性。

发明内容

[0003] 在本公开的第一方面,提供一种用于语言翻译的方法。该方法包括:确定源语言中的第一文本序列中多个词分别对应的多个相对跳跃步数分布,每个词对应的相对跳跃步数分布指示在将第一文本序列翻译成目标语言后,词在目标语言中的对应词在目标语言中的文本序列中的位置分布;基于多个相对跳跃步数分布以及多个词分别对应的多个位置向量来确定多个词对应的多个统计位置向量,位置向量指示词在第一文本序列中的位置;以及基于多个词对应的多个统计位置向量、多个词向量和多个位置向量,将第一文本序列翻译成在目标语言中的第二文本序列。

[0004] 在本公开的第二方面,提供一种用于语言翻译的装置。该装置包括:分布确定模块,被配置为确定源语言中的第一文本序列中多个词分别对应的多个相对跳跃步数分布,每个词对应的相对跳跃步数分布指示在将第一文本序列翻译成目标语言后,词在目标语言中的对应词在目标语言中的文本序列中的位置分布;向量确定模块,被配置为基于多个相对跳跃步数分布以及多个词分别对应的多个位置向量来确定多个词对应的多个统计位置向量,位置向量指示词在第一文本序列中的位置;以及文本翻译模块,被配置为基于多个词对应的多个统计位置向量、多个词向量和多个位置向量,将第一文本序列翻译成在目标语言中的第二文本序列。

[0005] 在本公开的第三方面,提供了一种电子设备。该电子设备包括至少一个处理单元;以及至少一个存储器,至少一个存储器被耦合到至少一个处理单元并且存储用于由至少一个处理单元执行的指令,指令在由至少一个处理单元执行时使电子设备执行本公开第一方面的方法。

[0006] 在本公开的第四方面,提供了一种计算机可读存储介质。该计算机可读存储介质上存储有计算机程序,其可由处理器执行以执行根据本公开的第一方面的方法。

[0007] 应当理解,发明内容部分中所描述的内容并非旨在限定本公开的实施例的关键特征或重要特征,也不用于限制本公开的范围。本公开的其它特征将通过以下的描述而变得容易理解。

附图说明

[0008] 在下文中,结合附图并参考以下详细说明,本公开各实现方式的上述和其他特征、优点及方面将变得更加明显。在附图中,相同或相似的附图标记表示相同或相似的元素,其中:

[0009] 图1示出了本公开的实施例能够在其中实现的示例环境的示意图;

[0010] 图2示出了根据本公开的一些实施例的用于语言翻译的过程的流程图;

[0011] 图3示出了根据本公开的一些实施例的计算相对跳跃步数分布的过程的示意图;

[0012] 图4示出了根据本公开的一些实施例的示例Transformer模型的示意图;

[0013] 图5示出了根据本公开的某些实施例的用于语言翻译的装置的示意性结构框图;以及

[0014] 图6示出了其中可以实施本公开的一个或多个实施例的计算设备的框图。

具体实施方式

[0015] 下面将参照附图更详细地描述本公开的实施例。虽然附图中示出了本公开的一些实施例,然而应当理解的是,本公开可以通过各种形式来实现,而且不应该被解释为限于这里阐述的实施例,相反,提供这些实施例是为了更加透彻和完整地理解本公开。应当理解的是,本公开的附图及实施例仅用于示例性作用,并非用于限制本公开的保护范围。

[0016] 在本公开的实施例的描述中,术语“包括”及其类似用语应当理解为开放性包含,即“包括但不限于”。术语“基于”应当理解为“至少部分地基于”。术语“一个实施例”或“该实施例”应当理解为“至少一个实施例”。术语“一些实施例”应当理解为“至少一些实施例”。下文还可能包括其它明确的和隐含的定义。

[0017] 需要说明的是,本公开的技术方案中,所涉及的用户个人信息的获取,存储和应用等,均符合相关法律法规的规定,且不违背公序良俗。

[0018] 可以理解的是,在使用本公开各实施例公开的技术方案之前,均应当根据相关法律法规通过适当的方式对本公开所涉及个人信息的类型、使用范围、使用场景等告知用户并获得用户的授权。

[0019] 例如,在响应于接收到用户的主动请求时,向用户发送提示信息,以明确地提示用户,其请求执行的操作将需要获得和使用到用户的个人信息,从而使得用户可以根据提示信息来自主地选择是否向执行本公开技术方案的操作的电子设备、应用程序、服务器或存储介质等软件或硬件提供个人信息。

[0020] 作为一种可选的但非限制性的实现方式,响应于接收到用户的主动请求,向用户发送提示信息的方式,例如可以是弹出窗口的方式,弹出窗口中可以以文字的方式呈现提示信息。此外,弹出窗口中还可以承载供用户选择“同意”或“不同意”向电子设备提供个人信息的选择控件。

[0021] 可以理解的是,上述通知和获取用户授权过程仅是示意性的,不对本公开的实施例构成限定,其他满足相关法律法规的方式也可应用于本公开的实施例中。

[0022] 如本文中所使用的,术语“模型”可以从训练数据中学习到相应的输入与输出之间的关联关系,从而在训练完成后可以针对给定的输入,生成对应的输出。模型的生成可以基于机器学习技术。深度学习是一种机器学习算法,通过使用多层处理单元来处理输入和提

供相应输出。神经网络模型是基于深度学习的模型的一个示例。在本文中,“模型”也可以被称为“机器学习模型”、“学习模型”、“机器学习网络”或“学习网络”,这些术语在本文中可互换地使用。

[0023] “神经网络”是一种基于深度学习的机器学习网络。神经网络能够处理输入并且提供相应输出,其通常包括输入层和输出层以及在输入层与输出层之间的一个或多个隐藏层。在深度学习应用中使用的神经网络通常包括许多隐藏层,从而增加网络的深度。神经网络的各个层按顺序相连,从而前一层的输出被提供作为后一层的输入,其中输入层接收神经网络的输入,而输出层的输出作为神经网络的最终输出。神经网络的每个层包括一个或多个节点(也称为处理节点或神经元),每个节点处理来自上一层的输入。

[0024] 通常,机器学习大致可以包括三个阶段,即训练阶段、测试阶段和应用阶段(也称为推理阶段)。在训练阶段,给定的模型可以使用大量的训练数据进行训练,不断迭代更新参数值,直到模型能够从训练数据中获得一致的满足预期目标的推理。通过训练,模型可以被认为能够从训练数据中学习从输入到输出之间的关联(也称为输入到输出的映射)。训练后的模型的参数值被确定。在测试阶段,将测试输入应用到训练后的模型,测试模型是否能够提供正确的输出,从而确定模型的性能。在应用阶段,模型可以被用于基于训练得到的参数值,对实际的输入进行处理,确定对应的输出。

[0025] 图1示出了本公开的实施例能够在其中实现的示例环境100的示意图。如图1所示,环境100可以包括电子设备110。

[0026] 电子设备110可以确定源语言文本序列102(也可称为源语言句子Src)对应的目标语言文本序列112(也可称为目标语言句子Trg)。也即,电子设备110可以对源语言文本序列102执行语言翻译任务以生成对应的目标语言序列112。这里的源语言文本序列102和目标语言文本序列112可以是任意适当的两个不同语言的文本序列。这里的源语言文本序列102可以使电子设备110本地的文本序列,也可以是电子设备110实时获取到的文本序列。

[0027] 电子设备110例如可以利用经训练的机器学习模型115来执行语言翻译任务。机器学习模型115例如可以包括统计机器翻译模型、词对齐模型、Transformer模型、卷积神经网络(CNN)、循环神经网络(RNN)等中的至少一项。机器学习模型115可以是电子设备110本地的模型,也可以是被安装在其他电子设备110的模型(例如安装在远端设备中)。

[0028] 电子设备110可以包括具有计算能力的任何计算系统,例如各种计算设备/系统、终端设备、服务端设备等。终端设备可以是任意类型的移动终端、固定终端或便携式终端,包括移动手机、台式计算机、膝上型计算机、笔记本计算机、上网本计算机、平板计算机、媒体计算机、多媒体平板、掌上电脑、便携式游戏终端、VR/AR设备、个人通信系统(Personal Communication System,PCS)设备、个人导航设备、个人数字助理(Personal Digital Assistant,PDA)、音频/视频播放器、数码相机/摄像机、定位设备、电视接收器、无线电广播接收器、电子书设备、游戏设备或者前述各项的任意组合,包括这些设备的配件和外设或者其任意组合。服务端设备可以是独立的物理服务器,也可以是多个物理服务器构成的服务器集群或者分布式系统,还可以是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、内容分发网络、以及大数据和人工智能平台等基础云计算服务的云服务器。服务端设备例如可以包括计算系统/服务器,诸如大型机、边缘计算节点、云环境中的计算设备,等等。

[0029] 应当理解,仅出于示例性的目的描述环境100中各个元素的结构和功能,而不暗示对于本公开的范围的任何限制。

[0030] 如前文所简要提及的,具有自然语言处理功能的应用或平台可以基于Transformer模型向用户提供自然语言处理服务。语言翻译任务是自然语言处理任务中的重要任务。人们期望能够提高利用Transformer模型实现语言翻译的准确性。由于Transformer模型在结构上主要依赖自注意力机制(Self-Attention, SAN),该机制不同于以往的卷积神经网络或者循环神经网络,它本身无法自动学习到输入文本序列中的位置信息,这会影响最终翻译结果的准确性。为了解决这个问题,提出了引入位置向量来弥补这个缺陷,即用一个向量来表示一个单词在文本序列中的绝对位置。通过引入位置向量,Transformer模型获得了较好的单个文本序列的编码能力。

[0031] 然而,尽管位置向量能够很好地弥补Transformer模型无法捕获单个文本序列中单词位置的缺陷,但是语言翻译任务涉及至少两个不同语种的文本序列,这两个文本序列可以分别称为源语言文本序列和目标语言文本序列。这两个文本序列中对应的词语的位置也可能存在差异。示例性的,若源语言文本序列为“这需要在安全商业方面分别达成协议”,Transformer模型输出的目标语言文本序列例如可以为“this requires separate security, commercial agreements”。若调整源语言文本序列中多个词的词序,使得调整后的源语言文本序列为“这需要分别达成协议在安全商业方面”,则Transformer模型输出的目标语言文本序列例如可以为“this requires separate agreements on security and commercial aspects”。由此可见,两个文本序列中对应的词语的位置之间的差异会严重影响模型的翻译性能,且Transformer模型的位置向量并不能从输入中捕获这种差异。也即不同语种间单词顺序的差异会影响翻译模型的性能且传统的Transformer模型并不能解决该问题。

[0032] 因此,传统的Transformer模型无法充分利用翻译数据中的统计信息。翻译数据中包含了源语言和目标语言之间的词对齐信息,传统的Transformer模型使用的是端到端的方式,无法高效地利用这些统计信息。此外,传统的Transformer模型使用的位置向量机制也只能用于捕获单个文本序列中的词序信息,并不能很好地处理翻译任务中不同语种间词序不同的问题,该问题会导致Transformer模型性能的下降。

[0033] 有鉴于此,本公开的实施例提供了一种用于语言翻译的方法。该方法包括:确定源语言中的第一文本序列中多个词分别对应的多个相对跳跃步数分布,每个词对应的相对跳跃步数分布指示在将第一文本序列翻译成目标语言后,词在目标语言中的对应词在目标语言中的文本序列中的位置分布;基于多个相对跳跃步数分布以及多个词分别对应的多个位置向量来确定多个词对应的多个统计位置向量,位置向量指示词在第一文本序列中的位置;以及基于多个词对应的多个统计位置向量、多个词向量和多个位置向量,将第一文本序列翻译成在目标语言中的第二文本序列。以此方式,本公开的实施例可以利用翻译数据中的词对齐信息,增强模型对翻译数据的利用率,降低不同语种中次序差异造成的翻译误差,有助于提高翻译结果的准确性。

[0034] 以下将继续参考附图描述本公开的一些示例实施例。

[0035] 图2示出了根据本公开的一些实施例的用于语言翻译的过程200的流程图。过程200可以被实现在电子设备110处。为便于讨论,将参考图1的环境100来描述过程200。

[0036] 在框210,电子设备110确定源语言中的第一文本序列中多个词分别对应的多个相对跳跃步数分布。每个词对应的相对跳跃步数分布指示在将第一文本序列翻译成目标语言后,词在目标语言中的对应词在目标语言中的文本序列中的位置分布。

[0037] 电子设备110可以获取要被翻译的源语言的第一文本序列。这里的源语言可以是任意适当的语言,这里的第一文本序列可以包括任意适当数目的多个词。在一些实施例中,电子设备110可以直接从其他电子设备处获取第一文本序列中多个词分别对应的多个相对跳跃步数分布。在一些实施例中,电子设备110还可以基于第一文本序列自行确定第一文本序列中多个词分别对应的多个相对跳跃步数分布。

[0038] 具体地,针对第一文本序列所包括的多个词中的每个词,电子设备110可以确定该词的多个相对跳跃步数,进而基于多个相对跳跃步数确定词对应的相对跳跃步数分布。这里的每个相对跳跃步数可以指示词在第一文本序列的位置相对于词在目标语言中的文本序列中的对应位置的差异。示例性的,若词A在第一文本序列中的位置为2,词A在目标语言中的文本序列中的对应位置为4,则相对跳跃步数例如可以为2。这表示词A在目标语言中的文本序列中的对应位置为在第一文本序列中的位置向右2个位置处。电子设备110例如可以使用变量 $\delta \in [-K, +K]$ 来表示每个词的相对跳跃步数,其中K表示最大跳跃步数。也即翻译成目标语言下的文本序列后,每个词的位置位于 $\{j-K, \dots, j+K\}$ 。示例性的,若一个词的跳跃步数 $\delta = -2$,则表示该词向左跳跃2步。若一个词的跳跃步数 $\delta = 2$,则表示该词向右跳跃2步。若一个词的跳跃步数 $\delta = 0$,则表示该词保持在原位置。

[0039] 关于确定多个相对跳跃步数的具体方式,在一些实施例中,电子设备110可以获取包括大量源语言-目标语言文本序列对的语料库。语料库中的每个源语言-目标语言文本序列对中的源语言文本序列均被标注有词在目标语言中的对应词。例如,若目标语言文本序列包括词A、B、C,则目标语言文本序列还标注有词A、B、C在目标语言中的对应词,例如为词a、b、c。

[0040] 若要确定目标词的多个相对跳跃步数,电子设备110可以从语料库获取其中每个源语言-目标语言文本序列对中的源语言文本序列至少包括目标词的多个源语言-目标语言文本序列对。示例性的,若要确定词A的多个相对跳跃步数,则电子设备110获取到的多个源语言-目标语言文本序列对中每个源语言文本序列至少包括词A。电子设备110可以从多个源语言-目标语言文本序列对中分别确定词的多个相对跳跃步数。

[0041] 图3示出了根据本公开的一些实施例的计算相对跳跃步数分布的过程的示意图。如图3所示,电子设备110获取到的语料库包括源语言-目标语言文本序列对301。源语言-目标语言文本序列对301中的源语言文本序列均被标注有词在目标语言中的对应词。示例性的,源语言文本序列“继续改革的努力”中与“继续”、“改革”、“的”、“努力”这多个词相关联的被标注有在目标语言中的对应词“continue”、“reform”、“to”和“efforts”。源语言-目标语言文本序列对301中与该源语言文本序列“继续改革的努力”对应的目标语言文本序列为“the efforts to continue reform”。

[0042] 可以基于目标词在源语言文本序列中的位置以及其的对应词在目标语言文本序列中的位置来确定目标词的相对跳跃位置。可以理解,目标词可以使源语言文本序列中的任意词。示例性的,“继续”这个词在源语言文本序列中的位置为1,其的对应词“continue”在目标语言文本序列中的位置为4,则可以确定“继续”这个词的相对跳跃步数为3。类似地,

可以确定“改革”这个词的相对跳跃步数为3,“的”这个词的相对跳跃步数为0,“努力”这个词包含两个相对跳跃步数,分别为-3和-2。

[0043] 关于上述确定多个相对跳跃步数的具体方式,在一些实施例中,这多个相对跳跃步数可以是人工获取的。备选的或者附加地,在一些实施例中,这多个相对跳跃步数还可以是电子设备110自行获取的。具体地,这多个相对跳跃步数可以是电子设备110利用经训练的词对齐模型所生成的。这个词对齐模型至少基于语料库来训练。这个词对齐模型可以是任意适当的模型。示例性的,电子设备110可以使用fast_align工具包,利用词对齐模型来确定多个相对跳跃步数。这里的多个相对跳跃步数也可以被称之为词对齐信息,在本文中,“相对跳跃步数”和“词对齐信息”可以互换地使用。

[0044] 在一些实施例中,由于语料库中包含多个源语言文本序列包括目标词的源语言-目标语言文本序列对,因此目标词可以包括多个相对跳跃步数,这多个相对跳跃步数中存在相同的相对跳跃步数。电子设备110可以统计多个相对跳跃步数中不同的相对跳跃步数的数目。示例性的,若针对5个源语言-目标语言文本序列对,目标词在这5个源语言-目标语言文本序列对种的相对跳跃步数分别为1,1,2,3,2,则电子设备110对这5个相对跳跃数目进行统计可以得到相对跳跃步数为1的数目为2,相对跳跃步数为2的数目为2,相对跳跃步数为3的数目为1。电子设备110例如可以用 $C(\delta|w_j)$ 表示整个语料库中单词 w_j 的相对跳跃步数为 δ 的出现次数。示例性的,如图3所示,若单词 w_j 为“继续”,则“继续”这个词的相对跳跃步数为3的数目可以表示为 $C(3|\text{继续})$ 。

[0045] 可以理解,每个源语言文本序列中均包括多个词,语料库中包含多个源语言-目标语言文本序列对,因此电子设备110可以获取到每个词的多个相对跳跃步数以及这多个相对跳跃步数中不同相对跳跃步数的数目。电子设备110例如可以获取到数据集302,数据集302中包括多个词中每个词的多个相对跳跃步数以及这多个相对跳跃步数中不同相对跳跃步数的数目。电子设备110后续可以直接基于这个数据集302确定目标词的目标相对跳跃步数的数目,这里的目标相对跳跃步数可以为目标词对应的多个相对跳跃步数中的任意相对跳跃步数。示例性的,电子设备110可以直接基于数据集302确定“继续”这个词的相对跳跃步数为3的数目为125,“改革”这个词的相对跳跃步数为3的数目为240等等。

[0046] 在一些实施例中,电子设备110进一步可以基于多个相对跳跃步数中不同的相对跳跃步数各自对应的数目 $C(\delta|w_j)$,以及多个相对跳跃步数对应的总数目 $\sum_k C(\delta|w_k)$,来确定词对应的相对跳跃步数分布。电子设备110例如可以基于如下公式来确定相对跳跃分布 $p(\delta|w_j)$:

$$p(\delta|w_j) = \frac{C(\delta|w_j)}{\sum_k C(\delta|w_k)} \quad (1)$$

[0048] 可以理解,电子设备110可以计算目标词的目标相对跳跃步数的数目和目标词的多个相对跳跃数目的中署目的比值,该比值即为目标词的目标相对跳跃步数的相对跳跃步数分布。相对跳跃步数分布可以指示源语言文本序列被翻译成目标语言文本序列后,源语言文本序列中的词的对应词在目标语言文本序列的各个位置的概率。

[0049] 在一些实施例中,电子设备110可以确定语料库中多个源语言文本序列的多个词分别对应的多个相对跳跃步数分布。电子设备110例如可以将确定的多个相对跳跃步数分布确定为数据集303。电子设备110后续可以直接继续目标词以及目标相对跳跃步数从数据

集303中查询对应的相对跳跃步数分布。示例性的,电子设备110可以直接基于数据集303确定“继续”这个词的相对跳跃步数为3对应的相对跳跃步数分布为0.059。

[0050] 返回参考图2,在框220,电子设备110基于多个相对跳跃步数分布以及多个词分别对应的多个位置向量来确定多个词对应的多个统计位置向量,位置向量指示词在第一文本序列中的位置。

[0051] 位置向量 pe_j 也可以称之为位置编码(Positional Encoding),其可以用词的位置信息对第一文本序列中的每个词进行二次表示。也即其可以表示每个词在第一文本序列中的位置信息。在一些实施例中,电子设备110可以使用三角函数来表示位置向量 pe_j 。示例性的,电子设备110可以基于如下公式来确定位置向量:

$$[0052] \quad pe_{(j,2i)} = \sin(j/10000^{2i/d_{model}}) \quad (2)$$

$$[0053] \quad pe_{(j,2i+1)} = \cos(j/10000^{2i/d_{model}}) \quad (3)$$

[0054] 其中j表示词语在文本序列中的位置,i表示向量中的维数。电子设备110可以基于这样的三角函数公式来分别确定第一文本序列中奇数位置(例如第1位、第3位等)和偶数位置(例如第2位、第4位等)的词对应的位置向量。

[0055] 电子设备110获取到多个词对应的多个相对跳跃步数分布 $p(\delta|w_j)$ 后,可以基于每个词对应的相对跳跃步数分布 $p(\delta|w_j)$ 以及位置向量 pe_j 来确定每个词对应的统计位置向量 spe_j 。具体地,电子设备110可以基于每个词对应的相对跳跃步数分布 $p(\delta|w_j)$ 以及位置向量 pe_j 来确定每个词对应的期望结果,该期望结果即为每个词对应的统计位置向量 spe_j 。示例性的,电子设备110例如可以基于如下公式来确定每个词对应的统计位置向量 spe_j :

$$[0056] \quad spe_j = \sum_{\delta=-K}^K (pe_{j+\delta} \cdot p(\delta|w_j)) \quad (4)$$

[0057] 其中, $pe_{j+\delta}$ 表示施加相对跳跃步数 δ 后的词的位置向量 pe_j 。由此,电子设备110可以依次确定多个词中每个词对应的统计位置向量 spe_j ,进而确定第一文本序列所包括的多个词对应的多个统计位置向量 spe_j 。

[0058] 继续参考图3,电子设备110获取数据集303后,电子设备110可以基于第一文本序列中包括的多个词对数据集303进行查询,以得到这多个词分别对应的多个相对跳跃步数分布 $p(\delta|w_j)$ 。示例性的,若第一文本序列包括多个词304,电子设备110若需要确定多个词304中的“市民”这个词的多个相对跳跃步数分布 $p(\delta|w_j)$,电子设备110可以在数据集303中基于“市民”这个词进行查询以得到查询结果305。查询结果305中包括“市民”这个词对应多个相对跳跃步数 δ 的多个相对跳跃步数分布 $p(\delta|w_j)$ 。电子设备110可以基于查询结果305来确定“市民”这个词对应的统计位置向量 spe_j 。

[0059] 在一些实施例中,电子设备110可以直接基于查询结果305所包括的多个相对跳跃步数分布 $p(\delta|w_j)$ 来确定统计位置向量 spe_j 。在一些实施例中,电子设备110可以将查询结果305所包括的多个相对跳跃步数分布 $p(\delta|w_j)$ 进行格式转换以得到数据集306。电子设备110进而可以利用数据集306来确定统计位置向量 spe_j 。需要注意的是,在一些实施例中,电子设备110可以利用查询结果305所包括的全部相对跳跃步数分布 $p(\delta|w_j)$ 来计算统计位置向量 spe_j 。在一些实施例中,电子设备110也可以进利用查询结果305中大于阈值的多个相对跳跃步数分布 $p(\delta|w_j)$ 来计算统计位置向量 spe_j 。“市民”这个词的统计位置向量例如可以为

向量307, dw_3 例如可以表示该词是源语言文本序列中的第三个词。

[0060] 返回参考图2, 在框230, 电子设备110基于多个词对应的多个统计位置向量、多个词向量和多个位置向量, 将第一文本序列翻译成在目标语言中的第二文本序列。

[0061] 在一些实施例中, 电子设备110可以分别基于第一文本序列的多个词对应的多个统计位置向量和多个词向量以及多个词对应的多个词向量和多个位置向量来分别确定第一特征表示和第二特征表示。电子设备110进而可以基于第一特征表示和第二特征表示来生成在目标语言中的第二文本序列。

[0062] 具体地, 针对多个词中的每个词, 电子设备110可以基于词对应的统计位置向量以及词向量确定第一特征表示。电子设备110例如可以基于如下公式来确定第一输入向量 T_0 :

$$[0063] \quad T_0 = x_j + \text{spe}_j \quad (5)$$

[0064] 电子设备110例如可以通过对统计位置向量 spe_j 以及词向量 x_j 进行相加以得到第一输入向量 T_0 。电子设备110可以基于第一输入向量 T_0 确定第一编码向量 T_e^N 。电子设备110进而可以至少基于第一编码向量 T_e^N 确定第一特征表示 T_d^N 。

[0065] 电子设备110还可以基于词对应的词向量以及位置向量确定第二特征表示。电子设备110例如可以基于如下公式来确定第二输入向量 H_0 :

$$[0066] \quad H_0 = x_j + \text{pe}_j \quad (6)$$

电子设备110例如可以通过对位置向量 pe_j 以及词向量 x_j 进行相加以得到第二输入向量 H_0 。电子设备110可以基于第二输入向量 H_0 确定第二编码向量 H_e^N 。电子设备110进而可以至少基于第二编码向量 H_e^N 确定第二特征表示 S_d^N 。

[0067] 在一些实施例中, 电子设备110可以基于门 (Gate) 机制来确定第一特征表示 T_d^N 和第二特征表示 S_d^N 分别对应的第一权重和第二权重。电子设备110可以分别利用第一权重和第二权重对第一特征表示 T_d^N 和第二特征表示 S_d^N 进行加权, 得到加权后的第一特征表示 T_d^N 和第二特征表示 S_d^N 。电子设备110进而可以通过对加权后的第一特征表示 T_d^N 和第二特征表示 S_d^N 进行融合来生成词在目标语言中的对应词。

[0068] 具体来说, 使用第一特征表示 T_d^N 和第二特征表示 S_d^N 来计算权重 $\text{gate} \in [0, 1]$, 所计算的权重可以用于平衡第一特征表示 T_d^N 和第二特征表示 S_d^N 在后续融合过程中的重要性。电子设备110例如可以基于如下公式来确定权重:

$$[0069] \quad \text{gate} = \text{sigmoid}([S_d^N; T_d^N]W) \quad (7)$$

$$[0070] \quad S_d^N = \text{gate} \odot S_d^N + (1 - \text{gate}) \odot T_d^N \quad (8)$$

[0071] 其中sigmoid表示使用sigmoid函数来确定权重gate。 $\text{gate} \odot S_d^N$ 表示使用权重gate对第二特征表示 S_d^N 进行加权所得到的加权后的第二特征表示 S_d^N , $(1 - \text{gate}) \odot T_d^N$ 表示使用权重 $(1 - \text{gate})$ 对第一特征表示 T_d^N 进行加权所得到的加权后的第一特征表示 T_d^N 。 S_d^N 表示对加权后的第一特征表示 T_d^N 和第二特征表示 S_d^N 进行融合所得到的融合结果。在一些实施例中, 电子设备110进而可以对融合结果 S_d^N 进行处理以生成第一文本序列中的词在目标语言中的对应词。电子设备110可以基于多个词中每个词在目标语言中的对应词, 生成第一文本序列在目标语言中的第二文本序列。

[0072] 在一些实施例中,电子设备110可以利用Transformer模型来基于多个词对应的多个统计位置向量、多个词向量和多个位置向量,将第一文本序列翻译成在目标语言中的第二文本序列。图4示出了根据本公开的一些实施例的示例Transformer模型400的示意图。如图4所示,Transformer模型400中可以包含N个编码器410、N个编码器420以及N个解码器430,N为大于等于1的正整数。

[0073] Transformer模型400包括输入层402和输入层404。电子设备110获取到多个词对应的多个统计位置向量 spe_j 、多个词向量 x_j 和多个位置向量 pe_j 后,可以将多个统计位置向量 spe_j 和多个词向量 x_j 作为输入层402的输入,将多个词向量 x_j 和多个位置向量 pe_j 作为输入层404的输入。

[0074] 在一些实施例中,电子设备110也可以仅获取到的多个相对跳跃步数 δ 以及第一文本序列。电子设备110将第一文本序列同第一文本序列中的多个词对应的多个相对跳跃步数 δ 提供给输入层402。输入层402的输入嵌入模块可以用于将第一文本序列中的多个词转换层可以表示每个词的语义信息的多个词向量 x_j 。示例性的,输入层402可以将多个词映射到多个向量来表示这多个词的语义信息,被映射的多个向量即为多个词向量 x_j 。输入层402还可以基于多个相对跳跃步数 δ 确定对应的统计位置信息,也可以称之为统计位置向量 spe_j 。输入层402可以对统计位置向量 spe_j 以及词向量 x_j 进行相加以得到第一输入向量 T_0 。

[0075] 同样的,电子设备110将第一文本序列同第一文本序列中的多个词对应的多个相对跳跃步数 δ 提供给输入层404。输入层404的输入嵌入模块可以用于将第一文本序列中的多个词转换层可以表示每个词的语义信息的多个词向量 x_j 。输入层404例如还可以基于三角函数来确定每个词对应的位置向量 pe_j 。输入层404可以对位置向量 pe_j 以及词向量 x_j 进行相加以得到第二输入向量 H_0 。

[0076] 第一输入向量 T_0 被提供给编码器410。编码器410例如可以通过多头注意力模块、ADD&NORM层、前馈模块对第一输入向量 T_0 进行编码以得到第一输入向量 T_0 对应的第一编码向量 T_e^N 。第二输入向量 i_0 被提供给编码器模型420。编码器420例如可以通过多头注意力模块、ADD&NORM层、前馈模块对第二输入向量 i_0 进行编码以得到第二输入向量 i_0 对应的第二编码向量 H_e^N 。

[0077] 第一编码向量 T_e^N 和第二编码向量 H_e^N 将被提供给解码器430。在一些实施例中,解码器430的输入除了第一编码向量 T_e^N 和第二编码向量 H_e^N 外,还包括先前的输出。这里是由于模型是逐词翻译的,也即模型会先基于第一个词对应的统计位置向量 spe_j 、词向量 x_j 和位置向量 pe_j 来输出第一个词在目标语言中的对应词。模型在生成第二个词在目标语言中的对应词时,先前的输出例如可以为第一个词的对应词。模型在生成第三个词在目标语言中的对应词时,先前的输出例如可以为第一个词的对应词和第二个词的对应词,或者第二个词的对应词。也即,先前的输出即为模型先前生成的对应词。

[0078] 相较于传统的仅包括一个多头注意力模块+ADD&NORM层的解码器,解码器430包括两个多头注意力模块+ADD&NORM层。多头注意力模块+ADD&NORM层也可以称之为ATT层。解码器430的两个ATT层可以分别用于处理第一编码向量 T_e^N 和第二编码向量 H_e^N 以输出第一特征表示 T_d^N 和第二特征表示 S_d^N 。第一特征表示 T_d^N 和第二特征表示 S_d^N 被提供给融合门模块(Fusion Gate)。融合门模块例如可以基于门机制来确定第一特征表示 T_d^N 和第二特征表示

S_a^N 各自的权重并对二者进行加权计算,融合门模块还可以对加权后的第一特征表示 T_a^N 和第二特征表示 S_a^N 进行融合以确定融合结果 S_a^N 。解码器430进而可以利用前馈模块、ADD&NORM层等输出融合结果 S_a^N 对应的解码结果。解码结果被提供给Transformer模型400的线性层和SOFTMAX层,线性层和SOFTMAX层也可以统称为Transformer模型400的输出层。输出层可以对解码结果进行处理以输出与输入的第一文本序列中的词对应的目标语言下的对应词。电子设备110可以基于多个词中每个词在目标语言中的对应词,生成第一文本序列在目标语言中的第二文本序列。由此,可以为Transformer模型设计了一种统计位置向量机制来辅助Transformer模型捕获不同语种句对间词序差异,可以将统计位置向量集成到Transformer中,以增强其翻译性能。

[0079] 综上所述,本公开的实施例可以引入第一文本序列的统计位置信息,并利用该统计位置信息来将第一文本序列翻译成在目标语言中的第二文本序列。以此方式,本公开的实施例可以利用翻译数据中的词对齐信息,增强模型对翻译数据的利用率,降低不同语种中次序差异造成的翻译误差,有助于提高翻译结果的准确性。

[0080] 本公开的实施例还提供了用于实现上述方法或过程的相应装置。图5示出了根据本公开的某些实施例的用于语言翻译的装置500的示意性结构框图。装置500可以被实现为或者被包括在电子设备110中。装置500中的各个模块/组件可以由硬件、软件、固件或者它们的任意组合来实现。

[0081] 如图5所示,装置500包括分布确定模块510,被配置为确定源语言中的第一文本序列中多个词分别对应的多个相对跳跃步数分布,每个词对应的相对跳跃步数分布指示在将第一文本序列翻译成目标语言后,词在目标语言中的对应词在目标语言中的文本序列中的位置分布。装置500还包括向量确定模块520,被配置为基于多个相对跳跃步数分布以及多个词分别对应的多个位置向量来确定多个词对应的多个统计位置向量,位置向量指示词在第一文本序列中的位置。装置500还包括文本翻译模块530,被配置为基于多个词对应的多个统计位置向量、多个词向量和多个位置向量,将第一文本序列翻译成在目标语言中的第二文本序列。

[0082] 在一些实施例中,分布确定模块510包括:步数确定模块,被配置为对于多个词中的每个词,确定词的多个相对跳跃步数,每个相对跳跃步数指示词在第一文本序列的位置相对于词在目标语言中的文本序列中的对应位置的差异;以及步数分布确定模块,被配置为基于多个相对跳跃步数确定词对应的相对跳跃步数分布。

[0083] 在一些实施例中,对于多个词中的每个词,步数确定模块包括:语料获取模块,被配置为从语料库获取多个源语言-目标语言文本序列对,每个源语言-目标语言文本序列对中的源语言文本序列至少包括词,并且每个源语言-目标语言文本序列对中的源语言文本序列被标注有词在目标语言中的对应词;以及跳跃步数确定模块,被配置为从多个源语言-目标语言文本序列对中分别确定词的多个相对跳跃步数。

[0084] 在一些实施例中,多个相对跳跃步数是利用经训练的词对齐模型所生成的,词对齐模型至少基于语料库来训练。

[0085] 在一些实施例中,对于多个词中的每个词,步数分布确定模块被进一步配置为:基于多个相对跳跃步数中不同的相对跳跃步数各自对应的数目,以及多个相对跳跃步数对应的总数目,来确定词对应的相对跳跃步数分布。

[0086] 在一些实施例中,向量确定模块520包括:期望结果确定模块,被配置为基于多个相对跳跃步数分布中每个相对跳跃步数分布以及与该相对跳跃步数分布对应的词的位置向量,确定与相对跳跃步数分布相关联的期望结果;以及确定模块,被配置为将与多个相对跳跃步数分布相关联的期望结果确定为多个统计位置向量。

[0087] 在一些实施例中,文本翻译模块530被进一步配置为:针对多个词中的每个词,基于词对应的统计位置向量以及词向量确定第一特征表示,基于词对应的词向量以及位置向量确定第二特征表示,通过对第一特征表示和第二特征表示进行加权融合来生成词在目标语言中的对应词;以及基于多个词中每个词在目标语言中的对应词,生成第一文本序列在目标语言中的第二文本序列。

[0088] 装置500中所包括的单元和/或模块可以利用各种方式来实现,包括软件、硬件、固件或其任意组合。在一些实施例中,一个或多个单元和/或模块可以使用软件和/或固件来实现,例如存储在存储介质上的机器可执行指令。除了机器可执行指令之外或者作为替代,装置500中的部分或者全部单元和/或模块可以至少部分地由一个或多个硬件逻辑组件来实现。作为示例而非限制,可以使用的示范类型的硬件逻辑组件包括现场可编程门阵列(FPGA)、专用集成电路(ASIC)、专用标准品(ASSP)、片上系统(SOC)、复杂可编程逻辑器件(CPLD),等等。

[0089] 应当理解,以上方法中的一个或多个步骤可以由适当的电子设备或电子设备的组合来执行。这样的电子设备或电子设备的组合例如可以包括图1中的电子设备110。

[0090] 图6示出了其中可以实施本公开的一个或多个实施例的电子设备600的框图。应当理解,图6所示出的电子设备600仅仅是示例性的,而不应当构成对本文所描述的实施例的功能和范围的任何限制。图6所示出的电子设备600可以用于实现图1的电子设备110或图5的装置500。

[0091] 如图6所示,电子设备600是通用电子设备的形式。电子设备600的组件可以包括但不限于一个或多个处理器或处理单元610、存储器620、存储设备630、一个或多个通信单元640、一个或多个输入设备650以及一个或多个输出设备660。处理单元610可以是实际或虚拟处理器并且能够根据存储器620中存储的程序来执行各种处理。在多处理器系统中,多个处理单元并行执行计算机可执行指令,以提高电子设备600的并行处理能力。

[0092] 电子设备600通常包括多个计算机存储介质。这样的介质可以是电子设备600可访问的任何可以获得的介质,包括但不限于易失性和非易失性介质、可拆卸和不可拆卸介质。存储器620可以是易失性存储器(例如寄存器、高速缓存、随机访问存储器(RAM))、非易失性存储器(例如,只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、闪存)或它们的某种组合。存储设备630可以是可拆卸或不可拆卸的介质,并且可以包括机器可读介质,诸如闪存驱动、磁盘或者任何其他介质,其可以能够用于存储信息和/或数据并且可以在电子设备600内被访问。

[0093] 电子设备600可以进一步包括另外的可拆卸/不可拆卸、易失性/非易失性存储介质。尽管未在图6中示出,可以提供用于从可拆卸、非易失性磁盘(例如“软盘”)进行读取或写入的磁盘驱动和用于从可拆卸、非易失性光盘进行读取或写入的光盘驱动。在这些情况中,每个驱动可以由一个或多个数据介质接口被连接至总线(未示出)。存储器620可以包括计算机程序产品625,其具有一个或多个程序模块,这些程序模块被配置为执行本公开的各

种实施例的各种方法或动作。

[0094] 通信单元640实现通过通信介质与其他电子设备进行通信。附加地,电子设备600的组件的功能可以以单个计算集群或多个计算机器来实现,这些计算机器能够通过通信连接进行通信。因此,电子设备600可以使用与一个或多个其他服务器、网络个人计算机(PC)或者另一个网络节点的逻辑连接来在联网环境中进行操作。

[0095] 输入设备650可以是一个或多个输入设备,例如鼠标、键盘、追踪球等。输出设备660可以是一个或多个输出设备,例如显示器、扬声器、打印机等。电子设备600还可以根据需要通过通信单元640与一个或多个外部设备(未示出)进行通信,外部设备诸如存储设备、显示设备等,与一个或多个使得用户与电子设备600交互的设备进行通信,或者与使得电子设备600与一个或多个其他电子设备通信的任何设备(例如,网卡、调制解调器等)进行通信。这样的通信可以经由输入/输出(I/O)接口(未示出)来执行。

[0096] 根据本公开的示例性实现方式,提供了一种计算机可读存储介质,其上存储有计算机可执行指令,其中计算机可执行指令被处理器执行以实现上文描述的方法。根据本公开的示例性实现方式,还提供了一种计算机程序产品,计算机程序产品被有形地存储在非瞬态计算机可读介质上并且包括计算机可执行指令,而计算机可执行指令被处理器执行以实现上文描述的方法。

[0097] 这里参照根据本公开实现的方法、装置、设备和计算机程序产品的流程图和/或框图描述了本公开的各个方面。应当理解,流程图和/或框图的每个方框以及流程图和/或框图中各方框的组合,都可以由计算机可读程序指令实现。

[0098] 这些计算机可读程序指令可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理单元,从而生产出一种机器,使得这些指令在通过计算机或其他可编程数据处理装置的处理单元执行时,产生了实现流程图和/或框图中的一个或多个方框中规定的功能/动作的装置。也可以把这些计算机可读程序指令存储在计算机可读存储介质中,这些指令使得计算机、可编程数据处理装置和/或其他设备以特定方式工作,从而,存储有指令的计算机可读介质则包括一个制造品,其包括实现流程图和/或框图中的一个或多个方框中规定的功能/动作的各个方面的指令。

[0099] 可以把计算机可读程序指令加载到计算机、其他可编程数据处理装置、或其他设备上,使得在计算机、其他可编程数据处理装置或其他设备上执行一系列操作步骤,以产生计算机实现的过程,从而使得在计算机、其他可编程数据处理装置、或其他设备上执行的指令实现流程图和/或框图中的一个或多个方框中规定的功能/动作。

[0100] 附图中的流程图和框图显示了根据本公开的多个实现的系统、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、程序段或指令的一部分,模块、程序段或指令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。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,两个连续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,框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或动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。

[0101] 以上已经描述了本公开的各实现,上述说明是示例性的,并非穷尽性的,并且也不

限于所公开的各实现。在不偏离所说明的各实现的范围和精神的情况下,对于本技术领域的普通技术人员来说许多修改和变更都是显而易见的。本文中所用术语的选择,旨在最好地解释各实现的原理、实际应用或对市场中的技术的改进,或者使本技术领域的其他普通技术人员能理解本文公开的各个实现方式。

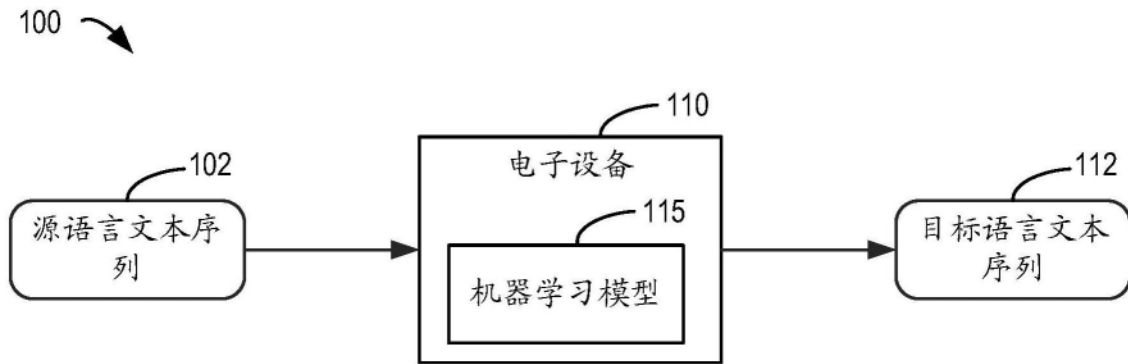


图1

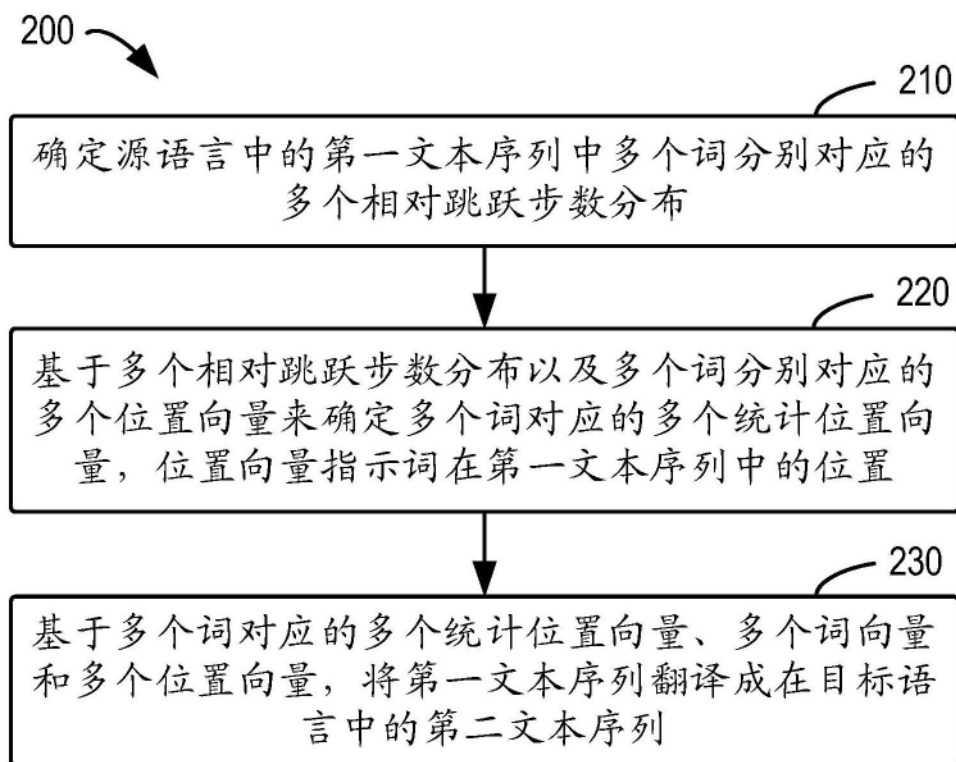


图2

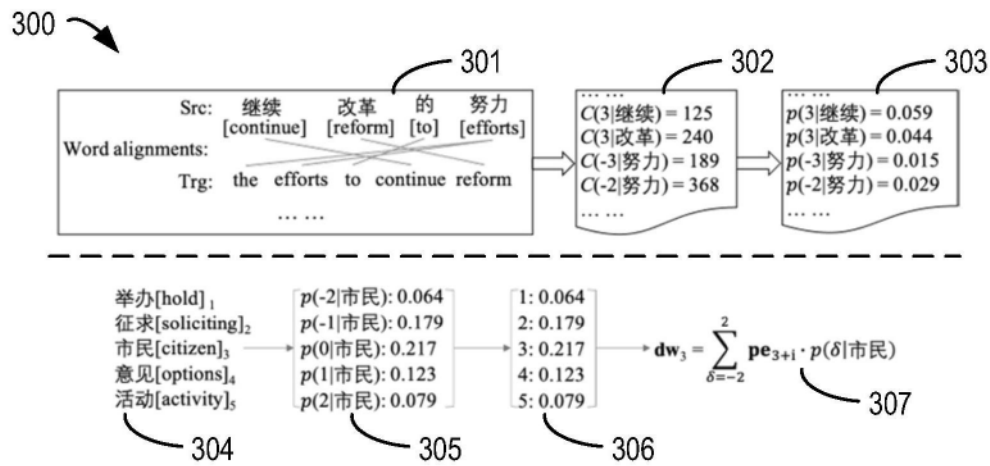


图3

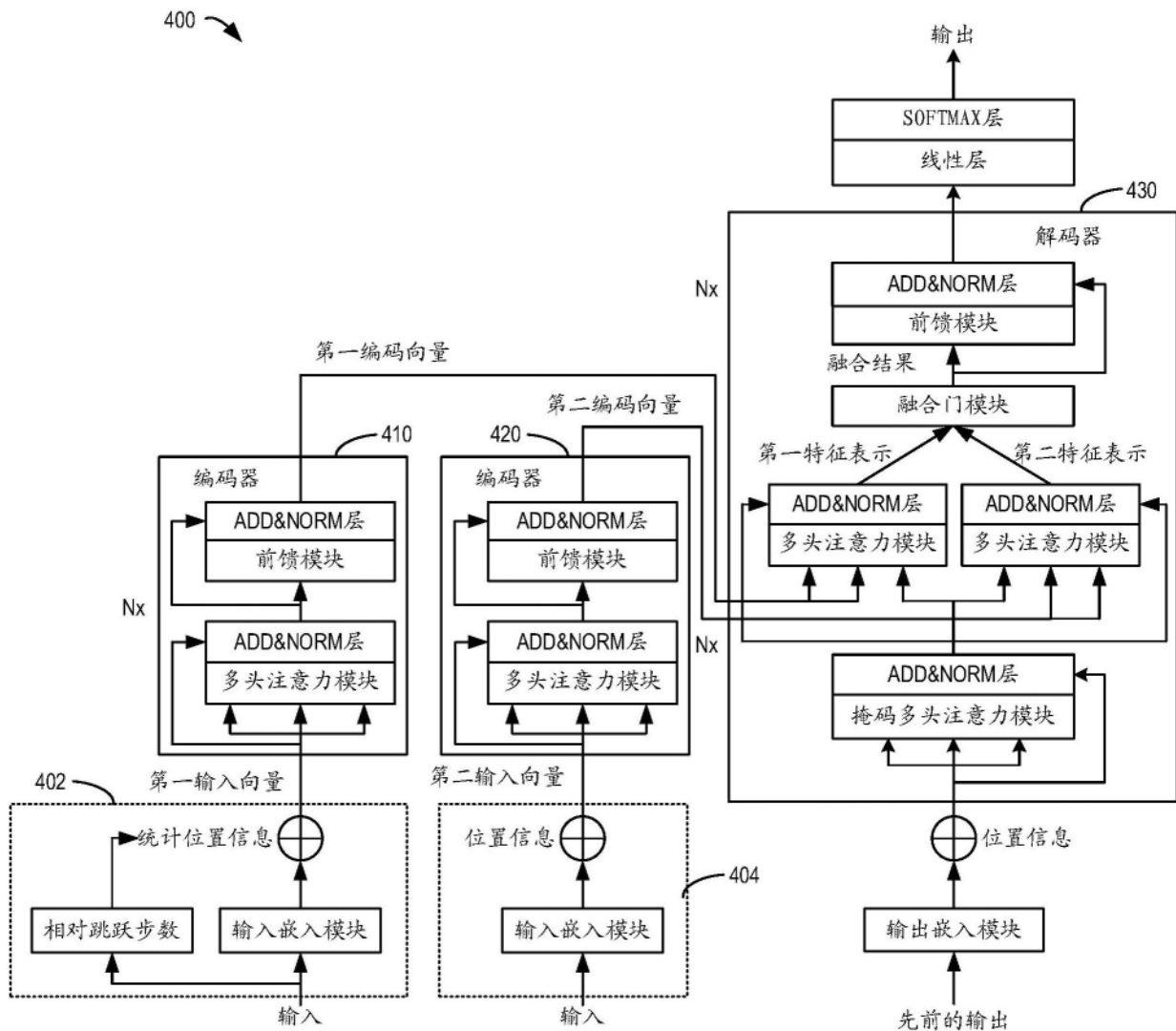


图4

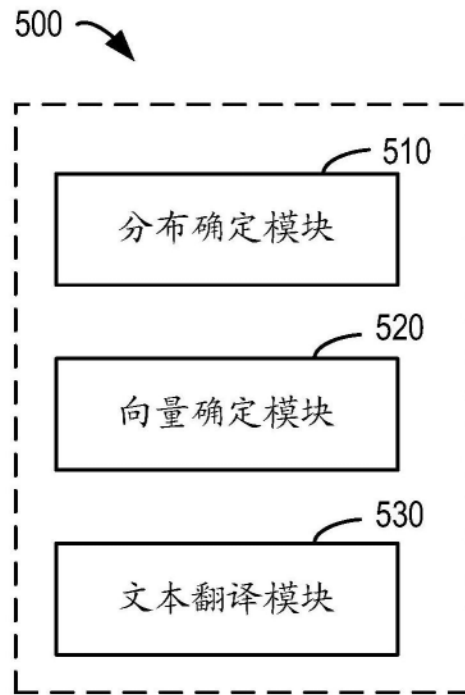


图5

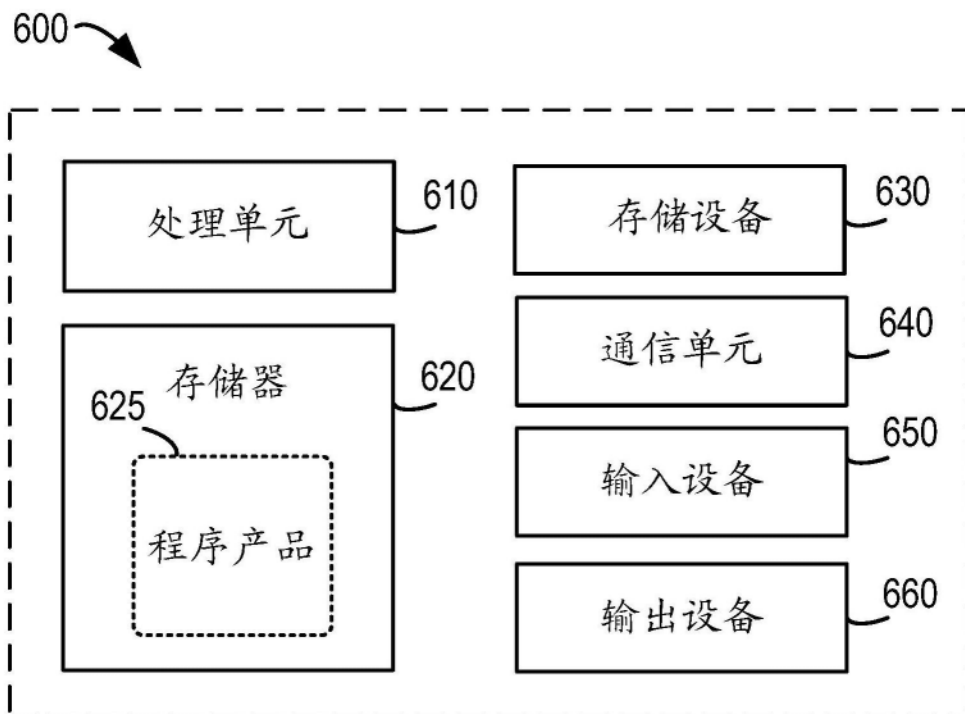


图6