(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 105068998 A
(43)申请公布日 2015.11.18

(21)申请号 201510455270.6

(22)申请日 2015.07.29

(71)申请人 百度在线网络技术(北京)有限公司
地址 100085 北京市海淀区上地十街10号百度大厦三层

(72)发明人 何中军 和为 吴华 王海峰

(74)专利代理机构 北京金律言科知识产权代理事务所(普通合伙) 11461
代理人 罗延红 杨移

(51)Int.Cl.
G06F 17/28(2006.01)
G06N 3/02(2006.01)

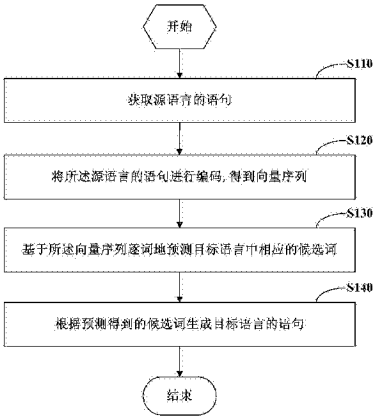
权利要求书2页 说明书8页 附图3页

(54)发明名称

基于神经网络模型的翻译方法及装置

(57)摘要

本发明实施例提供了一种基于神经网络模型的翻译方法及装置。其中,基于神经网络模型的翻译方法包括:获取源语言的语句;将所述源语言的语句进行编码得到向量序列;基于所述向量序列逐词地预测目标语言中相应的候选词;根据预测得到的候选词生成目标语言的语句。本发明实施例的基于神经网络模型的翻译方法及装置,能够结合多种翻译特征执行翻译,提高了翻译结果的翻译质量、流利程度以及可读性。



CN 105068998 A

1.一种基于神经网络模型的翻译方法，其特征在于，所述方法包括：

获取源语言的语句；

将所述源语言的语句进行编码，得到向量序列；

基于所述向量序列逐词地预测目标语言中相应的候选词；

根据预测得到的候选词生成目标语言的语句，

其中，在任一所述候选词的预测处理中，从预设的翻译词表中获取多个初选词，并且基于对数线性模型计算每个所述初选词的翻译概率，以根据所述翻译概率从所述多个初选词选取所述候选词，所述对数线性模型根据神经网络翻译模型的翻译特征和统计机器翻译模型的翻译特征建立的。

2.根据权利要求1所述的方法，其特征在于，所述从预设的翻译词表中获取多个初选词，并且基于对数线性模型计算每个所述初选词的翻译概率，以根据所述翻译概率从所述多个初选词选取所述候选词的处理包括：

从预设的翻译词表中获取M个词语作为初选词，M为正整数，

根据所述向量序列和在获取M个初选词之前预测出的候选词，计算每个所述初选词的神经网络翻译模型的翻译特征，

计算每个所述初选词的统计机器翻译模型的翻译特征，

根据计算得到的神经网络翻译模型的翻译特征和统计机器翻译模型的翻译特征，结合所述对数线性模型计算每个所述初选词的翻译概率值，

对计算得到的M个翻译概率值进行排序，并将与前N位的翻译概率值相应的初选词选取为所述候选词，N为正整数，且N＜M。

3.根据权利要求2所述的方法，其特征在于，所述基于所述向量序列逐词地预测目标语言中相应的候选词的处理还包括：依据所述神经网络翻译模型的翻译特征标记所述候选词之间的关联关系，

所述根据预测得到的候选词生成目标语言的语句的处理包括：

根据具有关联关系的所有候选词生成目标语言的语句。

4.根据权利要求3所述的方法，其特征在于，所述统计机器翻译模型的翻译特征包括双向单词翻译模型特征。

5.根据权利要求4所述的方法，其特征在于，所述统计机器翻译模型的翻译特征还包括N元语言模型特征和/或词惩罚特征。

6.根据权利要求1～5中任一项所述的方法，其特征在于，所述将所述源语言的语句进行编码得到向量序列的处理包括：

对源语言的语句进行分词，

将分得的多个分词分别转换成与每个所述分词对应的词向量，

由多个分词各自对应的词向量得到所述向量序列。

7.根据权利要求6所述的方法，其特征在于，所述获取源语言的语句的处理包括以下处理之一：

接收文本数据，并将所述文本数据作为所述源语言的语句，

接收语音数据，对所述语音数据进行语音识别得到经语音识别的文本数据，并将所述经语音识别的文本数据作为所述源语言的语句，

接收图片数据，对所述图片数据进行光学字符识别 OCR 得到经 OCR 识别的文本数据，并将所述经 OCR 识别的文本数据作为所述源语言的语句。

8. 一种基于神经网络模型的翻译装置，其特征在于，所述装置包括：

语句获取模块，用于获取源语言的语句；

语句编码模块，用于将所述源语言的语句进行编码得到向量序列；

候选词预测模块，用于基于所述向量序列逐词地预测目标语言中相应的候选词；

语句生成模块，用于根据预测得到的候选词生成目标语言的语句，

其中，在任一所述候选词的预测处理中，从预设的翻译词表中获取多个初选词，并且基于对数线性模型计算每个所述初选词的翻译概率，以根据所述翻译概率从所述多个初选词选取所述候选词，所述对数线性模型根据神经网络翻译模型的翻译特征和统计机器翻译模型的翻译特征建立的。

9. 根据权利要求 8 所述的装置，其特征在于，所述候选词预测模块包括：

初选词获取单元，用于从预设的翻译词表中获取 M 个词语作为初选词，M 为正整数，

神经网络模型特征计算单元，用于根据所述向量序列和在获取 M 个初选词之前预测出的候选词，计算每个所述初选词的神经网络翻译模型的翻译特征，

统计机器模型特征计算单元，用于计算每个所述初选词的统计机器翻译模型的翻译特征，

翻译概率计算单元，用于根据计算得到的神经网络翻译模型的翻译特征和统计机器翻译模型的翻译特征，结合所述对数线性模型计算每个所述初选词的翻译概率值，

候选词选取单元，用于对计算得到的 M 个翻译概率值进行排序，并将与前 N 位的翻译概率值相应的初选词选取为所述候选词，N 为正整数，且 $N < M$。

10. 根据权利要求 9 所述的装置，其特征在于，所述候选词预测模块还包括：候选词关联单元，用于依据所述神经网络翻译模型的翻译特征标记所述候选词之间的关联关系，

所述语句生成模块，用于根据具有关联关系的所有候选词生成目标语言的语句。

11. 根据权利要求 10 所述的装置，其特征在于，所述统计机器翻译模型的翻译特征包括双向单词翻译模型特征。

12. 根据权利要求 11 所述的装置，其特征在于，所述统计机器翻译模型的翻译特征还包括 N 元语言模型特征和 / 或词惩罚特征。

13. 根据权利要求 8 ～ 12 中任一项所述的装置，其特征在于，所述语句编码模块用于对源语言的语句进行分词，将分得的多个分词分别转换成与每个所述分词对应的词向量，由多个分词各自对应的词向量得到所述向量序列。

14. 根据权利要求 13 所述的装置，其特征在于，所述语句获取模块包括以下单元之一：

文本数据接收单元，用于接收文本数据，并将所述文本数据作为所述源语言的语句，

语音数据接收及识别单元，用于接收语音数据，对所述语音数据进行语音识别得到经语音识别的文本数据，并将所述经语音识别的文本数据作为所述源语言的语句，

图片数据接收及识别单元，用于接收图片数据，对所述图片数据进行光学字符识别 OCR 得到经 OCR 识别的文本数据，并将所述经 OCR 识别的文本数据作为所述源语言的语句。

基于神经网络模型的翻译方法及装置

技术领域

[0001] 本发明涉及机器翻译技术领域，尤其涉及一种基于神经网络模型的翻译方法及装置。

背景技术

[0002] 近年来，循环神经网络(Recurrent Neural Network，RNN)技术被广泛应用于机器翻译领域。与传统的统计机器翻译系统相比，基于循环神经网络的机器翻译系统能够充分利用全局语义信息，翻译质量显著提升。

[0003] 然而，基于循环神经网络的机器翻译技术也存在着明显的不足之处：

[0004] 一、词表受限。

[0005] 通常，RNN翻译模型仅能使用一个数量有限的单词词表（通常三万单词以内），导致词表外单词(Out-of-vocabulary，OOV)无法翻译。

[0006] 二、RNN翻译模型仅支持双语语句对进行训练，难以利用可有效提高翻译结果流利度的目标语言单语语料进行训练。

[0007] 三、无法增加或丰富更多的特征。

[0008] 具体来说，传统的统计机器翻译方法是基于多种特征执行翻译处理的，对于每一个翻译单元（单词、短语、翻译规则等）均计算每个特征的得分，最终结合多个特征的得分来确定最优翻译结果。然而，这些能够提高翻译质量的特征无法直接为RNN翻译模型所用。

[0009] 四、漏词现象严重，易生成较短翻译结果，影响翻译结果的可读性。

发明内容

[0010] 本发明实施例的目的在于，提供一种基于神经网络模型的翻译方法及装置，结合多种翻译特征执行翻译，以提高翻译结果的翻译准确性。

[0011] 为实现上述发明目的，本发明的实施例提供了一种基于神经网络模型的翻译方法，包括：获取源语言的语句；将所述源语言的语句进行编码，得到向量序列；基于所述向量序列逐词地预测目标语言中相应的候选词；根据预测得到的候选词生成目标语言的语句，其中，在任一所述候选词的预测处理中，从预设的翻译词表中获取多个初选词，并且基于对数线性模型计算每个所述初选词的翻译概率，以根据所述翻译概率从所述多个初选词选取所述候选词，所述对数线性模型根据神经网络翻译模型的翻译特征和统计机器翻译模型的翻译特征建立的。

[0012] 优选地，所述从预设的翻译词表中获取多个初选词，并且基于对数线性模型计算每个所述初选词的翻译概率，以根据所述翻译概率从所述多个初选词选取所述候选词的处理包括：从预设的翻译词表中获取M个词语作为初选词，M为正整数；根据所述向量序列和在获取M个初选词之前预测出的候选词，计算每个所述初选词的神经网络翻译模型的翻译特征；计算每个所述初选词的统计机器翻译模型的翻译特征；根据计算得到的神经网络翻译模型的翻译特征和统计机器翻译模型的翻译特征，结合所述对数线性模型计算每个所述

初选词的翻译概率值；对计算得到的M个翻译概率值进行排序，并将与前N位的翻译概率值相应的初选词选取为所述候选词，N为正整数，且N＜M。

[0013] 优选地，所述基于所述向量序列逐词地预测目标语言中相应的候选词的处理还包括：依据所述神经网络翻译模型的翻译特征标记所述候选词之间的关联关系，所述根据预测得到的候选词生成目标语言的语句的处理包括：根据具有关联关系的所有候选词生成目标语言的语句。

[0014] 优选地，所述统计机器翻译模型的翻译特征包括双向单词翻译模型特征。

[0015] 优选地，所述统计机器翻译模型的翻译特征还包括N元语言模型特征和/或词惩罚特征。

[0016] 优选地，所述将所述源语言的语句进行编码得到向量序列的处理包括：对源语言的语句进行分词，将分得的多个分词分别转换成与每个所述分词对应的词向量，由多个分词各自对应的词向量得到所述向量序列。

[0017] 优选地，所述获取源语言的语句的处理包括以下处理之一：接收文本数据，并将所述文本数据作为所述源语言的语句；接收语音数据，对所述语音数据进行语音识别得到经语音识别的文本数据，并将所述经语音识别的文本数据作为所述源语言的语句；接收图片数据，对所述图片数据进行光学字符识别OCR得到经OCR识别的文本数据，并将所述经OCR识别的文本数据作为所述源语言的语句。

[0018] 本发明的实施例还提供了一种基于神经网络模型的翻译装置，包括：语句获取模块，用于获取源语言的语句；语句编码模块，用于将所述源语言的语句进行编码得到向量序列；候选词预测模块，用于基于所述向量序列逐词地预测目标语言中相应的候选词；语句生成模块，用于根据预测得到的候选词生成目标语言的语句，其中，在任一所述候选词的预测处理中，从预设的翻译词表中获取多个初选词，并且基于对数线性模型计算每个所述初选词的翻译概率，以根据所述翻译概率从所述多个初选词选取所述候选词，所述对数线性模型根据神经网络翻译模型的翻译特征和统计机器翻译模型的翻译特征建立的。

[0019] 优选地，所述候选词预测模块包括：初选词获取单元，用于从预设的翻译词表中获取M个词语作为初选词，M为正整数；神经网络模型特征计算单元，用于根据所述向量序列和在获取M个初选词之前预测出的候选词，计算每个所述初选词的神经网络翻译模型的翻译特征；统计机器模型特征计算单元，用于计算每个所述初选词的统计机器翻译模型的翻译特征；翻译概率计算单元，用于根据计算得到的神经网络翻译模型的翻译特征和统计机器翻译模型的翻译特征，结合所述对数线性模型计算每个所述初选词的翻译概率值；候选词选取单元，用于对计算得到的M个翻译概率值进行排序，并将与前N位的翻译概率值相应的初选词选取为所述候选词，N为正整数，且N＜M。

[0020] 优选地，所述候选词预测模块还包括：候选词关联单元，用于依据所述神经网络翻译模型的翻译特征标记所述候选词之间的关联关系，所述语句生成模块，用于根据具有关联关系的所有候选词生成目标语言的语句。

[0021] 优选地，所述统计机器翻译模型的翻译特征包括双向单词翻译模型特征。

[0022] 优选地，所述统计机器翻译模型的翻译特征还包括N元语言模型特征和/或词惩罚特征。

[0023] 优选地，所述语句编码模块用于对源语言的语句进行分词，将分得的多个分词分

别转换成与每个所述分词对应的词向量，由多个分词各自对应的词向量得到所述向量序列。

[0024] 优选地，所述语句获取模块包括以下单元之一：文本数据接收单元，用于接收文本数据，并将所述文本数据作为所述源语言的语句；语音数据接收及识别单元，用于接收语音数据，对所述语音数据进行语音识别得到经语音识别的文本数据，并将所述经语音识别的文本数据作为所述源语言的语句；图片数据接收及识别单元，用于接收图片数据，对所述图片数据进行光学字符识别 OCR 得到经 OCR 识别的文本数据，并将所述经 OCR 识别的文本数据作为所述源语言的语句。

[0025] 本发明实施例提供的基于神经网络模型的翻译方法及装置，对获取到的源语言的语句进行编码得到向量序列，再基于向量序列逐词地预测目标语言中相应的候选词，在任一所述候选词的预测处理过程中，利用对数线性模型，将统计机器翻译模型的翻译特征引入神经网络翻译模型，从而计算从预设的翻译词表中获取的每个初选词的翻译概率，由此基于翻译概率计算结果从多个初选词中选取候选词，最终根据预测的候选词自动地生成目标语言的语句，极大地提高了翻译的准确性、可读性及流畅性。同时，生成的翻译结果长度适宜，降低了漏词率。

附图说明

[0026] 图 1 是示出本发明实施例一的基于神经网络模型的翻译方法的流程图；

[0027] 图 2 是示出本发明实施例一的基于神经网络模型的翻译方法中神经网络翻译模型的应用场景示例性示意图；

[0028] 图 3 是示出本发明实施例一的基于神经网络模型的翻译方法中 OOV 的预测处理示例性示意图；

[0029] 图 4 是示出本发明实施例二的基于神经网络模型的翻译装置的逻辑框图。

具体实施方式

[0030] 本发明的基本构思是，提供一种结合传统的统计机器翻译方法和 RNN 翻译方法的翻译方式：利用对数线性模型将神经网络翻译模型的翻译特征和统计机器翻译模型的翻译特征相结合，计算从预设的翻译词表中预测出的多个初选词的得分（即翻译概率），从经排序（得分从大到小顺序排列）的多个初选词中选取候选词。

[0031] 然后，根据经上述预测处理得到候选词，自动地生成长度适宜的目标语言语句，解决了 RNN 翻译方法中存在的词表受限、难以利用单语语料进行训练、无法丰富特征、倾向于生成较短翻译结果的问题，显著提升翻译质量，且提高翻译结果的可读性及流利程度。

[0032] 下面结合附图详细描述本发明实施例的基于神经网络模型的翻译方法以及使用所述方法的装置。

[0033] 实施例一

[0034] 图 1 是示出本发明实施例一的基于神经网络模型的翻译方法的流程图。可在实施例二所述的装置上执行所述方法。

[0035] 参照图 1，在步骤 S110，获取源语言的语句。

[0036] 根据本发明的示例性实施例，步骤 S110 包括以下处理之一：

[0037] 接收文本数据，并将所述文本数据作为所述源语言的语句。

[0038] 接收语音数据，对所述语音数据进行语音识别得到经语音识别的文本数据，并将所述经语音识别的文本数据作为所述源语言的语句。

[0039] 接收图片数据，对所述图片数据进行光学字符识别 OCR 得到经 OCR 识别的文本数据，并将所述经 OCR 识别的文本数据作为所述源语言的语句。

[0040] 在步骤 S120，将所述源语言的语句进行编码，得到向量序列。

[0041] 具体地，可以对源语言的语句进行分词，将分得的多个分词分别转换成与每个所述分词对应的词向量，由多个分词各自对应的词向量得到所述向量序列。

[0042] 图 2 是示出本发明实施例一的基于神经网络模型的翻译方法中神经网络翻译模型的应用场景示例性示意图。参照图 2，以获取到的源语言语句“北京的出租车司机很热情”为例，对该语句进行分词处理得到六个分词，图 2 中每个方框内的字或词语代表一个分词，“0.123,0.264…0.465”是经编码处理后得到的向量序列。

[0043] 在步骤 S130，基于所述向量序列逐词地预测目标语言中相应的候选词。其中，在任一所述候选词的预测处理中，从预设的翻译词表中获取多个初选词，并且基于对数线性模型计算每个所述初选词的翻译概率，以根据所述翻译概率从所述多个初选词选取所述候选词，所述对数线性模型根据神经网络翻译模型的翻译特征和统计机器翻译模型的翻译特征建立的。

[0044] 具体地，在本步骤中，所述从预设的翻译词表中获取多个初选词，并且基于对数线性模型计算每个所述初选词的翻译概率，以根据所述翻译概率从所述多个初选词选取所述候选词的处理可包括：从预设的翻译词表中获取 M 个词语作为初选词，M 为正整数；根据所述向量序列和在获取 M 个初选词之前预测出的候选词，计算每个所述初选词的神经网络翻译模型的翻译特征；计算每个所述初选词的统计机器翻译模型的翻译特征；根据计算得到的神经网络翻译模型的翻译特征和统计机器翻译模型的翻译特征，结合所述对数线性模型计算每个所述初选词的翻译概率值。对计算得到的 M 个翻译概率值进行排序，并将与前 N 位的翻译概率值相应的初选词选取为所述候选词，N 为正整数，且 N ＜ M。

[0045] 在具体的实现方式中，仍以图 2 为例进行说明，首先从预设的翻译词表中获取多个词语（the，beijing，taxi 等）作为初选词，根据向量序列（0.123,0.264…0.465）计算每个所述初选词的神经网络翻译模型的翻译特征。这里，由于是预测目标语句中位于第一位的候选词，在此之前没有预测出的候选词，因此计算时无需考虑之前预测出的候选词。

[0046] 然而，在后续的候选词的预测处理中均需考虑之前预测出的候选词。也就是说，预测出目标语言语句中位于第一位的候选词 the，beijing 与 tax 之后，将分别使用这三个候选词预测后面的候选词 taxi，driver 与’s。在预测位于第二位的候选词时，“the taxi”，“the driver”以及“beijing’s”三者得分最高，参与后续的预测处理过程，而位于第一位的候选词“taxi”后续不会再被参考。

[0047] 优选地，可通过以下公式（1）执行所述根据所述向量序列和在获取 M 个初选词之前预测出的候选词，计算每个所述初选词的神经网络翻译模型的翻译特征的处理：

[0048] $$h_{rnn} = \sum_{j=1}^{J} \log\left(p\left(e_j \middle| \{e_{j-1}, \ldots, e_1\}, \vec{f}\right)\right) \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\text{式（1）}$$

[0049] 其中，h_{rnn}为神经网络翻译模型的翻译特征，e_j为目标语言的语句中第 j 个初选词，

e_{j-1},…,e_1为获取初选词 e_j之前预测出的前 j-1 个候选词，$\vec{f}$为所述向量序列。

[0050] 如前所述，在 RNN 翻译方法中存在词表受限的问题。图 3 是示出本发明实施例一的基于神经网络模型的翻译方法中 OOV 的预测处理示例性示意图。参照图 3，单词“坑洞”对于神经网络翻译模型来说是一个 OOV，无法得到准确的译文。

[0051] 在本实施例中，对于 OOV 用“UNK”标记，在后续的处理步骤中，使用预设的单词翻译表对标记为“UNK”的单词进行预测候选词处理。具体地，为了生成 OOV 的准确翻译，首先找到源语言语句中对应的词，图 3 中 $\alpha_{13}=0.1$ 表示“陷在”到“UNK”的单词对齐概率是 0.1，$\alpha_{23}=0.7$ 表示“坑洞”到“UNK”的单词对齐概率是 0.7，$\alpha_{33}=0.2$ 表示“中”到“UNK”的单词对齐概率是 0.2，α_{23}的值最大。由此可知，“UNK”对应源语言语句中的“坑洞”一词，然后，“坑洞”相应的目标语言语句中的候选词可从大规模对齐双语语料库训练出来的单词翻译表中获得，如图 3 所示的 hole、pothole 等。

[0052] 其次，计算每个所述初选词的统计机器翻译模型的翻译特征。为了优化单词互译程度、提高翻译质量，优选地，所述统计机器翻译模型的翻译特征可包括双向单词翻译模型特征。可通过以下公式 (2) 和公式 (3) 执行所述计算每个所述初选词的双向单词翻译模型特征的处理：

[0053] $$h_{tp1}=\sum_{j=1}^{J}\sum_{i=1}^{I}\alpha_{ji}\log\left(p(e_j|f_i)\right)\delta(e_j,f_i)\cdots\cdots\text{式 (2)}$$

[0054] $$h_{tp2}=\sum_{j=1}^{J}\sum_{i=1}^{I}\alpha_{ji}\log\left(p(f_i|e_j)\right)\delta(f_i,e_j)\cdots\cdots\text{式 (3)}$$

[0055] 其中，h_{tp1}为源语言到目标语言的单词翻译模型特征，h_{tp2}为目标语言到源语言的单词翻译模型特征，$p(e_j|f_i)$ 为源语言的单词 f_i到目标语言的初选词 e_j的翻译概率，$p(f_i|e_j)$ 为目标语言的初选词 e_j到源语言的单词 f_i的翻译概率，α_{ji}为源语言的单词 f_i到目标语言的初选词 e_j的单词对齐概率，所述单词对齐概率是基于神经网络翻译模型生成的。$\delta(e_j,f_i)$ 为目标语言的初选词 e_j和源语言的单词 f_i之间的翻译参数，$\delta(e_j,f_i)=1$ 时，目标语言的初选词 e_j和源语言的单词 f_i之间互为翻译，$\delta(e_j,f_i)=0$ 时，目标语言的初选词 e_j和源语言的单词 f_i之间不互为翻译。$\delta(f_i,e_j)$ 为源语言的单词 f_i和目标语言的初选词 e_j之间的翻译参数，$\delta(f_i,e_j)=1$ 时，源语言的单词 f_i和目标语言的初选词 e_j之间互为翻译，$\delta(f_i,e_j)=0$ 时，源语言的单词 f_i和目标语言的初选词 e_j之间不互为翻译。

[0056] 进一步地，所述统计机器翻译模型的翻译特征还可以包括 N 元语言模型特征和 / 或词惩罚特征。由于在 RNN 翻译方法中通过求译文中每个单词的翻译概率的乘积来计算所述单词的得分，而每个单词的翻译概率都是一个小于 1 的值，因此译文长度越长得分越低，反之译文长度越短得分越高，这导致 RNN 翻译系统更倾向于生成短译文。因此，通过词惩罚特征对过短的翻译结果进行惩罚，可以优化翻译结果的长度，生成长度适宜的译文。

[0057] 这里，利用大规模单语语料库通过 N 元语言模型特征优化了翻译结果局部的流利程度。需要说明的是，N 元语言模型中当前词出现的概率仅和其之前的 N-1 个词有关系。例如，当 N 取值为 2 时，当前词的出现概率仅和其前一个词有关系。具体地，可通过以下公式 (4) 执行所述计算每个所述初选词的 N 元语言模型特征的处理：

[0058] $$h_{lm}=\sum_{j=1}^{J}\log\left(p(e_j|e_{j-1},\ldots,e_{j-n+1})\right)\cdots\cdots\text{式 (4)}$$

[0059] 其中，h_{lm}为N元语言模型特征，e_j为目标语言的语句中第j个初选词，e_{j1}，…，e_{jn+1}为获取初选词e_j之前预测出的前n-1个候选词。

[0060] 最后，可通过以下公式(5)执行所述根据计算得到的神经网络翻译模型的翻译特征和统计机器翻译模型的翻译特征，结合所述对数线性模型计算每个所述初选词的翻译概率值的处理：

[0061] $$p(e|f)=\frac{\exp\left(\sum_{i=1}^{m}\lambda_i h_i(f,e)\right)}{\sum_{e'}\exp\left(\sum_{i=1}^{m}\lambda_i h_i(f,e')\right)} \quad \cdots\cdots(5)$$

[0062] 其中，m为翻译特征的个数，$h_i(f,e)$为第i个翻译特征，λ_i为第i个翻译特征对应的特征权重，e为所述初选词，f为源语言的单词，p(e|f)为初选词e的翻译概率。需要说明的是，公式(5)中的e'是求和函数中的临时变量，和求和符号一起使用表示对所有的初选词求和。

[0063] 在具体的实现方式中，依据公式(1)至公式(4)分别计算得到的神经网络翻译模型的翻译特征h_{rnn}、统计机器翻译模型的翻译特征h_{tp1}、h_{tp2}和h_{lm}，将上述多个翻译特征以及各自对应的特征权重代入公式(5)，计算每个所述初选词的得分(即翻译概率值)，就可以采用得分排序方式从多个初选词中选取候选词。这里，翻译特征不同，其对应的特征权重所起的作用也不同。在实际应用中，由于源语言的单词f一旦给定，公式(5)中分母部分的计算结果是一个常量，因此只需计算分子部分即可。

[0064] 在步骤S140，根据预测得到的候选词生成目标语言的语句。

[0065] 为了能够生成可读性及流畅性强的语句，进一步地，步骤S130还可以包括：依据所述神经网络翻译模型的翻译特征标记所述候选词之间的关联关系。仍以图2为例，由于任一所述候选词的预测处理均是基于神经网络翻译模型的，而神经网络翻译模型的重要特点就是考虑上下文的关联关系，使得翻译结果更加流畅。图2示出的候选词之间的箭头线代表了候选词之间的关联关系。例如，位于第二位的候选词taxi与位于第一位的候选词the是相关联的。

[0066] 相应地，根据本发明的示例性实施例，步骤S140包括：根据具有关联关系的所有候选词生成目标语言的语句。例如，参照图2，重复执行前述候选词的预测处理步骤，直至预测到设定的结束词(EOS)为止。图2中使用栈的结构描述候选词及各个候选词之间的关系，第一个栈(Stack-1)中的词语是位于目标语言的语句中第一位的候选词，第i个栈(Stack-i)的词语是位于目标语言的语句中第i位的候选词，依据图2示出的候选词之间的箭头线就可以查找到具有关联关系的所有候选词，从而生成目标语言的语句。

[0067] 本发明实施例提供的基于神经网络模型的翻译方法，在获取源语言的语句之后，对其进行编码得到向量序列，再基于向量序列逐词地预测目标语言中相应的候选词，在任一所述候选词的预测处理过程中，首先从预设的翻译词表中获取多个初选词，基于结合了神经网络翻译模型的翻译特征和统计机器翻译模型的翻译特征的对数线性模型，计算每个所述初选词的翻译概率，根据翻译概率计算结果从多个初选词选取候选词，最终，根据经上述预测处理得到候选词，从而自动地生成长度适宜的目标语言语句，提高了翻译结果的准确度、可读性和流畅性。同时，降低了漏词率。

[0068] 实施例二

[0069] 图4是示出本发明实施例二的基于神经网络模型的翻译装置的逻辑框图。可用于

执行如图 1 所示实施例的方法步骤。

[0070] 参照图 4，所述基于神经网络模型的翻译装置包括语句获取模块 410、语句编码模块 420、候选词预测模块 430 和语句生成模块 440。

[0071] 语句获取模块 410 用于获取源语言的语句。

[0072] 进一步地，所述语句获取模块 410 可包括以下单元之一：

[0073] 文本数据接收单元（未示出）用于接收文本数据，并将所述文本数据作为所述源语言的语句。

[0074] 语音数据接收及识别单元（未示出）用于接收语音数据，对所述语音数据进行语音识别得到经语音识别的文本数据，并将所述经语音识别的文本数据作为所述源语言的语句。

[0075] 图片数据接收及识别单元（未示出）用于接收图片数据，对所述图片数据进行光学字符识别 OCR 得到经 OCR 识别的文本数据，并将所述经 OCR 识别的文本数据作为所述源语言的语句。

[0076] 语句编码模块 420 用于将所述源语言的语句进行编码得到向量序列。

[0077] 具体地，所述语句编码模块 420 用于对源语言的语句进行分词，将分得的多个分词分别转换成与每个所述分词对应的词向量，由多个分词各自对应的词向量得到所述向量序列。

[0078] 候选词预测模块 430 用于基于所述向量序列逐词地预测目标语言中相应的候选词。其中，在任一所述候选词的预测处理中，从预设的翻译词表中获取多个初选词，并且基于对数线性模型计算每个所述初选词的翻译概率，以根据所述翻译概率从所述多个初选词选取所述候选词，所述对数线性模型根据神经网络翻译模型的翻译特征和统计机器翻译模型的翻译特征建立的。

[0079] 根据本发明的示例性实施例，所述候选词预测模块 430 可包括：

[0080] 初选词获取单元（未示出）用于从预设的翻译词表中获取 M 个词语作为初选词，M 为正整数。

[0081] 神经网络模型特征计算单元（未示出）用于根据所述向量序列和在获取 M 个初选词之前预测出的候选词，计算每个所述初选词的神经网络翻译模型的翻译特征。

[0082] 统计机器模型特征计算单元（未示出）用于计算每个所述初选词的统计机器翻译模型的翻译特征。

[0083] 翻译概率计算单元（未示出）用于根据计算得到的神经网络翻译模型的翻译特征和统计机器翻译模型的翻译特征，结合所述对数线性模型计算每个所述初选词的翻译概率值。

[0084] 候选词选取单元（未示出）用于对计算得到的 M 个翻译概率值进行排序，并将与前 N 位的翻译概率值相应的初选词选取为所述候选词，N 为正整数，且 $N < M$。

[0085] 为了优化单词互译程度，提高翻译质量，优选地，所述统计机器翻译模型的翻译特征可包括双向单词翻译模型特征。

[0086] 可选地，所述统计机器翻译模型的翻译特征还包括 N 元语言模型特征和 / 或词惩罚特征。通过 N 元语言模型特征优化了翻译结果局部的流利程度，结合词惩罚特征，对过短的翻译结果进行惩罚，进而优化翻译结果长度。

[0087] 语句生成模块 440 用于根据预测得到的候选词生成目标语言的语句。

[0088] 为了能够生成可读性及流畅性强的语句，进一步地，所述候选词预测模块 430 还可以包括：候选词关联单元（未示出）用于依据所述神经网络翻译模型的翻译特征标记所述候选词之间的关联关系。

[0089] 相应地，所述语句生成模块 440 用于根据具有关联关系的所有候选词生成目标语言的语句。

[0090] 本发明实施例提供的基于神经网络模型的翻译装置，在获取源语言的语句之后，对其进行编码得到向量序列，再基于向量序列逐词地预测目标语言中相应的候选词，在任一所述候选词的预测处理过程中，首先从预设的翻译词表中获取多个初选词，基于结合了神经网络翻译模型的翻译特征和统计机器翻译模型的翻译特征的对数线性模型，计算每个所述初选词的翻译概率，根据翻译概率计算结果从多个初选词选取候选词，最终，根据经上述预测处理得到候选词，从而自动地生成长度适宜的目标语言的语句，显著提升了翻译质量，同时提高了翻译结果的可读性和流畅性，降低了漏词率。

[0091] 在本发明所提供的几个实施例中，应该理解到，所公开的装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，所述模块的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式。

[0092] 另外，在本发明各个实施例中的各功能模块可以集成在一个处理模块中，也可以是各个模块单独物理存在，也可以两个或两个以上模块集成在一个模块中。上述集成的模块既可以采用硬件的形式实现，也可以采用硬件加软件功能模块的形式实现。

[0093] 上述以软件功能模块的形式实现的集成的模块，可以存储在一个计算机可读取存储介质中。上述软件功能模块存储在一个存储介质中，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备等）或处理器（processor）执行本发明各个实施例所述方法的部分步骤。而前述的存储介质包括：U 盘、移动硬盘、只读存储器（Read-Only Memory，ROM）、随机存取存储器（Random Access Memory，RAM）、磁碟或者光盘等各种可以存储程序代码的介质。

[0094] 以上所述，仅为本发明的具体实施方式，但本发明的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本发明揭露的技术范围内，可轻易想到变化或替换，都应涵盖在本发明的保护范围之内。因此，本发明的保护范围应以所述权利要求的保护范围为准。

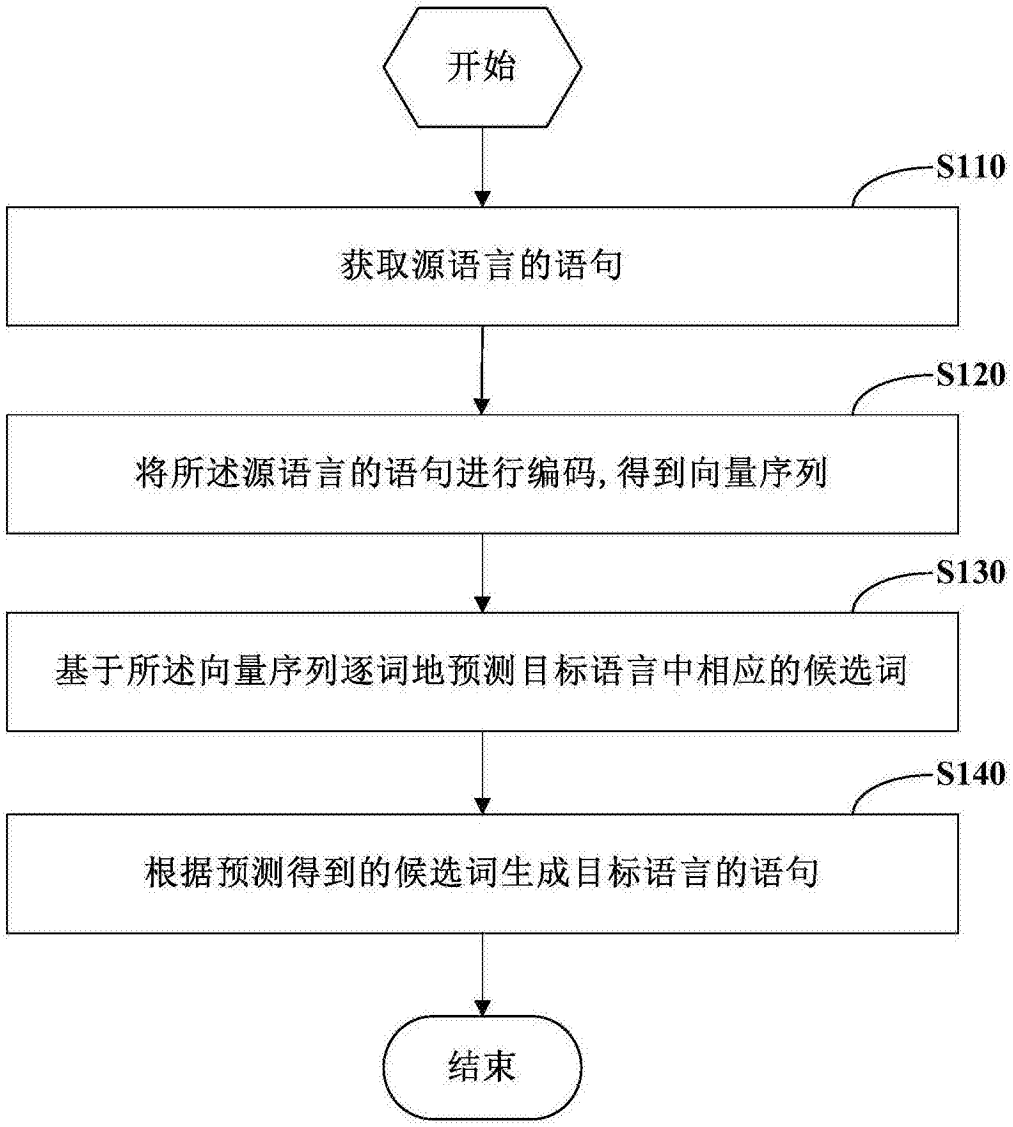

开始
S110
获取源语言的语句
S120
将所述源语言的语句进行编码,得到向量序列
S130
基于所述向量序列逐词地预测目标语言中相应的候选词
S140
根据预测得到的候选词生成目标语言的语句
结束

图 1

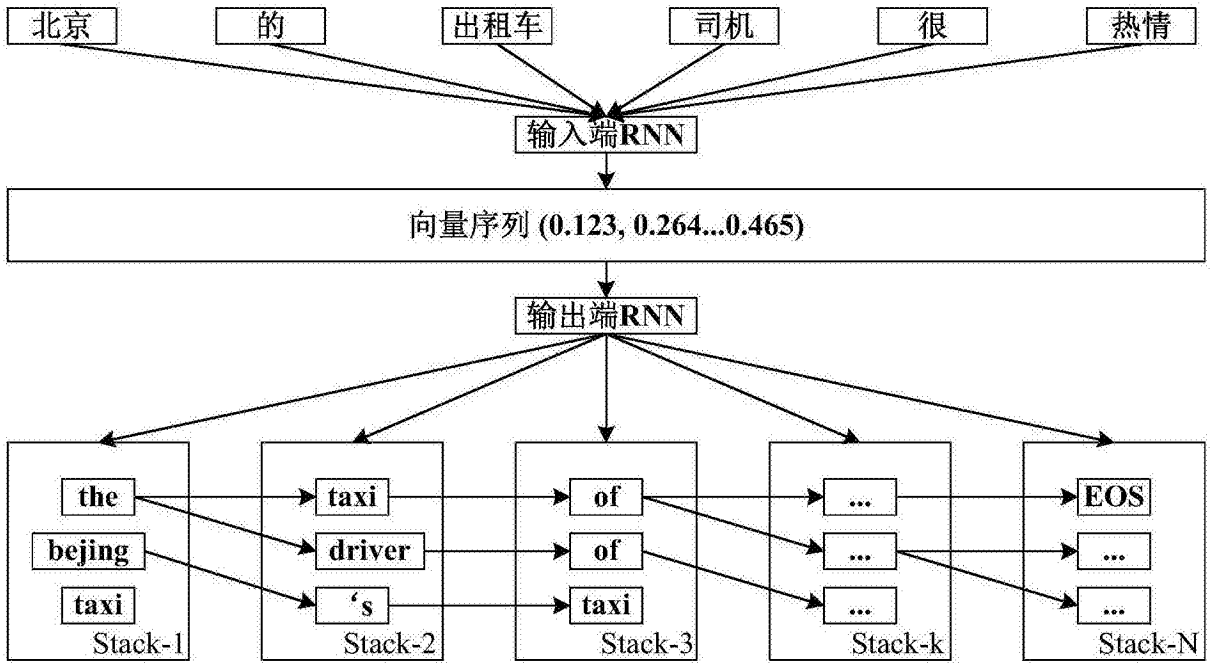

北京
的
出租车
司机
很
热情
输入端RNN
向量序列 (0.123, 0.264...0.465)
输出端RNN
the
bejing
taxi
Stack-1
taxi
driver
‘s
Stack-2
of
of
taxi
Stack-3
...
...
...
Stack-k
EOS
...
...
Stack-N

图 2

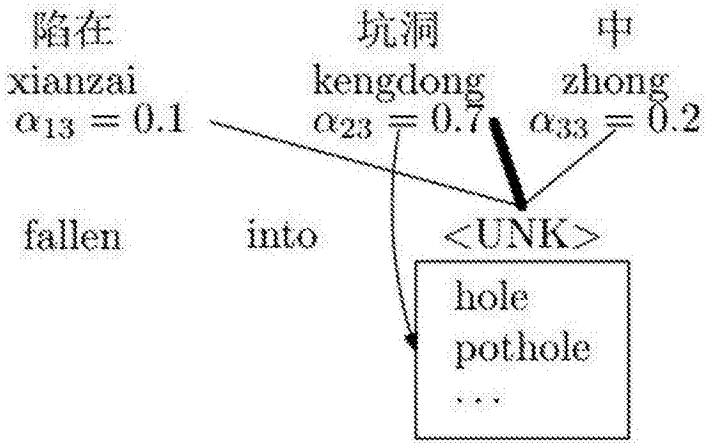

陷在
xianzai
$\alpha_{13} = 0.1$
坑洞
kengdong
$\alpha_{23} = 0.7$
中
zhong
$\alpha_{33} = 0.2$
fallen
into
<UNK>
hole
pothole
...

图 3

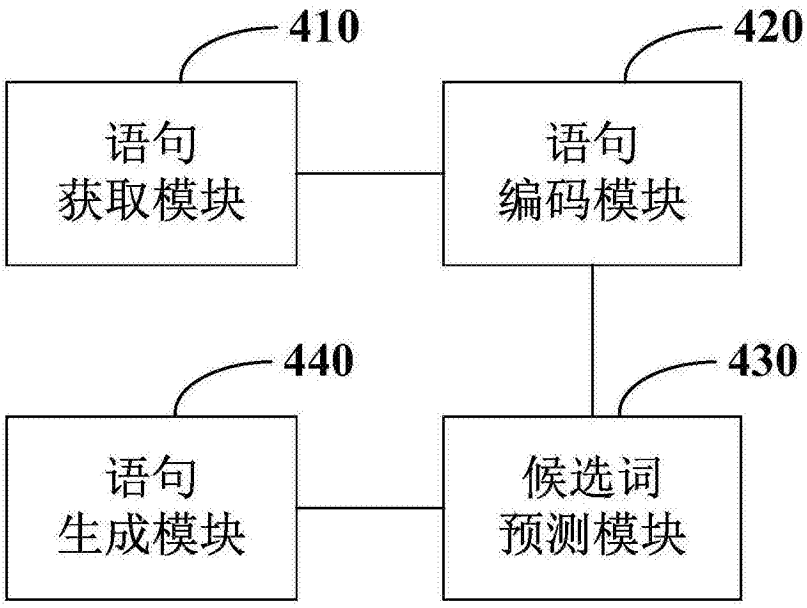


图 4