

文章编号: 1003-0077(2018)06-0036-08

融合先验信息的蒙汉神经网络机器翻译模型

樊文婷, 侯宏旭, 王洪彬, 武 静, 李金廷

(内蒙古大学 计算机学院, 内蒙古 呼和浩特 010021)

摘 要:神经网络机器翻译模型在蒙古文到汉文的翻译任务上取得了很好的效果。神经网络翻译模型仅利用双语语料获得词向量,而有限的双语语料规模却限制了词向量的表示。该文将先验信息融合到神经网络机器翻译中,首先将大规模单语语料训练得到的词向量作为翻译模型的初始词向量,同时在词向量中加入词性特征,从而缓解单词的语法歧义问题。其次,为了降低翻译模型解码器的计算复杂度以及模型的训练时间,通常会限制目标词典大小,这导致大量未登录词的出现。该文利用加入词性特征的词向量计算单词之间的相似度,将未登录词用目标词典中与之最相近的单词替换,以缓解未登录词问题。最终实验显示在蒙古文到汉文的翻译任务上将译文的 BLEU 值提高了 2.68 个 BLEU 点。

关键词: 重现神经网络; 未登录词; 词向量; 词性标注

中图分类号: TP391 **文献标识码:** A

Mongolian-Chinese Neural Machine Translation with Priori Information

FAN Wenting, HOU Hongxu, WANG Hongbin, WU Jing, LI Jinting

(College of Computer Science, Inner Mongolia University, Hohhot, Inner Mongolia 010021, China)

Abstract: Neural machine translation (NMT) has become a prominent model in Mongolian-Chinese translation task. We implement neural machine translation model with priori information. On one hand, we train word representations using large-scale monolingual corpus to act as the initial word vectors. On the other hand, we add part-of-speech feature for word vector to solve the problem of grammatical ambiguity. To solve the out of vocabulary problem, we use word embedding to calculate the similarity of words, then replace the out-of-vocabulary words by the most similar words who are covered by the target vocabulary. In the task of Mongolian-Chinese machine translation, experimental results show that BLEU increased 2.68 points.

Keywords: recurrent neural network; out-of-vocabulary; word embedding; part-of-speech

0 引言

神经网络机器翻译模型被提出之后成为机器翻译的一个研究热点^[1-2]。神经网络机器翻译模型是基于词、短语和句子的连续表示,连续的词向量更能准确地表示词的形态、语义和语法信息,刻画近义词之间的关系^[3-4]。蒙古文的构词形式是在词干后不断缀接词缀来构成词汇,且词形变化丰富,词汇的连续表示更容易捕捉到蒙古文词的形态和构词信息。目前在蒙古文到汉文的机器翻译任务上基于注意力

机制的神经机器翻译模型取得了较好的效果,但仍存在一些不足。

在神经网络翻译模型的框架下,训练翻译模型往往需要大规模的训练语料。模型的训练仅仅依赖双语训练集,网络学习的特征全部依赖双语训练集。但是蒙古文这样的低资源语种并不能获得大规模蒙汉双语平行语料用来训练翻译模型,仅使用双语语料训练翻译模型又学习不到足够多的特征。针对蒙古文—汉文平行语料规模较小的情况,本文将外部特征融合到蒙古文—汉文神经网络翻译模型的训练过程中,来指导网络训练,缓解蒙古文—汉文翻译中因

收稿日期: 2017-07-26 定稿日期: 2017-10-18

基金项目: 国家自然科学基金(61362028)

训练语料不足导致网络学习到的特征不足的问题。为此,我们将先验信息融合到神经网络的训练中。

首先,词向量表示的精确度对网络的翻译性能有很大影响,神经网络翻译模型以 one-hot 为初始词向量,且初始词向量对模型有较大的影响。在网络训练过程中,词向量表示不断调整。词向量的学习与翻译模型的训练同时进行,然而双语平行语料的获得要比单语语料更为困难,有限的双语语料限制了词向量对词本身所包含信息的学习。大规模的单语训练语料有助于更精准的词向量表示。蒙古文属于低资源语言,可利用的双语句对齐语料较少,与之相比汉文的单语语料则较为丰富。本文利用大规模的汉文单语语料来训练得到较为准确的汉文词向量表示。其次,蒙古文到汉文的翻译任务中可利用的双语平行语料较少,网络自动学习到的特征较少。针对上述神经网络机器翻译模型在蒙古文到汉文的翻译任务上存在的问题,本文提出将可利用的先验信息融合到蒙古文到汉文的神经网络翻译模型中。第一,使用大规模单语语料预先训练词向量,用得到的词向量作为神经网络机器翻译模型的初始词向量。这样词向量的训练不仅限于有限双语平行语料,还可以利用大规模的单语语料,为翻译模型提供语义表达更精准的词向量表示。第二,神经网络机器翻译模型训练时仅使用语料文本信息,并没有使用外部的标注特征。本文采用有效的外部特征来增加蒙古文到汉文翻译任务中的可用信息。这些丰富的先验特征更有利于网络学习到语言特征,充分发挥神经网络特征学习的优势。本文提出在词向量中融入词性特征,可以起到消除歧义的作用。

对于神经网络翻译模型来说,目标语言词典的大小对翻译结果的影响较大。如果目标语言词典包含训练集中的所有词,翻译效果会有很大程度提升,但训练时间和规模会相应增加。因此,翻译中通常会限制目标词典大小,这就导致翻译中未被词典覆盖的词会增加。在翻译过程中,通常使用未知词 (UNK) 来代替所有的未登录词。神经网络的主要优势是可以利用长距离依赖信息,而词的连续表示是在翻译模型的训练过程中学习而成,所以未登录词的出现会影响整个句子的翻译以及词向量的学习。本文采用的方法是在词典规模不变的情况下,让词典包含更多目标语言单词信息、提高词典利用率。该方法利用加入词性特征的词向量计算词与词之间的相似度,然后将未登录词用目标词典中与之最相近的词替代。

为了证明本文提出的融合先验信息的神经网络翻译模型的有效性,在蒙古文到汉文翻译任务上进行了实验比较,使用基于注意力机制的重现神经网络作为基线系统。采用大规模单语蒙古文以及汉文语料预先训练词向量方法的实验结果比基线系统提高了 1.98 个 BLEU 点;此外,在预先训练得到的词向量中融入词性特征的实验结果比未加入词性特征的实验结果提高 0.7 个 BLEU 点;相同参数下,利用加入词性特征的词向量计算相似度做低频词替换实验结果比基线系统提高 0.6 个 BLEU 点;相比基线系统,本文将蒙古文—汉文的翻译结果提高 2.68 个 BLEU 点。

1 相关研究

1.1 基于注意力机制的神经网络翻译模型

目前翻译效果较好的是基于注意力机制的重现神经网络编码器—解码器翻译模型,在文献[5-7]中有详细介绍。本文使用含有记忆单元的注意力机制的重现神经网络^[8]。基于重现神经网络编码器—解码器的机器翻译模型是一种端到端翻译模型。它由编码器和解码器两部分构成。源语言句子 $X = (x_1, \dots, x_M)$ 作为编码器的输入, (h_1, \dots, h_M) 表示编码器的输出。解码器利用编码器的输出和注意力机制产生的对齐权重产生向量序列 (c_1, \dots, c_M) , 从而生成目标语言句子。目标语言句子 $Y = (y_1, \dots, y_T)$ 的联合翻译概率定义为:

$$p(Y | X) = \prod_{t=1}^T p(y_t | y_{<t}, c_t) \quad (1)$$

每一个目标单词的翻译概率为:

$$p(y_t | y_{<t}, c_t) = \text{softmax}(g(y_{t-1}, s_t, c_t)) \quad (2)$$

其中, c_t 是编码层输出 (h_1, \dots, h_M) 的加权和:

$$c_t = \sum_i^M \alpha_{ti} h_i \quad (3)$$

M 表示源语言句子长度, T 表示目标语言句子长度, g 是非线性激活函数。 s_t 表示解码器隐层的输出, c_t 由编码器的输出和注意力机制的对齐权重得到, $\{y_1, \dots, y_{t-1}\}$ 表示已经翻译出来的目标语言单词, 权值 α_{ti} 由编码器的输出和解码器上一步的隐层输出计算得到。

1.2 受限的目标词典大小

神经网络机器翻译模型预测当前目标单词时, 会计算目标语言词典中每一个单词作为当前预测单

词的概率,目标词典的大小就是网络的搜索空间。为了减小网络训练时间,通常会限制目标语言词典大小。这样会导致训练语料中的低频词不会被词典包含从而变为未登录词。从式(2)可看出,计算目标单词的预测概率时,会依赖前面已经预测出来的单词。若译文的前一个单词是 UNK 则会影响后面目标单词的预测,UNK 不仅会影响预测下一个词的准确率,还会影响整个网络的训练。

2 融合先验信息的神经网络翻译模型

本节介绍将先验信息融合到神经网络翻译模型中的具体方法。首先,将大规模单语语料预先训练得到的词向量作为神经网络机器翻译模型的初始词向量。其次,在词向量中融入词性特征,可以起到消除语义、词义上的歧义的作用。针对神经网络的未登录词问题,利用加入词性特征的词向量计算词与词之间的相似度,然后将未登录词用目标词典中与之最相近的词替代。

2.1 词向量预训练

神经网络机器翻译模型中词向量学习仅使用有限双语平行语料,且词向量的初始值采用随机赋值。训练词向量时语料的规模对词向量准确性影响较大。本文采用大规模的单语语料训练得到的词向量可以更准确地表示词汇本身携带的信息,将得到的词向量作为神经网络机器翻译模型的初始词向量。第四节的实验结果表明该方法的有效性。

自然语言相关任务中最常见的第一步是创建一个词典并把每个词顺序编号,也就是词向量表示方法中的 one-hot 表示。向量的维度等于词典的大小,这种方法无法捕捉词与词之间的相似度,且容易发生维数灾难。最早由 Hinton 在 1986 年提出一种分布式词向量表示方法,该方法可以利用词向量较为精确的表示单词的词义。2013 年 Mikolov Tomas 提出两种学习单词连续向量表示的新模型^[9-11]。2016 年 Facebook 提出一种学习连续词向量的方法^[12]。本文词向量预先训练采用的是 Facebook 提出的连续词向量学习的方法。图 1 表示词向量训练模型。图中 W 是给定的词典,假设存在一个 w_1, w_2, \dots, w_T 的词组序列,则最大化公式如式(4)所示。

$$\sum_{t=1}^T \sum_{c \in C_t} \log p(w_c | w_t) \quad (4)$$

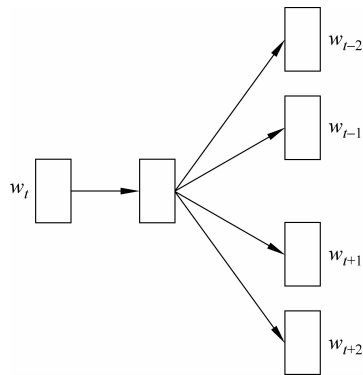


图 1 词向量训练模型

C_t 是 w_t 的上下文, $|C_t|$ 是上下文窗口大小, $|C_t|$ 越大则考虑的上下文越多,词向量的表示相对更准确。同时,训练时间也会增加,每个词向量表征了单词的上下文分布。 s 是计算得分函数, $s(w, c)$ 表示已知单词 w 预测下一个单词 c 的得分。根据得分定义上下文单词预测概率如式(5)所示。

$$p(w_c | w_t) = \frac{es(w_t, w_c)}{\sum_{j=1}^w es(w_t, j)} \quad (5)$$

得分函数 s 定义为:

$$s(w, c) = \sum_{g \in G_w} z_g^T v_c \quad (6)$$

z_g 是 w 的 n -gram 片段的向量表示, v_c 是 c 的向量表示。集合 G_w 表示给定一个单词 w 的 n -gram 片段集合。

2.2 融合词性特征

近年来,研究者将词的形态信息或主题信息融入到词向量表示中,来丰富词向量的信息。本文在神经网络翻译模型中融入词性特征来指导网络的训练,从而得到更准确的翻译结果。

自然语言中词含有多个词义,在各个层面上存在歧义。语法层面上,一个词可以有多种词性,语义层面上,一个词有多个词义^[13-15]。根据上下文关系得到的词性标注特征可以消除句子中词的语法兼类现象,有助于在特定句子中选择正确的词义。通常可以在语言模型或翻译模型中加入外部特征信息^[16-17],起到消除歧义的作用。神经网络翻译模型仅使用词本身作为学习信息,事实上更多的先验信息可以指导神经网络机器翻译模型的学习,从而提高模型对语境知识的学习能力。本文提出一种方法将汉文的词性标注信息融入到神经网络机器翻译训练过程中。首先,在预先训练得到的词向量中融入

词性特征。然后,把得到的词向量作为神经网络机器翻译模型的初始词向量。利用词向量将词性特征带入到网络训练中,具体融合方法如图 2 所示。首先,得到词性对应的词性向量。之后,将其与词向量进行融合。

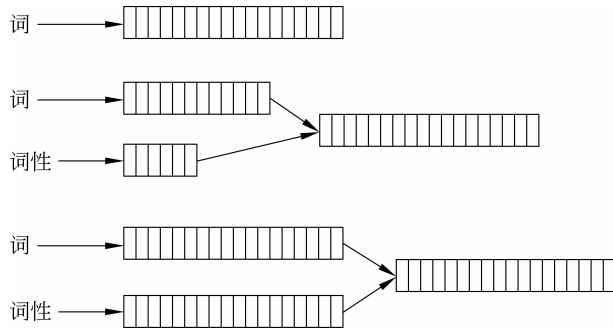


图 2 词性向量与词向量融合方法

本文采用的词性特征融合方法有两种:第一,训练维度相同的词向量和词性向量,根据汉文语料的词性标注信息,将两个向量直接相加得到最终词向量。作为网络的初始向量。第二,使用向量拼接方法训练维度不同的词向量和词性向量,将得到的两个向量直接拼接作为最终词向量,两个向量的维度之和是最终词向量的维度。针对拼接方法,词性向量维度与词向量维度的比例会影响加入词性的翻译效果。如果词性向量维度过低,则带入的信息过少对翻译结果的提升不明显。如果词性向量维度过高,则会导致词向量携带词本身信息被削弱。在本文的实验部分对词向量与词性向量不同拼接维度进行了分析。

2.3 低频词替换解决未登录词问题

未登录词问题是神经网络机器翻译研究的重要组成部分。2015 年, Jean S 提出一种在使用大规模目标端词典的前提下,提高翻译结果且不增加训练时间的方法^[18]。该方法使用子采样技术,把总的训练语料分成多个集合,分别为其建立目标端词典。Quoc V. Le 提出了一种后处理的方法来缓解未登录词问题^[19],增加对齐算法,利用对齐将未被翻译出来的目标词用与之对应的源语言单词替换。Rohan Chitnis 提出了利用哈夫曼编码的方法编码目标语言词典从而使词典覆盖更多的目标端单词^[20]。本文采用在词典规模不变的情况下,让词典包含更多目标语言单词信息,提高词典利用率。该方法利用加入词性特征的词向量计算词与词之间的相似度,

然后将未登录词用目标词典中与之最相近的词替代,减少了未登录词的数量,且不会增加网络的训练时间。

具体方法是:首先得到训练语料中出现的所有词的融合词性特征的分布式词向量,再根据词典大小把所有的词划分为两部分。一部分是被词典覆盖的词,另一部分是未被词典覆盖的词。前者称为高频词,后者称为低频词。其次,利用词向量计算低频词与高频词的相似度。最后,根据相似度替换训练语料中的低频词。具体的操作流程如图 3 所示。

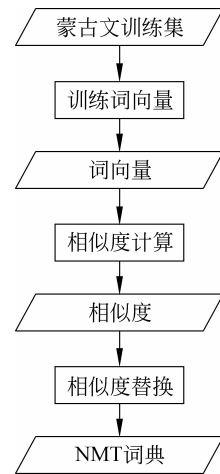


图 3 低频词替换方法流程图

低频词 w_1 和低频词 w_2 的词向量可以分别表示为 $w_1 = (x_{11}, x_{12}, \dots, x_{1N})$, $w_2 = (x_{21}, x_{22}, \dots, x_{2N})$, 其中 N 表示词向量的维度。单词之间相似度的计算为:

$$\text{sim}(w_1, w_2) = \frac{\sum_{k=1}^n x_{1k} x_{2k}}{\sqrt{\sum_{k=1}^n x_{1k}^2} \sqrt{\sum_{k=1}^n x_{2k}^2}} \quad (7)$$

如果训练语料中的单词 w_i 不在目标语言词典中,则在目标语言词典中搜索与之最为相近的词替换:

$$i = \arg \min_{k \in |V|, v_k \in V} \{ \text{sim}(w_i, v_k) \} \quad (8)$$

$$S_{\text{train}} = \text{replace}(w_i, v_i) \quad (9)$$

$|V|$ 是目标语言词典的大小, v_i 是目标词典中的单词。低频词替换的过程中会引入一些噪声,所以在替换的过程中加入一些限制条件,以免引入过多的噪声。首先,每个高频词只可以被用于一次替换。如果多次使用同一个高频词替换,则会影响高频词本身词义的表达。其次,为了减少噪声的引入我们不对所有低频词进行替换。

3 实验分析

3.1 实验数据及参数配置

本文在蒙古文到汉文的翻译任务上验证所提出方法的有效性。使用 CWMT2009 蒙汉双语平行语料作为训练数据集,双语平行语料中选择长度为 2~50 的汉文和蒙古文句子。蒙古文的形态多变低频词较多。传统的基于统计的蒙古文到汉文翻译方法中,通常采用基于词典、规则和统计相结合的蒙古文词切分方法,从而缓解蒙古文的数据稀疏问题^[21-22]。然而神经网络机器翻译模型优势主要体现在对于特征的学习能力,数据稀疏并不是影响其效果的最主要问题。在本文的实验中不对蒙古文做词缀切分或词缀提取等预处理工作。中文语料处理采用 Zhang 等人提出的 ICTCLAS 中文分词系统对中文语料按词切分^[23]。本文采用 BLEU 值作为翻译结果的评测^[24]指标,为了表明方法的有效性,本文所有实验结果都使用均值,而不是最高值,实验数据集具体信息如表 1 所示。

表 1 实验语料统计信息

数据集	训练集	开发集	测试集
双语句对	65 752	1 000	1 000
蒙古文总词数	571 075	7 371	7 555
汉文总词数	591 512	8 670	8 792
汉文总字数	722 099	10 484	10 556

基线系统是基于注意力机制的重现神经网络的编码器—解码器翻译模型^[5]。系统参数配置如下:词向量维度 500,隐层单元数量 500,批大小 32,优化算法采用 adadelta。

3.2 融合词性特征对比实验

使用大规模单语语料来获得更准确的词向量,本文使用 Fasttext 开源工具获得词向量表示。使用大规模单语语料训练得到融合词性特征的词向量,将其作为神经网络机器翻译模型的初始词向量。实验结果如表 2 所示。实验中汉文按词切分,词向量预先训练的翻译结果比基线系统的 BLEU 值提高了 1.98 个点。且使用预先训练词向量作为翻译模型的初始词向量后,模型训练时间比基线系统更短。主要原因是,首先预先训练词向量使用了大规模单语语料从而获得更准确的词向量表示,也可以

得到更好的翻译结果。其次与随机初始化词向量比较,经过训练的词向量有利于模型初始训练对单词的选择,也可以减小翻译模型的初始错误率。

表 2 预训练词向量对比实验结果

系 统	目标词典大小	词向量维度	BLEU
基线	22 270	500	29.22
基线+预训练	22 270	500	31.20

为了证明融合词性标注信息到神经网络中的有效性,本文对汉文做分词处理后使用斯坦福开源标注工具 Stanford-postagger 获取汉文词性标注。目标词典包含所有训练集中的单词,词典大小为 22 270。从表 3 中的实验结果可以看出,本文提出的两种词性特征融合的方法对翻译结果都有明显提高,向量相加的方法相对更好一些。词性向量与词向量直接相加的方法比仅使用词向量的 BLEU 值提高了 0.7 个点,拼接方法将 BLEU 提高了 0.62 个点。其中词性维度不同,拼接效果也有差异,词性向量为 50 维的翻译结果比 100 维的效果更好。主要原因是词性向量维度过高时会导致原本的词向量维度变低,以至于词向量没有携带更多的词信息,但是相比于未加词性特征的翻译结果也有一定提高。

表 3 词向量中融入词性实验结果

实验系统	融合方法	词向量维度	BLEU
基线+预训练	—	500	31.20
基线+预训练+词性	拼接	400+100	31.66
基线+预训练+词性	拼接	450+50	31.82
基线+预训练+词性	相加	500	31.90

表 4 是不同系统的翻译结果,从翻译结果中可以很明显看到翻译质量相比于基线系统都有明显的改善。相比于基线系统的翻译结果,加入词向量预先训练的翻译结果的语义更为完整。而加入词性特征的翻译结果比仅利用词本身词向量的翻译结果更为流畅,更符合汉语的语言习惯。

表 4 翻译结果比较

参考译文	1. 该建筑计划于五月份完工。
	2. 还能参加城市游吗?
	3. 你认识草原新牧民永红吗?
基线	1. 该工程计划于五月呼和浩特。
	2. 还有市区旅游吗?
	3. 您认识草原上生产牧民吗?

续表	
基线+预训练	1. 该建筑物的计划在 6 月结束。
	2. 还能参加这个城市之旅吗?
	3. 你认识草原新的新牧民。
基线+预训练+标注	1. 该工程计划将于 6 月完成。
	2. 还能参加城市之旅吗?
	3. 你认识草原新牧民吗?

3.3 低频词替换对比实验

由于模型训练时目标词典大小的限制,原本在训练语料中出现的单词可能变为未登录词。表 5 是第三部分针对是否在算法中引入限制条件的实验结果。从表中可以看到采用低频词替换方法,但是未加限制条件的翻译评测结果与基线系统的评测结果差异很小。主要原因是由于不加限制条件,该方法引入的噪声太大,抵消了低频词替换的积极作用。当加入限制条件后翻译效果则会有明显提升,相比于基线系统 BLEU 值提高了 0.6 个点。实验结果证明在相同目标词典大小的情况下,本文使用的低频词替换方法在没有增加模型训练时间的前提下,有效的缓解了未登录词问题,提高了翻译译文的质量。

表 5 加入限制条件的实验结果

系统	词典大小	限制条件	BLEU
基线	12 850	—	24.54
基线+方法 1	12 850	否	24.70
基线+方法 1	12 850	是	25.14

本文采用词向量相似度将低频词替换为高频词。从表 6 可以看出,替换过程中会出现词义不相近的替换实例。例如,“私心”与“核心”作为近义词替换是不正确的。但是从替换后的句子来看,相比于直接将句子中的低频词“私心”去掉,将其替换为“核心”可以保证句子结构的完整性。表 7 显示,目标端词典大小相同时,使用高频词替换替低频词方法后的翻译结果要优于直接从词典中去掉低频词的翻译结果。虽然低频词与高频词替换过程中会出现词义不一致的替换,但从翻译结果来看与该方法引入的噪声相比,其积极作用要更大,有效地提高了基于神经网络机器翻译模型从蒙古文到汉文的翻译质量。对于低频词与高频词替换过程中出现词义不一

致的替换问题,会在之后的工作中继续研究,进而提高替换准确率及翻译质量。

表 6 低频词高频词替换实例

低频词→高频词	替 换 实 例
绚烂→灿烂	这里的日落太绚烂了→这里的日落太灿烂了
紊乱→捣乱	心绪紊乱→心绪捣乱
私心→核心	我有私心那就冤枉我了→我有核心那就冤枉我了

表 7 低频词替换实验结果

系统	词典大小	BLEU/%
基线	12 850	24.70
基线+低频词替换	12 850	25.14

低频词与高频词替换比例会影响翻译结果。低频词替换比例相对较高,会影响高频词原本的词义。低频词替换比例较低则对翻译效果影响不明显。经过一组实验找到更合适的低频词与高频词的比例,分别替换训练语料中全部词频小于等于 2、小于等于 1 和部分词频小于等于 1 的词。表 8 的对比实验结果证明,当替换全部小于等于 2 的低频词时翻译结果与相同目标词典大小的基线系统没有差异。主要原因是采用低频词替换方法虽然缓解了未登录词问题,但替换比例太大对高频词原有词义有消极影响,引入的噪声会过大。替换小于等于 1 频词的翻译结果与相同目标词典大小的基线系统相比 BLEU 值提高了 0.6 个点。采用替换部分 1 频词的翻译结果与相同目标词典大小的基线系统相比有一定提高。虽提高有限,但译文质量达到了包含全部训练语料中单词的译文质量。

表 8 低频词替换实验结果对比

系统	目标词典	未登录词	BLEU
基线	10 000	12 270	24.30
基线+方法 1	10 000	12 270	24.25
基线	12 850	9 420	24.54
基线+方法 1	12 850	9 420	25.14
基线	22 270	0	29.22
基线+方法 1	15 000	7 270	29.36

4 总结

本文主要针对基于神经网络翻译模型在蒙古文到汉文翻译任务上的不足提出相应解决方法, 主要将先验信息融合到蒙古文到汉文的神经网络中指导翻译模型训练。使用大规模单语语料预先训练词向量来提高词向量的精确度。为了消除句子中单词的语法兼类等歧义问题, 将汉文训练语料的词性特征加入到神经网络机器翻译模型训练过程中。神经网络机器翻译模型在训练中限制了目标语言词典大小, 导致原本在训练语料中出现过的词变为未登录词。本文提出通过在相同词典规模下提高词典的利用率, 从而有效缓解因为目标词典规模大小限制带来的未登录词问题。最终将蒙古文到汉文的译文 BLEU 值提高了 2.68 个点。本文提出将蒙古文—汉文的外部信息融合到神经网络翻译模型的训练中, 而对于模型本身的改进较少, 在之后的研究中将实现对神经网络模型本身的改进。在未来的研究中, 主要着重在蒙古文到汉文的翻译任务上把已有文本的先验信息融入到网络训练过程中, 用以指导网络训练。例如在网络训练过程中融入句法特征来指导网络学习句子内部的句法结构信息。

参考文献

- [1] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks[C]//Proceedings of Advances in neural information processing systems, 2014; 3104-3112.
- [2] Kalchbrenner N, Blunsom P. Recurrent Continuous Translation Models [C]//Proceedings of EMNLP, 2013; 3(39): 413.
- [3] Collobert R, Weston J. A unified architecture for natural language processing: Deep neural networks with multitask learning[C]//Proceedings of the 25th International Conference on Machine learning. ACM, 2008; 160-167.
- [4] 郑亚楠, 珠杰. 基于词向量的藏文词性标注方法研究[J]. 中文信息学报, 2017, 31(1): 112-117.
- [5] Cho K, Van Merriënboer B, Bahdanau D, et al. On the properties of neural machine translation: Encoder-decoder approaches [J]. arXiv preprint arXiv, 2014, 1409.1259.
- [6] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv, 2014, 1409.0473.
- [7] Wu Y, Schuster M, Chen Z, et al. Google's neural machine translation system: Bridging the gap between human and machine translation[J]. arXiv preprint arXiv: 2016.1609.08144.
- [8] Gers, Felix A, Jürgen Schmidhuber. Learning to forget: Continual prediction with LSTM[J]. Neural Computation, 2000, 12(10): 2451-2471.
- [9] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv, 2013, 1301.3781.
- [10] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]//Proceedings of Advances in neural information processing systems, 2013; 3111-3119.
- [11] Mikolov T, Yih W, Zweig G. Linguistic regularities in continuous space word representations[C]//Proceedings of HLT-NAACL. 2013; 746-751.
- [12] Bojanowski P, Grave E, Joulin A, et al. Enriching word vectors with subword information [J]. arXiv preprint arXiv: 2016.1607.04606.
- [13] 白栓虎, 夏莹, 黄昌宁. 汉语语料库词性标注方法研究[J]. 机器翻译研究进展, 1992; 408-418.
- [14] 丁伟伟, 常宝宝. 基于语义组块分析的汉语语义角色标注[J]. 中文信息学报, 2009, 23(5): 53-61.
- [15] 张仰森, 黄改娟, 苏文杰. 基于隐最大熵原理的汉语词义消歧方法[J]. 中文信息学报, 2012, 26(3): 72-78.
- [16] Chen W, Matusov E, Khadivi S, et al. Guided alignment training for topic-aware neural machine translation[J]. arXiv preprint arXiv: 2016.1607.01628.
- [17] Sennrich R, Haddow B. Linguistic input features improve neural machine translation[J]. arXiv preprint arXiv: 2016.1606.02892.
- [18] Jean S, Cho K, Memisevic R, et al. On using very large target vocabulary for neural machine translation [J]. arXiv preprint arXiv: 2014.1412-2007.
- [19] Luong M T, Sutskever I, Le Q V, et al. Addressing the rare word problem in neural machine translation [J]. arXiv preprint arXiv: 2014.1410.8206.
- [20] Chitnis R, DeNero J. Variable-Length Word Encodings for Neural Translation Models[C]// Proceedings of EMNLP. 2015; 2088-2093.
- [21] 那顺乌日图. 蒙古文词根、词干、词尾的自动切分系统[J]. 内蒙古大学学报(人文社会科学版), 1997, 02: 53-57.
- [22] 史建国, 侯宏旭, 飞龙. 基于词典、规则的斯拉夫蒙古文词切分系统的研究[J]. 中文信息学报, 2015, 29(1): 197-202.
- [23] Zhang R, Yasuda K, Sumita E. Improved statistical machine translation by multiple Chinese word segmentation[C]//Proceedings of the Third Workshop on Statistical Machine Translation. Association for

Computational Linguistics Ohio: Association for Computational Linguistics, 2008, 216-223.

[24] 张剑, 吴际, 周明. 机器翻译评测的新进展[J]. 中文信息学报, 2003, 17(6): 1-8.



樊文婷(1992—), 硕士研究生, 主要研究领域为自然语言处理。

E-mail: 1583679655@qq.com



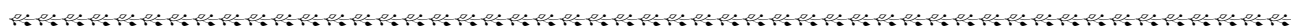
侯宏旭(1972—), 通信作者, 博士, 教授, 博士生导师, 主要研究领域为自然语言处理、信息检索。

E-mail: cshhx@imu.edu.cn



王洪彬(1989—), 硕士研究生, 主要研究领域为自然语言处理。

E-mail: 929004459@qq.com



(上接第 35 页)

[22] Feng M, Xiang B, Glass M R, et al. Applying deep learning to answer selection: A study and an open task[C]//Proceedings of, 2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU). IEEE, 2015: 813-820.

[23] 史顺良, 任育新. 语言学类学术文章标题的结构及其语用功能: 调查与分析[J]. 外语教学, 2010, 31(4): 28-28.

[24] 曹杨, 赵硕. 科技论文标题的结构和语言特征——以和为例[J]. 外语教学, 2014, 35(2): 35-39.

[25] 姜亚军. 复合式论文标题之争: 评述与启示[J]. 外语教学, 2010, 31(6): 29-32.

[26] Berant J, Srikumar V, et al. Modeling biological processes for reading comprehension [C]//Proceedings of Empirical Methods in Natural Language Processing, 2014: 1499-1510.

[27] 王智强, 李茹, 梁吉业等. 基于汉语篇章框架语义分析的阅读理解问答研究[J]. 计算机学报, 2016, 39(4): 795-807.



关勇(1990—), 硕士, 主要研究领域为自然语言处理。

E-mail: guanyong0130@163.com



吕国英(1964—), 硕士, 副教授, 主要研究领域为自然语言处理。

E-mail: english@sxu.edu.cn



李茹(1963—), 博士, 教授, 主要研究领域为中文信息处理、数据智能预测。

E-mail: liru@sxu.edu.cn