基于参数迁移的蒙汉神经机器翻译模型

王宇飞　苏依拉*　赵亚平　孙晓骞　仁庆道尔吉
（内蒙古工业大学信息工程学院　内蒙古 呼和浩特 010080）

摘　要　目前在神经网络机器翻译的研究中，大规模的语料库数据是得到理想译文效果的必备条件之一。对于英语、德语等大语种而言，有着丰富的资源和数据供研究人员使用，但是对于很多小语种如蒙古语，其平行语料库数据较少。为了缓解该问题，把参数迁移引入到蒙汉神经机器翻译中，将训练好的英汉神经机器翻译中的模型参数迁移到蒙汉翻译任务中。实验结果表明，迁移学习策略的引入对蒙汉翻译结果有所提升。

关键词　神经网络　机器翻译　参数迁移　蒙汉翻译

中图分类号　TP3　　**文献标志码**　A　　DOI：10.3969/j.issn.1000-386x.2020.09.014

MONGOLIAN-CHINESE NEURAL MACHINE TRANSLATION MODEL BASED ON PARAMETER TRANSFER

Wang Yufei　Su Yila*　Zhao Yaping　Sun Xiaoqian　Ren Qingdaoerji
(*College of Information Engineering*, *Inner Mongolia University of Technology*, *Hohhot 010080*, *Inner Mongolia*, *China*)

Abstract　At present, in the research of machine translation based on neural networks, large-scale corpus data is one of the necessary conditions for achieving ideal translation results. For large languages such as English, German and other languages, there are abundant resources and data for researchers to use. However, for many small languages, such as Mongolian, the data of parallel corpus are less. In order to alleviate this problem, parameter transfer is introduced into Mongolian-Chinese neuro-machine translation. The model parameters of the trained English-Chinese neuro-machine translation was transferred to the Mongolian-Chinese translation task. The experimental results show that the introduction of transfer learning strategy can improve the results of Mongolian-Chinese translation.

Keywords　Neural network　Machine translation　Parameter transfer　Mongolian-Chinese translation

0　引　言

机器翻译是自然语言处理中的一个最早的研究分支，从 1946 年第一台计算机诞生之日起，机器翻译就作为计算机自然语言处理的一个重要研究领域受到了人们的广泛关注[1]。

机器翻译的发展历程主要经历了基于规则的机器翻译[2]、基于实例的机器翻译[3]和基于统计的机器翻译[4]。20 世纪 90 年代初，随着互联网的兴起，统计机器翻译在自然语言处理领域取得了较为理想的结果，基于统计的机器翻译方法逐渐成为主流。之后随着深度学习的兴起及其在计算机视觉等领域取得了不错的进展，深度学习[5]在自然语言处理中得到了广泛的关注，为解决自然语言处理中稀疏特征容易拟合、对语言结构缺乏描述能力等问题提出新的思路。2014 年以来，端到端的神经网络机器翻译[6-7]得到了迅速发展，基于神经网络技术的机器翻译方法在多个平行语言对中表现显著超过了传统的统计机器翻译[8]。

端到端神经机器翻译应用一种“编码器-解码器”的框架，它的核心思想是通过训练神经网络，找到源语言文本和目标语言文本之间的一种映射关系，直接将

收稿日期：2019－06－24。国家自然科学基金项目（61363052，61502255）；内蒙古自治区自然科学基金项目（2016MS0605）；内蒙古自治区民族事务委员会基金项目（MW-2017-MGYWXXH-03）。**王宇飞**，硕士生，主研领域：人工智能，模式识别。**苏依拉**，教授。**赵亚平**，硕士生。**孙晓骞**，硕士生。**仁庆道尔吉**，教授。

源语言映射到目标语言。“编码器-解码器”方法简化了传统的机器翻译方法的过程。同时神经网络机器翻译很好地缓解了机器翻译中存在的长距离依赖问题。对于不同种类的语言而言,可能存在独有的语法和语言规则,其中最明显的区别之一就是不同语言之间的句子、词语顺序存在的差异,由此带来了将源语言映射为目标语言时的长距离调序问题。神经网络机器翻译通过引入长短时记忆 LSTM(Long Short Term Memory)循环神经网络和注意力机制[9]来捕获长距离依赖[10],从而显著提高了机器翻译译文的流畅性和可读性。

内蒙古地区和蒙古国的官方用语为蒙古语,但是由于地域、经济和技术发展等各方面因素的制约,蒙古语相关的自然语言处理研究起步较晚、蒙古语语法特殊性以及蒙古语形态变化相对复杂使得蒙汉翻译的研究进展相对缓慢[11]。目前相关的蒙汉机器翻译取得了一定的研究成果,但无论是传统的基于统计的机器翻译还是基于神经网络的机器翻译方法,其得到较好的翻译模型通常都需要大量的、高质量的以及包含各个领域文本内容的双语平行语料库作为神经网络的训练数据。

蒙汉平行语料库相较于大语种例如英汉(英语和汉语)、英德(英语和德语)等而言,目前还处于较为匮乏的阶段,而平行语料库规模的不足直接影响到神经网络训练模型的结果,在这种情况下的蒙汉机器翻译的译文质量依然有待提高。因此,如何利用目前数量、规模有限的平行语料得到效果较为理想的机器翻译模型已经成为一个十分重要的研究方向和内容。蒙古语属于乏资源语言,因此本文将迁移学习策略引入到蒙汉神经机器翻译中来缓解平行语料库缺乏的问题。首先利用大规模的英汉平行语料库进行训练,得到质量较高的翻译模型后将其参数迁移至蒙汉神经机器翻译的过程中,即使在蒙汉平行语料库规模较小的条件下,相较于之前直接训练的结果,迁移策略对最终的翻译译文质量也有一定提升。同时为缓解未登录词的影响,进一步提高翻译效果,本文对汉语语料进行分词,对蒙古语语料进行 BPE 分词预处理,对蒙汉平行语料使用 Word2vec 预训练词向量。

1 相关技术

1.1 语料库的预处理

向量化单词或词语是将语料输入神经网络的首要工作,英语文本中单词之间本身存在空格作为分隔符,因此无须再进行基础的分词处理。而汉语并没有天然的标志将词语分隔开来,大部分是短语或者整个句子之间才有标点符号。如果把整个短语或句子直接转化为一个固定维度向量的表现形式,这会降低句子的语言表达力,加大神经网络学习语言模型的难度,所以汉语语料库的预处理是必要的。

本文采用基于统计的分词方法,其方法为在给定大量已经分词的文本的前提下,利用统计机器学习模型来学习语言的词语切分规律,以此来实现对未知文本的切分。本文采用 jieba 分词对汉语语料进行处理。

蒙古语在构词上属于黏着语并且形态丰富[12],黏着语的一个特点是通过在词根的前缀、中缀和后缀接其他构词成分作为派生新词的手段。所以在蒙汉机器翻译的实践中,未登录词的问题也不容忽视。针对该问题,本文采用字节对编码(Byte Pair Encoding,BPE)技术对蒙古语语料库进行预处理操作,只对语料库中低频词进行分词操作,从而提高低频词的子词的共现次数,这样对蒙汉翻译中的未登录词问题达到一定的缓解作用。本文以 5 000 操作数对蒙古语语料进行了 BPE 的预处理,图 1 为未进行分词处理的蒙古语,图 2 是经过 BPE 分词处理后的蒙古语语料。

图 1 未经过分词的蒙古语语料

图 2 经过 BPE 处理的蒙古语语料

目前很多自然语言处理任务中常用单词作为词汇单元。但对于蒙古语而言这种方法是有很大缺陷的,蒙古语可能组合的单词形式的数量非常多,所以翻译结果中出现未登录词的情况较多。很多的机器翻译模型系统用 < unk > 标记来替代未登录词[13]。神经机器翻译的研究人员主要通过未登录词替换和采样等方法处理受限词汇量问题[14]。Sennrich 等[15]提出基于子词单元(subword unit)的神经机器翻译,他们使用 BPE 的方式来获得子词词典。为了缓解蒙汉机器翻译过程出现严重的未登录词现象,本文使用 BPE 技术对蒙汉平行语料处理后原词汇被符号@@分隔成对应的子词。

1.2　Word2vec

自然语言是一套用来表达含义的复杂系统。在这套系统中,词是表义的基本单元,同时是自然语言用来表达含义的基本单元。机器学习使用向量来表示词,即词向量。要将符号形式的词语转化为数值形式,需要进行词嵌入。Word2vec 就是词嵌入(word embedding)方法的一种。该方法由谷歌提出,不需要大量的人工标记样本就可以得到质量较为理想的词向量。

在 Word2vec 之前,自然语言处理通常把词语转换为离散的单独的符号,即 one-hot 表示。该方法为语料库中的每个词语单独地赋予一个只有一个中间值 1 其余值为 0 的向量。但是 one-hot 表示的词向量之间相对独立,无法体现相近词之间的关系,且词向量的维度取决于语料库中词语数量,如果语料库规模较大,包含词语数量的规模也十分可观,这样就会使得向量矩阵过于稀疏,并造成维度灾难。

Word2vec 本质上是一个简化的神经网络,其输入是 one-hot 词向量,在隐藏层是线性单元,其输出层使用 Softmax 分类器,输出维度与输入维度一致。而 Word2vec 即隐藏层的输出,因为参与计算的 one-hot 向量的特点,所以 Word2vec 也即输入层和隐藏层之间的权重。本文使用 Word2vec 时选择 Skip-gram 词向量模型方法,图 3 为其基本内部结构示意图。

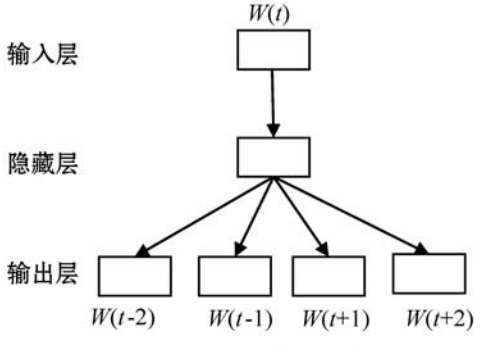

图 3　Skip-gram 模型结构

图 3 中,t 为输入序列中第 t 个位置的词语,中心词 $W(t)$ 作为神经网络的输入值,经过神经网络的隐藏层、输出层输出预测中心词前后出现若干词的相应的概率。本文采用基于层次的 Softmax 归一化方法进行 Skip-gram 词向量模型建模,Skip-gram 模型需要计算出两个窗口大小的词汇也就是相对于中心词上下文 $2n$ 个词汇的概率。

1.3　迁移学习策略

迁移学习一般先通过训练原任务获取一定的知识并将其记忆,然后将这些存储下来的知识迁移应用到与原任务相近的任务中。迁移学习策略允许借用大量已有的标记数据来训练网络并将其学到的知识迁移到标记数据较少的神经网络模型中,从而减少应用任务的训练数据量[16-17]。

传统自然语言处理的机器学习方法,首先针对特定的语言,通过相应的大量平行语料库进行模型的训练,然后将翻译模型应用到该特定的语言翻译任务中。而迁移学习与之相比,不再要求其基本条件,一是用来训练机器学习模型的数据和测试数据必须是同分布的并且数据间是相互独立的关系;二是用于训练的数据集必须保证一定的规模才可能得到效果理想的模型。

迁移学习属于机器学习方法中的一种。迁移学习中域(Domain)是学习的主体,由数据特征和分布组成。域又分为包含已有知识的源域和将要学习的目标域。迁移学习就是研究如何把在源域中学习到的知识迁移运用于目标域中。该方法按照迁移方法不同可以分为基于实例的迁移(Instance Based Transfer Learning,IBTL)、基于特征的迁移(Feature Based Transfer Learning,FBTL)、基于模型的迁移(Parameter Based Transfer Learning,PBTL)以及基于关系的迁移(Relation Based Transfer Learning,RBTL)[18]。本文在蒙汉机器翻译中研究基于模型的迁移,又称参数迁移方法,即在源域和目标域中存在可以共享的模型参数。具体方法是,首先对源域进行训练,得到效果较好的模型后,将该模型参数记录并迁移应用到目标域中,再根据目标域继续学习新的模型。

如图 4 所示,在一般的监督学习中,针对不同的任务,需要收集大量与任务相关的有标记的训练数据得到各自独立的模型。相比于这种策略,迁移学习可以实现在少量的标记数据条件下得到满意的模型。迁移学习将训练模型 A 获取的知识存储下来,应用于新的任务/域中,这里是对模型 B 的训练,以达到提高模型

B 性能的目的,图 5 为迁移学习模型示意图。

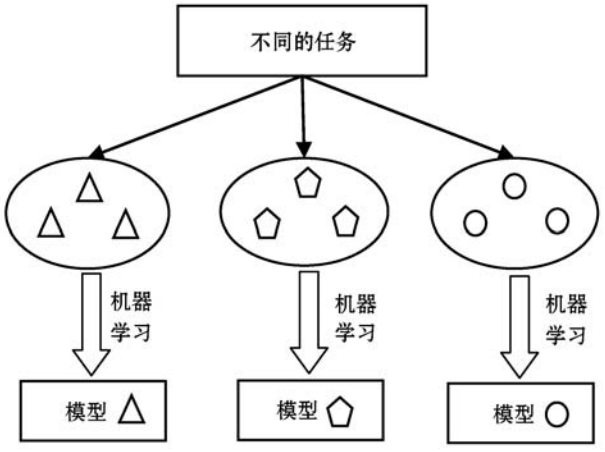

图 4 传统机器学习模型示意图

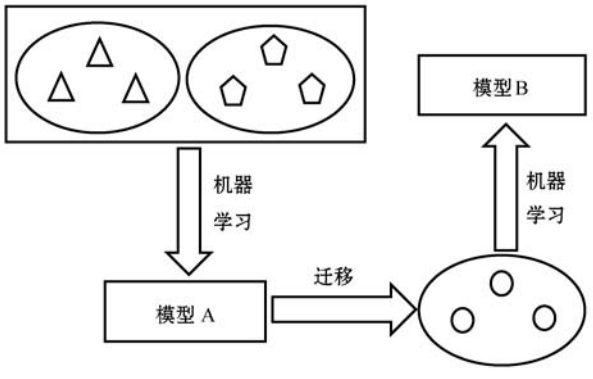

图 5 迁移学习模型示意图

迁移学习策略十分适用于已有标记数据缺乏的任务中,目前除了少量平行语料数据资源丰富的语言外(如英汉),很多语言对双语平行语料资源匮乏的问题普遍存在,没有足够的标记数据[19-20],迁移学习的引入将缓解这一困难。

1.4 长短时记忆网络

(1) 单向 LSTM。自然语言处理是典型的序列到序列的机器学习任务,对于这种任务,使用最多的是循环神经网络(Recurrent Neural Networks,RNN)。RNN 能够记忆输入网络的数据序列,通过这些信息 RNN 来进行一些预测。长短时记忆(Long Short-Term Memory,LSTM)是 RNN 的一种,它在基础的 RNN 上进行了改进。LSTM 拥有更长时间的记忆能力,有效克服了机器翻译中存在的长距离依赖问题,显著提升机器翻译译文的流利度和可读性。

LSTM 在普通 RNN 基础上加入判断信息是否有用的记忆单元,这个记忆单元被称为 block。该 block 主要通过门(gate)机制进行信息的控制。一个 block 中放置三个门控单元以及一个记忆细胞,包括遗忘门控制单元、输入门控制单元和输出门控制单元。LSTM 记忆单元结构如图 6 所示。

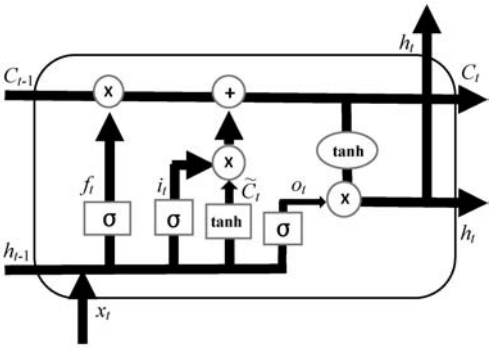

图 6 LSTM 结构图

图 6 中,h_t表示 t 时刻 block 的输出,x_t是当前输入的新信息。遗忘门中,f_t由 h_{t-1}和 x_t得到,它用来计算 C_{t-1}中信息的遗忘程度。上界为 1 表示完全记忆,下界为 0 表示完全遗忘,其计算公式为:

$$f_t = \sigma(\boldsymbol{W}_f \cdot [h_{t-1}, x_t] + \boldsymbol{b}_f) \tag{1}$$

输入门中,i_t用来控制新输入状态信息的更新程度,通过 h_{t-1}和 x_t计算得到,新状态信息为 $\tilde{C}$,其计算公式如下:

$$i_t = \sigma(\boldsymbol{W}_i \cdot [h_{t-1}, x_t] + \boldsymbol{b}_i) \tag{2}$$

$$\tilde{C}_t = \tanh(\boldsymbol{W}_C \cdot [h_{t-1}, x_t] + \boldsymbol{b}_C) \tag{3}$$

当前的新信息 C_t可以通过以上的状态计算,即遗忘一些信息并加入要记忆的新信息,计算公式如下:

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \tag{4}$$

输出门中,o_t用以控制哪些信息需要作为输出,计算公式如下:

$$o_t = (\boldsymbol{W}_0 [h_{t-1}, x_t] + \boldsymbol{b}_o) \tag{5}$$

$$h_t = o_t \times \tanh(C_t) \tag{6}$$

式中:$\boldsymbol{W}$ 为权重矩阵;$\boldsymbol{b}$ 为偏置矩阵。

一般 RNN 在训练中可能会发生梯度消失的情况,后面的时间节点对以前时间节点的感知力下降,这样两个距离较远的节点之间的联系就被削弱。而 LSTM 可以记忆符合算法或规则的信息,其他信息则会被遗忘。当神经网络中的信息进入隐藏节点时,记忆单元通过三个门机制来控制信息的传播,只有符合门控的信息才会被记忆并且向后传递,否则遗忘门会将信息丢弃,得以保留的信息则会根据需求继续向后传播或者直接输出结果。

(2) 双向 LSTM。单向 LSTM 对句子进行建模存在一个问题:无法编码从后到前的信息。而在机器翻译中,对某个词的翻译往往需要借助上下文相关信息,此时双向 LSTM 就可以更好地捕捉到双向语义依赖。

双向 LSTM 是由前向 LSTM 与后向 LSTM 组合而成。从 1 时刻到 i 时刻正向计算一遍,得到并保存每个时刻向前隐藏层的输出。还要沿着时刻 i 到时刻 1 反向计算一遍,得到并保存每个时刻向后隐藏层的输

出。最后在每个时刻结合双向计算相应时刻输出的结果得到最终的输出。双向 LSTM 结构图如图 7 所示。

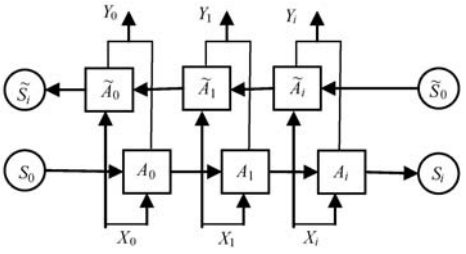

图 7　双向 LSTM 结构图

图 7 中,X_i 为 i 时刻的输入,双向 LSTM 会保留两个值,正向计算过程的隐藏层状态 A_i 和反向计算过程的隐藏层状态 $\tilde{A}_i$,最终的输出值 Y_i 取决于上述两个值。即正向计算时,隐藏层 S_i 与 S_{i-1} 有关,反向计算时,S_i 与 S_{i+1} 有关。

1.5　注意力机制

在普通的“编码器-解码器”架构中,解码和编码都依赖于内部一个固定长度向量(中间语义向量 context)。注意力机制打破了这个限制,它通过保留 LSTM 编码器对输入序列的中间输出结果,然后训练一个模型来对这些输入进行选择性学习,并且在输出时将输出序列与之相关联。Bahdanau 等[9] 把注意力机制用到了神经网络机器翻译上,在生成目标序列中的每个词时,传统的做法是只用编码器最后一个时刻的输出,但加入注意力机制后,用到的 context 是源序列通过编码器隐藏层的加权和。通过这种方式,使源序列的每个词与目标序列的词建立了联系。图 8 为神经机器翻译注意力机制示意图。

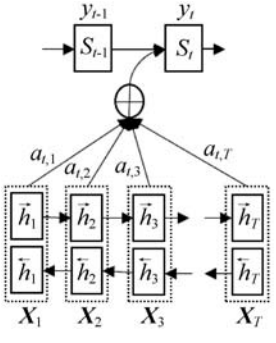

图 8　注意力机制

图 8 中,设有长度为 T 的词向量序列 $\boldsymbol{X}=(\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_T)$,源语言词的向量 $\boldsymbol{X}_2$ 同时包含了上一个词 $\boldsymbol{X}_1$ 和后一个词 $\boldsymbol{X}_3$ 的语义向量信息。其中:a 表示源语言对齐权重;$\boldsymbol{S}$ 表示神经网络隐藏层状态向量;Y 表示神经网络的输出序列;h 为编码器输入序列。基于注意力机制的对齐权重计算公式如下:

$$a_{ts}=\frac{\exp(score(h_t,h_s))}{\sum_{s'=1}^{s}\exp(score(h_t,h_s))} \tag{7}$$

式中:$score(h_t,h_s)=h_t\boldsymbol{W}h_s$,$h_t$ 表示目标端隐藏节点状态信息,h_s 表示源语言端隐藏层状态信息,$\boldsymbol{W}$ 为权重矩阵。将注意力机制加入到神经机器翻译模型中,在生成目标语言词时,计算得到当前词和源语言端所有词的对齐的权重,即生成当前词时只关注源语言端的部分词。注意力机制在一定程度上使得神经机器翻译模型在处理长句子上的能力得到提升。

1.6　TensorFlow 深度学习框架

TensorFlow 是由谷歌团队开源的机器学习框架。它拥有一个由工具、库和社区资源组成的全面、灵活的生态系统,使研究人员能够在机器学习中推动最先进的技术,开发人员可以轻松地构建和部署基于机器学习的应用程序。图 9 为 TensorFlow 深度学习框架示意图。

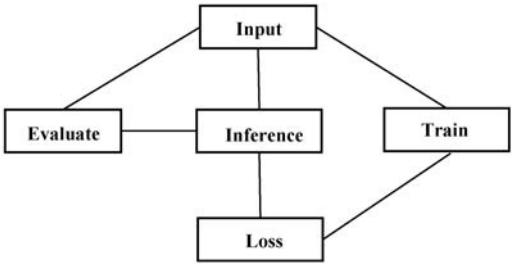

图 9　TensorFlow 模型训练框架示意图

如图 9 所示,Input 为神经网络输入,通常是训练模型的数据;Inference 表示神经网络经过前向推断预测模块;Evaluate 表示神经网络模型评估模块;Loss 是计算神经网络损失代价模块;Train 表示神经网络的训练模块,主要通过计算各个节点的梯度以及通过梯度来更新模型中的参数。神经网络中的经验风险最小化以下的目标函数:

$$L=\frac{1}{N}\sum_i L_i+\lambda R(\boldsymbol{W}) \tag{8}$$

式(8)用于计算整个训练数据集的损失值,由神经网络模型中所有训练样本 N 的损失值 L_i 的均值和正则化损失值 $R(\boldsymbol{W})$ 两部分之和构成,λ 为权重系数,i 表示第 i 个训练样本。

1.7　TensorFlow 中的参数迁移

本文进行的基于参数迁移的蒙汉机器翻译模型是搭建在 TensorFlow 环境下,使用 TensorFlow 内部的存储器对象来保存模型。利用导入预训练的模型来实现参数权重的迁移,将训练好的英汉神经网络参数权重迁移蒙汉翻译模型中,即蒙汉神经机器翻译时网络各节点参数不再是随机初始化,而是将训练好的英汉模型参数导入蒙汉模型中进行初始化。

TensorFlow 深度学习框架中计算图和相关参数是分开存储的,因此导入预训练的模型需要分两步:首先需要构造神经网络模型图;然后加载权重参数。

当预训练模型训练完成后，Checkpoint 文件（对应的扩展名为. ckpt）会存储神经网络中所有节点的权重、偏置和梯度等相关变量的值。通过调用 TensorFlow 中 Saver 对象的 save( )方法来保存全部网络参数。然后在新的神经网络中，通过 restore( )方法来加载预训练模型的参数，在此基础上继续进行训练。图 10 为 TensorFlow 参数迁移的过程示意图。

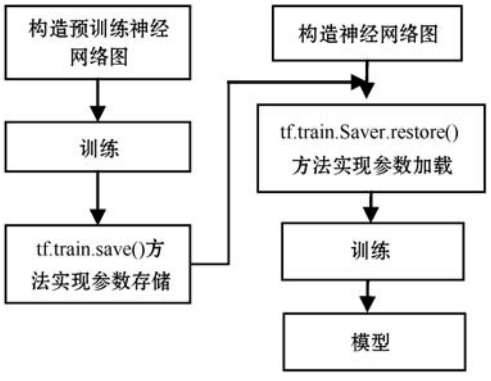

图 10 TensorFlow 参数迁移

2 实 验

2.1 实验数据

本文实验数据包含两大部分：一部分是用于英汉翻译模型的英汉平行语料库；另一部分是用于蒙汉翻译模型的蒙汉平行语料库。每部分的训练集、验证集和测试集划分及规模如表 1 所示。

表 1 实验数据集划分

平行语料库	训练集	验证集	测试集
英汉	160 万句对	16 万句对	10 万句对
蒙汉	145 620 句对	12 000 句对	4 000 句对

2.2 实验流程

本文具体实验流程如图 11 所示。

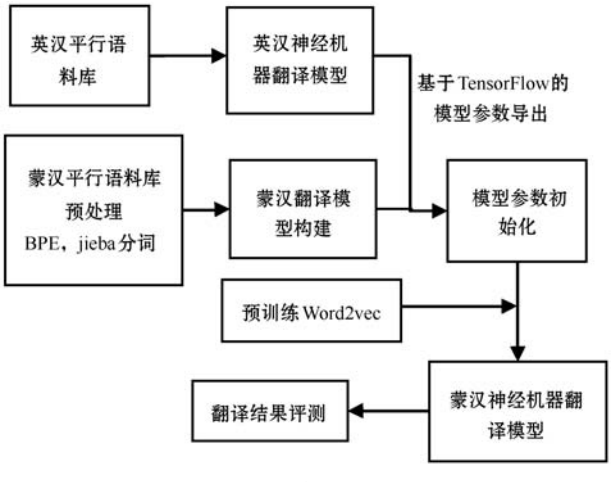

图 11 实验流程

TensorFlow 平台搭建神经网络模型，通过英汉平行语料库训练得到英汉翻译模型，将该神经网络模型的参数存储并导出。对蒙古语和汉语分别进行预处理得到处理后的平行语料库，同样在 TensorFlow 平台搭建用于蒙汉翻译的神经网络模型，把导出的参数导入该神经网络中进行参数初始化，然后再输入处理后的蒙汉平行语料库进行训练，最后得到翻译模型并进行翻译结果评测。

2.3 实验结果分析

本文在相同数量蒙汉平行语料库的前提下，进行了基于 TensorFlow 的传统蒙汉神经机器翻译（未进行参数迁移）模型训练、英汉神经机器翻译模型训练、基于 Word2vec 模型预训练得到词向量，并完成基于参数迁移的蒙汉神经机器翻译模型训练。本文对英汉进行了 185 000 步训练得到模型参数，对蒙汉进行了 100 000 步训练，记录了每 10 000 步 BLEU 值以进行实验对比。在 Tensorflow 框架中，Checkpoint 文件存储了神经网络中所有结点的参数权重、偏置值、函数梯度等相关变量的值。本文在利用英汉平行语料训练神经机器翻译模型后会得到相关模型文件，其中包括了存储神经网络结构图的. meta 文件和存储神经网络参数权重的. ckpt 文件。图 12 为本文 BLEU 值随训练步数变化的情况。

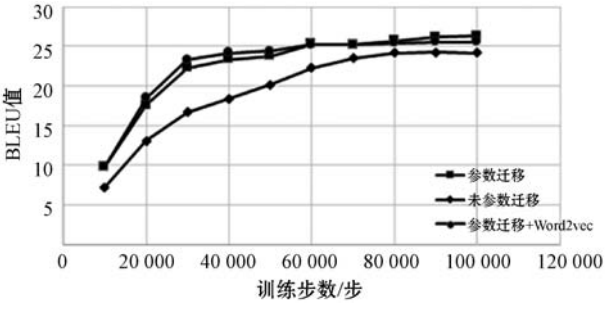

图 12 BLEU 值与训练步数关系

通过图 12 对比实验，可以分析出，使用了参数迁移的蒙汉神经机器翻译的最终结果（BLEU 值）要优于传统的未使用迁移策略的机器翻译结果（提高了 2.1BLEU 值）。可以看到，在经过英汉参数迁移的基础上进行训练，BLEU 值提高的速度在前期要快于完全依靠蒙汉语料进行神经网络初始化的翻译模型。以 BLEU 值达到 20 为目标点，引入参数迁移的翻译模型在训练不到 30 000 步就可以完成，但是未引入参数迁移的传统翻译模型需要训练约 50 000 步。

因为翻译模型训练之前参数初始化的影响，将同属于翻译任务的英汉翻译模型的参数引入蒙汉翻译模型的初始化，使得模型在训练之前已有了一定的参数基础，所以再进行训练时，其学习速率会有所

提升。并且使用 Word2vec 预训练的词向量后,最终的结果也要优于传统的翻译效果,其 BLEU 值相比提高了 1.4。

3 结　语

目前神经机器翻译在大语种之间互译的相关研究趋于成熟,例如英德、英汉等语种之间的互译效果已经达到了很好的效果。但是对于蒙古语这种小语种来说,现有的平行语料库规模有限,短时期内获取语料的难度也相对较大,这就限制了蒙汉神经机器的发展。本文将迁移学习引入到蒙汉神经机器翻译中,利用大语种拥有大规模语料库的优势,将其模型参数迁移。并且为了提高翻译效果,使用 Word2vec 预训练了词向量,将词汇表示成固定维度的实数向量,相近语义的词在词向量空间上距离相近,这样在一定程度上提升了双语词向量的表达能力。实验结果表明,引入迁移学习对蒙汉翻译效果有一定提升(BLEU 值),说明对蒙汉机器翻译中存在的平行语料库不足的问题有一定的缓解。

同时,本文在源域中选择的是英语到汉语的翻译任务,而目标域是蒙古语到汉语,两者虽然都属于机器翻译任务,但是英语和蒙古语在语法、语言结构上仍然存在较大差别。迁移学习在机器翻译中应用时应存在一定语言相关性,因此在源域选择与蒙古语语法相近的语言例如日语,这样迁移到目标域可能会达到更好的翻译效果,但是目前获得较大规模的日汉平行语料库还是相对困难的。平行语料库不足依然是蒙汉神经机器发展的主要阻碍之一,而迁移学习的特性能在一定程度上缓解这个问题。因此,接下来的工作一是尽可能获取蒙汉平行语料,二是继续进行迁移学习方面的研究,获取与蒙古语相近的其他拥有较大规模的语言的语料库,并继续尝试其他模式的迁移方法在蒙汉神经机器上的研究。

参 考 文 献

[1] 戴新宇,尹存燕,陈家骏,等. 机器翻译研究现状与展望[J]. 计算机科学,2004,31(11):176-179,184.

[2] 李洪政,朱筠,晋耀红. 汉英专利机器翻译中动-介兼类词自动识别[J]. 计算机工程与应用,2015,51(11):6-11.

[3] 赵红梅,吕雅娟,贲国生,等. 第七届全国机器翻译研讨会机器翻译评测总结[J]. 中文信息学报,2012,26(1):22-30.

[4] 杜金华,张萌,宗成庆,等. 中国机器翻译研究的机遇与挑战——第八届全国机器翻译研讨会总结与展望[J]. 中文信息学报,2013,27(4):1-8.

[5] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786):504-507.

[6] Kalchbrenner N, Blunsom P. Recurrent continuous translation models[C]//2013 Conference on Empirical Methods in Natural Language Processing, 2013.

[7] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks[EB]. arXiv: 1409.3215, 2014.

[8] Zhang J, Zong C. Deep neural networks in machine translation: an overview[J]. IEEE Intelligent Systems, 2015, 30(5):16-25.

[9] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[EB]. arXiv:1409.0473,2014.

[10] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation,1997,9(8):1735-1780.

[11] 王斯日古楞,斯琴图,那顺乌日图. 基于短语的汉蒙统计机器翻译研究[J]. 计算机工程与应用,2010,46(14):138-142.

[12] 乌兰,达胡白乙拉,关晓炟,等. 蒙古语短语结构树的自动识别[J]. 中文信息学报,2014,28(5):162-169,181.

[13] 韩冬,李军辉,熊德意,等. 基于子字单元的神经机器翻译未登录词翻译分析[J]. 中文信息学报,2018,32(4):74-79,119.

[14] 刘洋. 神经机器翻译前沿进展[J]. 计算机研究与发展,2017,54(6):1144-1149.

[15] Sennrich R, Haddow B, Brich A. Neural machine translation of rare words with subword units[C]//54rd Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2016:1715-1725.

[16] Zoph B, Yuret D, May J, et al. Transfer learning for low-resource neural machine translation[C]//2016 Conference on Empirical Methods in Natural Language Processing, 2016: 1568-1575.

[17] 庄福振,罗平,何清,等. 迁移学习研究进展[J]. 软件学报,2015,26(1):26-39.

[18] Pan J. Feature-based transfer learning with real-world applications[D]. Hong Kong University of Science and Technology,2010.

[19] Ciresan D C, Meier U, Schmidhuber J. Transfer learning for latin and chinese characters with deep neural networks[C]// The 2012 International Joint Conference on Neural Networks (IJCNN), 2012.

[20] 王强,杜权,肖桐,等. 基于 Transfer 和 Triangulation 融合的中介语统计机器翻译方法[J]. 中文信息学报,2017,31(4):36-43.