

文章编号: 1003-0077 (2011) 00-0000-00

藏汉神经网络机器翻译研究*

李亚超^{1,2}, 熊德意², 张民², 江静¹, 马宁¹, 殷建民³

(1. 甘肃省民族语言智能处理重点实验室(西北民族大学), 甘肃 兰州 730030;

2. 苏州大学 计算机科学与技术学院, 江苏 苏州 215000;

3. 潍坊北大青鸟华光照排有限公司, 山东 潍坊 261000)

摘要: 神经网络机器翻译是最近几年提出的机器翻译方法, 在多数语言对上逐渐超过了统计机器翻译方法, 成为当前机器翻译研究前沿热点。该文实验了基于注意力的藏汉神经网络机器翻译, 采用迁移学习方法缓解藏汉平行语料数量不足问题。实验结果显示, 该文提出的迁移学习方法简单有效, 相比短语统计机器翻译方法, 提高了3个 BLEU 值。从译文分析中可以看出藏汉神经网络机器翻译的译文比较流畅, 远距离调序能力较强, 同时也存在过度翻译、翻译不充分、翻译忠实度较低等神经网络机器翻译的共同不足之处。

关键词: 藏语; 神经网络机器翻译; 注意力机制; 循环神经网络; 迁移学习

中图分类号: TP391

文献标识码: A

Research on Tibetan-Chinese Neural Machine Translation

LI Yachao^{1,2}, XIONG Deyi², ZHANG Min², JIANG Jing¹, MA Ning¹, YIN Jianmin³

(1. Key Laboratory of National Language Intelligent Processing (Northwest Minzu University), Lanzhou, Gansu 730030, China;

2. School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215000, China;

3. Weifang Beida Jade Bird Huaguang Information Technology Co., Ltd., Weifang, Shandong 261000, China)

Abstract: Neural machine translation (NMT), a new machine translation method based on sequence-to-sequence learning, which uses neural network to complete the translation from the source language to the target language directly, has surpassed statistical machine translation (SMT) in several language pairs gradually. The experiment has been carried out on attention-based NMT for Tibetan-Chinese machine translation, and the transfer learning was applied on the experiment too. Experimental results show that the transfer learning method proposed is simple and effective, compared with the phrase-based SMT, gains of 3 BLEU scores. From the analysis of translations, it can be seen that the translations of Tibetan-Chinese NMT are relatively fluent, the ability of long-distance ordering is stronger, at the same time, and there are common shortcomings of NMT such as over translation, under translation, low translation fidelity and so on.

Key words: Tibetan; neural machine translation; attention mechanism; recurrent neural network; transfer learning

1 引言

机器翻译研究如何利用计算机自动地实现不同语言之间的相互转换, 是自然语言处理的重要研究领域。机器翻译方法可以分为基于规则的机器翻译、基于实例的机器翻译、统计机器翻译, 以及当前的神经网络机器翻译等。互联网与移动互联网的兴起, 为机器翻译研究提供了巨大的应用空间。产业界如谷歌、微软、百度等公司都对机器翻译投入巨大资源进行研

* 收稿日期:

定稿日期:

基金项目: 国家自然科学基金(61403269, 61432013, 61525205); 西北民族大学中央高校基本科研业务费专项资金资助项目(31920170154, 31920170153); 甘肃省高等学校科研项目(2016B-007); 中国民族语言文字信息技术重点实验室开放基金课题资助(KFJJ201605)。

究,同时也对外提供了相关的翻译服务,如谷歌翻译、百度翻译等。国内外的科研机构同样把该领域作为重点的研究方向。机器翻译成为当前自然语言处理研究的热点之一,不仅具有研究价值,同样具有实用价值。

神经网络起源于 20 世纪 40 年代,发展于 50-70 年代,由于种种限制 80-90 年代为神经网络研究的低潮期^[1]。2006 年, Hinton 等人^[2]解决了神经网络的训练难题,此后,随着计算能力的提高,以及训练数据量的增加,神经网络成功应用在多个领域。近年来,神经网络在图像识别、语音识别等领域取得巨大成功,同时学者们也将该技术应用在自然语言处理任务上,如语言模型、词语表示、序列标注等任务,并取得了令人鼓舞的成绩^[3-5]。

在机器翻译研究上,神经网络机器翻译实现源语言到目标语言的直接翻译,极大地提高了翻译效果,是目前机器翻译研究的前沿热点。基于神经网络的机器翻译方法源于 20 世纪 90 年代^[6],由于资源限制并没有成为主流方法。在深度学习兴起之后,神经网络通常用于统计机器翻译的词对齐、翻译规则抽取等^[7]。从 2014 年开始, Sutskever、Cho、Jean 等人^[8-11]提出单纯采用神经网络实现机器翻译的新方法,称为神经网络机器翻译。该方法实现了源语言到目标语言的直接翻译,并在多个语言对上超过了传统的统计机器翻译方法^[12],逐渐成为当前主流的机器翻译方法。

藏语机器翻译相关研究主要集中在统计机器翻译,以及藏语机器翻译相关基础研究。如藏语统计机器翻译短语抽取^[13]、基于短语的统计翻译等^[14-15];基于树到串的藏语机器翻译^[16];规则和统计相结合的汉语到藏语机器翻译^[17];面向藏语机器翻译的动词处理^[18]、短语句法研究^[19]、功能组块识别^[20]、藏文数词识别与翻译^[21]等。从整体上看,藏语机器翻译研究较为滞后,有关神经网络机器翻译在藏语上的应用及实际效果,并没有见到公开发表。

本文后续部分安排如下:第二部分详细介绍神经网络机器翻译以及迁移学习在藏汉神经网络机器翻译上的应用,第三部分进行实验与分析,第四部分为全文的总结和下一步工作安排。

2 神经网络机器翻译

机器翻译把翻译问题看作求解概率问题,即给定源语言 s ,求目标语言 t 在源语言 s 出现下的条件概率 $p(t|s)$,模型参数从双语平行语料中学习得到。神经网络机器翻译在翻译建模上完全采用神经网络完成源语言到目标语言的直接翻译,不需要经过统计机器翻译的词对齐、翻译规则抽取、调序等步骤。本小节简单介绍神经网络机器翻译,以及迁移学习方法在藏汉神经网络机器翻译上的应用。

2.1 基本神经网络机器翻译模型

神经网络机器翻译源于序列到序列学习^[8-10],本文以蒙特利尔大学提出的翻译模型为例说明^[9-10]。编码器解码器模型(Encoder-Decoder)是神经网络机器翻译模型之一,编码器读取源语言句子,编码为维数固定的向量,解码器读取该向量,依次生成目标语言词语序列,如图 1 所示。神经网络机器翻译模型主要部分如下所示, x 表示输入, h 表示隐藏状态, y 表示输出。

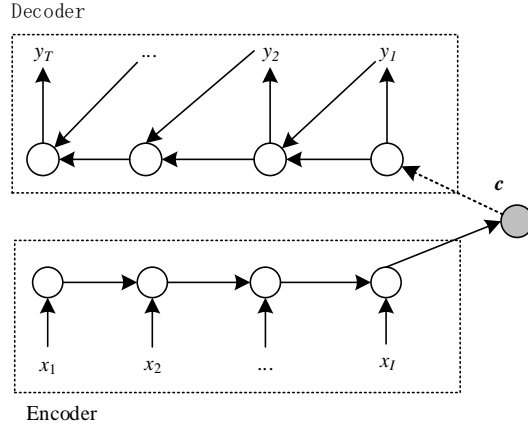


图 1 编码器解码器模型

编码器：编码器读取输入 $\mathbf{x} = (x_1, x_2, \dots, x_I)$ ，将其编码为隐藏状态 $\mathbf{h} = (h_1, h_2, \dots, h_I)$ ，编码器通常采用循环神经网络（Recurrent Neural Network, RNN）实现，更新方式如下所示：

$$h_i = f(x_i, h_{i-1}) \quad (1)$$

$$\mathbf{c} = q(\{h_1, \dots, h_I\}) \quad (2)$$

\mathbf{c} 为源语言句子表示， f 和 q 是非线性函数。

解码器：在给定源语言表示 \mathbf{c} 和前驱输出序列 $\{y_1, \dots, y_{t-1}\}$ 条件下，解码器依次生成目标语词 y_t ，如下所示：

$$p(\mathbf{y}) = \prod_{t=1}^T p(y_t | \{y_1, \dots, y_{t-1}\}, \mathbf{c}) \quad (3)$$

$\mathbf{y} = (y_1, y_2, \dots, y_T)$ ，解码器通常同样采用循环神经网络，形式如下所示：

$$p(y_t | \{y_1, \dots, y_{t-1}, \mathbf{c}\}) = g(y_{t-1}, s_t, \mathbf{c}) \quad (4)$$

g 是非线性函数用来计算 y_t 的概率， s_t 是解码器的隐藏状态。

模型训练：编码器和解码器可以进行联合训练，形式如下：

$$L(\theta) = \max_{\theta} \frac{1}{N} \sum_{n=1}^N \log p_{\theta}(y_n | x_n) \quad (5)$$

θ 是模型的参数，通常采用梯度下降法进行计算， (x_n, y_n) 表示双语训练语料。

2.2 基于注意力的神经网络机器翻译

普通神经网络机器翻译模型将源语言句子表示成一个固定向量，该方法存在不足之处，比如大小固定的向量并不能充分表达出源语言句子语义信息。基于注意力机制的神经网络机器翻译将源语言句子编码为向量序列，在生成目标语言词语时，通过注意力机制动态寻找与生成该词相关的源语言词语信息^[14]，因而大大增强了神经网络机器翻译的表达能力，在实验中显著提高了翻译效果，如图 2 所示。

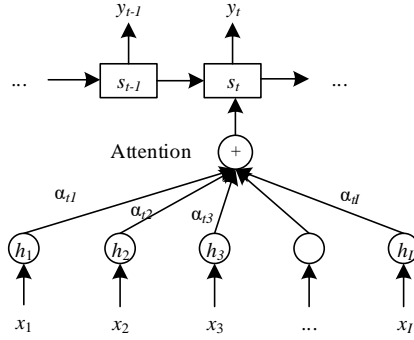


图2 注意力机制图示

当采用注意力机制时，公式3重新定义为：

$$p(y_t|\{y_1, \dots, y_{t-1}, \mathbf{x}\}) = g(y_{t-1}, s_t, c_t) \quad (6)$$

s_t 是 t 时刻循环神经网络的隐藏状态，由如下公式得出：

$$s_t = f(s_{t-1}, y_{t-1}, c_t) \quad (7)$$

g 、 f 是非线性函数，上下文向量(Context Vector) c_t 依赖于源语言编码序列(h_1, h_2, \dots, h_l)， h_i 包含第 i 个输入词上下文信息。 c_t 计算方法如下：

$$c_t = \sum_{j=1}^l \alpha_{tj} h_j \quad (8)$$

α_{tj} 是 h_j 的权重，计算方法如下：

$$\alpha_{tj} = \frac{\exp(e_{tj})}{\sum_{k=1}^l \exp(e_{tk})} \quad (9)$$

$e_{tj} = a(s_{t-1}, h_j)$ 是对齐模型，计算出 t 时刻生成词与第 j 个源语言词的匹配程度。

基于注意力的神经网络机器翻译在解码时只关注源语言句子中一部分区域，能够增强长句子表示能力。相比普通的神经网络机器翻译，该方法在解码时融合了更多的源语言端信息，可以显著提升机器翻译效果，是目前神经网络机器翻译的主流方法。

2.3 资源稀缺条件下的藏汉神经网络机器翻译

在训练语料较少情况下，神经网络机器翻译效果显著低于统计机器翻译方法，这在多个语言对上得到验证^[23-24]。迁移学习^[22]能够将学习到的知识应用到相近任务上，可以减少应用任务的训练数据量。在神经网络机器翻译中，迁移学习主要用在资源稀缺语言翻译上，可以将资源丰富语言对的翻译模型参数，如英汉、英法等语言，迁移到资源稀缺语言上，如藏语到汉语的翻译等，并在一些实验中得到了实际验证。

藏汉机器翻译同样存在训练语料不足问题，为此本文根据 Zoph 等人^[23]思想，首先，采用大规模英汉平行语料训练得到英语到汉语神经网络机器翻译模型。然后，在训练藏语到汉语神经网络机器翻译模型时，采用英汉翻译模型参数初始化藏汉翻译模型参数。最后，对由英汉翻译模型参数初始化后的模型，采用藏汉平行语料进行训练。

与 Zoph 等人^[23]方法不同的是，本文方法对藏汉翻译模型的所有参数均采用英汉模型参数初始化，且在初始化时不要求两种翻译模型的汉语词向量 (Word Embedding) 一致。这种迁移学习方法简单，对翻译模型本身不做任何改变，并具有语言无关优点，因而更加适合训练语料极度缺少的特殊情况。

3 实验与分析

3.1 实验设置

本实验采用的平行语料为 2011 年机器翻译研讨会 (CWMT 2011) 藏汉机器翻译评测语料, 训练语料 10 万句, 测试语料 650 句。迁移学习采用英汉平行语料, 共 125 万句, 从 LDC 语料中抽取得到。为了相对公平的对比, 统计机器翻译语言模型训练采用大小 100M 的汉语文本语料, 神经网络机器翻译不采用额外单语语料。

对训练语料句子长度限制为 50 词以下, 双语词向量维数为 620, 隐藏层大小为 1000, 解码时 Beamsize 大小为 10, 系统选择 adadelata 方法^[28]优化参数, 训练中 mini-batch 大小为 80 句, Dropout^[29]设置为 0.5。为了减少未登录词问题, 藏语、汉语词典大小设置为 3 万, 约覆盖 99% 的词语。语料训练轮数最多为 60 轮。

本文采用的基线系统为东北大学开发的 Niutrans 短语统计机器翻译系统^[22], 用“Niutrans”表示。神经网络机器翻译系统为本课题组基于 Bahdanau^[23]工作所开发的基于注意力的神经网络机器翻译系统, 循环单元采用门限循环单元 (Gated Recurrent Units, GRU)^[9], 用“NMT”表示。本文提出的采用迁移学习实现的藏汉神经网络机器翻译系统用“NMT+Trans”表示。

由于测试语料限制, 在统计机器翻译中, 测试语料同样作为训练短语翻译模型的开发集, 即统计机器翻译结果是开发集上的结果。在神经网络机器翻译中测试语料同样作为开发集, 但是不影响到翻译模型, 仅用于选择最优的翻译模型。以 BLEU-4 作为评测标准, 采用基于词的评测。

3.2 主要实验结果

表 1 实验结果

翻译系统	BLEU
Niutrans	33.40
NMT	30.39
NMT+Trans	36.52

从实验结果来看基线神经网络机器翻译系统的 BLEU 值低于短语统计机器翻译系统。因为训练数据量较少, 神经网络机器翻译模型并不能获得高质量的词汇语义映射关系。在数据量较少情况下, 神经网络机器翻译效果低于短语统计机器翻译, 这种情况同样出现在英汉、英法等其他语言翻译中^[23-24]。

本文通过迁移学习将语料资源规模较大的英汉神经网络机器翻译模型参数迁移到藏汉神经网络机器翻译中。相比基线神经网络机器翻译, 本文方法提高了 6 个 BLEU 值, 同时也显著超过了短语统计机器翻译系统结果。

根据 Liu^[24]的实验, 在 30 万句的汉英语料上, 神经网络机器翻译比短语统计机器翻译低了 10 个 BLEU 值。本文的平行语料更少, 但是神经网络机器翻译与统计机器翻译效果仅相差了 3 个 BLEU 值。我们认为是该训练语料题材很集中, 且测试集与训练集题材也很一致。因此, 两种翻译方法并没有出现巨大差距。

3.3 模型训练轮数对翻译结果影响

在神经网络机器翻译中, 要对训练语料整体训练多次, 平行语料完整训练一遍称为一轮 (Epoch), 通常经过 20 轮左右的训练, 即可得到稳定的模型。因为藏汉平行语料较少, 在

参考译文 5	会计账簿、财务会计报告。
SMT	会计UNK财务会计的报告关于。
NMT+Trans	UNK、财务的报告。

从翻译示例 1 可以看出，神经网络机器翻译相比统计机器翻译，可以很好地处理汉语语序问题，翻译结果很流畅。

从翻译示例 2 可以看出，相对参考译文“本法 有关规定”，神经网络机器翻译系统翻译成“本 法律 有关规定”，统计机器翻译系统的翻译结果为“本法 有关规定”，能够看出神经网络机器翻译具有很强的词义联想能力。

从翻译示例 3 可以看出，虽然神经网络机器翻译的译文很流畅，但是存在与原文语义相差较大的问题，翻译忠实度较低，并且存在重复翻译现象。

翻译示例 4 句型结构比较复杂，神经网络机器翻译显然能够更好的理解原文意思，并且翻译结果和更为流利。

翻译示例 5 中出现了未登录词，神经网络机器翻译和统计机器翻译均未能够准确翻译出“账簿”这个词，且神经网络机器翻译出现严重的翻译不充分现象，即原文多个词语并没有翻译被完整翻译出来。

从以上翻译结果可以看出，藏汉神经网络机器翻译虽然存在神经网络机器翻译的不足之处，如翻译不充分、过度翻译，对原文的翻译忠实度较低等缺点。但是从整体上看翻译结果较为流畅，对翻译调序处理效果很好，且对原文语义理解能力更强，译文质量整体上优于统计机器翻译方法。这种实验结果符合神经网络机器翻译在其他语言上应用的效果。

4 结论与下一步工作

本文研究了神经网络机器翻译在藏汉翻译上的应用，并采用迁移学习方法将英汉神经网络机器翻译模型参数迁移到藏汉神经网络机器翻译模型中。该方法显著提高了基线神经网络机器翻译系统效果，并超过了短语统计机器翻译，取得了最好的翻译结果。

从实验结果可以得出，藏汉神经网络机器翻译同样具有神经网络机器翻译的共同优点和不足之处。由于语料资源较少，基本的神经网络机器翻译模型效果仍然低于统计机器翻译方法。本文提出的迁移学习方法简单有效，能够使藏汉神经网络机器翻译显著超过短语统计机器翻译方法。该方法对神经网络模型不做任何改变，并具有语言无关性，理论上可以应用到维吾尔语、蒙古语等其他资源稀缺语种神经网络机器翻译中。

由于藏汉平行语料较少，语料资源是藏汉机器翻译的最大障碍。在后续研究中，我们将采用其他方法提高资源稀缺条件下的藏汉神经网络机器翻译效果，以及研究本文方法在其他语言上的应用。

参考文献

- [1] 焦李成, 杨淑媛, 刘芳等. 神经网络七十年: 回顾与展望[J]. 计算机学报, 39(8), 2016:1697-1716.
- [2] Geoffrey E. Hinton, Simon Osindero, Yee-Whye Teh. A fast learning algorithm for deep belief nets [J]. Neural Computation, 2006, 18:1527-1554.
- [3] Yoshua Bengio, Réjean Ducharme, Pascal Vincent, et al. A Neural Probabilistic Language Model [J]. Journal of Machine Learning Research. 2003, 3(6):1137-1155.
- [4] Tomas Mikolov, Ilya Sutskever, Kai Chen, et al. Distributed Representations of Words and Phrases and their Compositionality [J]. Advances in Neural Information Processing Systems, 2013, 26, 3111-3119.
- [5] Ronan Collobert, Jason Weston, Leon Bottou, et al. Natural Language Processing (almost) from Scratch [J]. Journal of Machine Learning Research, 2011, 12(1):2493-2537.

- [6] Ramn P. Neco, Mikel L. Forcada. Asynchronous translations with recurrent neural nets[C]// Proceedings of International Conference on Neural Networks, Houston, USA, 1997:2535–2540.
- [7] Jiajun Zhang, Chengqing Zong. Deep neural networks in machine translation: an overview [J]. *Intelligent Systems IEEE*, 2015, 30(5):16–25.
- [8] Ilya Sutskever, Oriol Vinyals, Quoc V. Le. Sequence to sequence learning with neural networks[C]// Proceedings of the Neural Information Processing Systems (NIPS 2014), Montreal, Canada, 2014:3104–3112.
- [9] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, et al. Learning phrase representations using rnn encoder–decoder for statistical machine translation. *arXiv preprint/1406.1078v2*, 2014.
- [10] Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau, et al. On the properties of neural machine translation: encoder-decoder approaches[C]// Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation. Doha, Qatar, 2014: 103–111.
- [11] Sebastien Jean, Kyunghyun Cho, Yoshua Bengio. On using very large target vocabulary for neural machine translation[C]// Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL 2015), Beijing, China, 2015:1–10.
- [12] Marcin Junczys-Dowmunt, Tomasz Dwojak and Hieu Hoang. Is Neural Machine Translation Ready for Deployment? A Case Study on 30 Translation Directions. *arXiv:1610.01108v2*, 2016.
- [13] 董晓芳. 藏汉统计机器翻译短语抽取技术研究[D]. 西北民族大学, 2013.
- [14] 位素东. 基于短语的藏汉在线翻译系统研究[D]. 西北民族大学, 2015.
- [15] Xiaodong Yan, Xiaobing Zhao. Research on Sino-Tibetan Machine Translation Based on the Reusing of Domain Ontology [J]. *International Journal of Online Engineering*, 2013, 9(4):105.
- [16] 华却才让. 基于树到串藏语机器翻译若干关键技术研究[D]. 陕西师范大学, 2014.
- [17] Liu Huidan, Zhao Weina, Yu Xin et al. A Chinese to Tibetan Machine Translation System with Multiple Translating Strategies [J]. *Himalayan Linguistics*, 2016, 15(1):148-166.
- [18] 华果才让. 汉藏机器翻译中的藏语动词研究[D]. 青海师范大学, 2014.
- [19] 万福成, 于洪志, 吴玺宏, 等. 面向机器翻译的藏语短语句法研究[J]. *计算机工程与应用*, 2015, 51(13):211-215.
- [20] 王天航. 面向机器翻译的藏语功能组块识别研究[D]. 北京理工大学, 2016.
- [21] 孙萌, 华却才让, 刘凯, 等. 藏文数词识别与翻译[J]. *北京大学学报(自然科学版)*, 2013, 49(1):75-80.
- [22] Karl Weiss, Taghi M. Khoshgoftaar and DingDing Wang. A Survey of Transfer Learning [J]. *Journal of Big Data*, 2016, 3(1):9.
- [23] Barret Zoph, Deniz Yuret, Jonathan May, et al. Transfer learning for low-resource neural machine translation[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP 2016). Austin, USA, 2016:1568–1575.
- [24] Lemao Liu, Masao Utiyama, Andrew Finch, et al. Neural machine translation with supervised attention[C]//Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics. Osaka, Japan, 2016:3093–3102.
- [25] Tong Xiao, Jingbo Zhu, Hao Zhang and Qiang Li. NiuTrans: an Open Source Toolkit for Phrase-based and Syntax-based Machine Translation[C]// Proceedings of the ACL 2012 System Demonstrations. 2012:19-24.
- [26] Dzmitry Bahdanau, KyungHyun Cho and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *Arxiv: 1409.0473v6*, 2014.
- [27] Zhaopeng Tu, Zhengdong Lu, Liu Yang, et al. Modeling coverage for neural machine translation[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics

(ACL 2016), Berlin, Germany, 2016:76-85.

[28] Matthew D. Zeiler. ADADELTA: An Adaptive Learning Rate Method [J]. Computer Science, 2012.

[29] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research [J], 2014, 15:1929–1958.

作者简介: 李亚超 (1986—), 男, 博士研究生, 讲师, 主要研究领域为机器翻译。Email: harry_lyc@foxmail.com; 熊德意 (1979—), 男, 博士, 教授, 主要研究领域为机器翻译。Email: dyxiong@suda.edu.cn; 张民 (1970—), 男, 博士, 教授, 博士生导师, 主要研究领域为机器翻译。Email: minzhang@suda.edu.cn。

作者照片:

作者一 李亚超	作者二 熊德意	作者三 张民
		