



# 基于改进字节对编码的汉藏机器翻译研究

头旦才让<sup>1,2</sup>, 仁青东主<sup>2</sup>, 尼玛扎西<sup>2\*</sup>, 于永斌<sup>3</sup>, 邓权芯<sup>3</sup>

(1. 青海师范大学藏文信息处理教育部重点实验室 西宁 810008; 2. 西藏大学信息科学技术学院 拉萨 850000;  
3. 电子科技大学信息与软件工程学院 成都 610054)

**【摘要】**该文通过改进字节对编码算法,提出了带字数阈值的藏文字节对编码算法,优化了基于注意力机制的汉藏神经机器翻译模型。收集整理了 100 万汉藏句对和 20 万汉藏人名地名词典,训练了汉藏神经机器翻译模型。通过测试和验证,模型的 BLEU 值达到 36.84。该模型的命名实体翻译效果优于已商用汉藏在线翻译系统。同时,该文的神经机器翻译模型已部署于汉藏机器翻译网站,实现了汉藏神经机器翻译系统的应用推广。

**关键词** 注意力机制; 字节对编码; 汉藏神经机器翻译; 命名实体识别  
中图分类号 TP391.1 文献标志码 A doi:10.12178/1001-0548.2020218

## Research on Chinese-Tibetan Machine Translation Model Based on Improved Byte Pair Encoding

THUPTEN Tsering<sup>1,2</sup>, RINCHEN Dhondub<sup>2</sup>, NYIMA Tashi<sup>2\*</sup>, YU Yong-bin<sup>3</sup>, and DENG Quan-xin<sup>3</sup>

(1. Key Laboratory of Tibetan Information Processing, Ministry of Education, Qinghai Normal University Xining 810008;  
2. School of Information Science and Technology, Tibet University Lhasa 850000;  
3. School of Information and Software Engineering, University of Electronic Science and Technology of China Chengdu 610054)

**Abstract** In order to optimize Chinese-Tibetan neural machine translation (NMT) based on attention mechanism, this paper proposes a Tibetan byte-pair encoding algorithm with maximum byte threshold to improve the original byte-pair encoding algorithm. By collecting one million Chinese-Tibetan sentence pairs and dictionaries with 200, 000 Chinese-Tibetan names and places, we train the Chinese-Tibetan NMT model using attention mechanism. Our model has a better translation result in named entity compared with commercial using of Chinese-Tibetan online translation and it achieves 36.84 in bilingual evaluation understudy (BLEU) score. Our work has already deployed in Chinese-Tibetan machine translation system web which will promote the spread and application of Chinese-Tibetan NMT system.

**Key words** attention mechanism; byte pair Encoding; Chinese-Tibetan neural machine translation; named entity recognition

机器翻译是利用计算机自动地将一种自然语言转换为相同含义的另一种自然语言的过程<sup>[1]</sup>。机器翻译在语言形态上分为语音翻译和文本翻译,其历史发展已从基于规则的机器翻译、基于统计的机器翻译发展至基于神经网络的机器翻译(即神经机器翻译)。目前,神经机器翻译已经取代统计机器翻译,成为 Google、微软、百度、搜狗等商用在线机器翻译系统的核心技术<sup>[2]</sup>。

神经机器翻译最早在 2013 年被提出,但是存

在长距离重新排序和梯度爆炸消失等问题,翻译效果不理想<sup>[3]</sup>。2014 年,文献 [4] 提出了编码器和解码器框架,引入了长短时记忆模型,解决了长距离重新排序和梯度爆炸消失等问题,同时神经机器翻译的主要难题变成了固定长度向量问题。2015 年,文献 [5] 将注意力机制应用到机器翻译中,解决了固定长度向量问题。

注意力机制的模型将注意力放在一些相关性高的词上,编码器和解码器之间通过注意力机制连

收稿日期: 2020-05-07; 修回日期: 2021-01-11

基金项目: 国家自然科学基金(61262051); 国家社会科学基金(19XYY021); 国家重点研发计划重点专项(2017YFB1402200); 青海省科技计划(2017-GX-146, 2017-ZJ-767)

作者简介: 头旦才让(1988-), 男, 博士生, 主要从事藏语计算语言学方面的研究。

通信作者: 尼玛扎西, E-mail: niqiongda@163.com

接<sup>[6]</sup>,在翻译目标单词时检测其与源端语句相关的部分,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,是目前神经机器翻译的主流方法,应用广泛。2018年,文献[7]使用自注意力机制来增强序列标注模型的全局表示能力,从序列标注任务端减少汉文分词对随后翻译对影响的方法。在此基础上,文献[8]提出了格到序列的神经机器翻译模型。通过实验,该模型在翻译性能上显著优于传统的基于注意力机制的序列到序列基线系统。同年,文献[9]提出简单循环单元的注意力机制模型。2019年,文献[10]提出了一个稀疏注意力模型,解决了注意力权重分布问题。2020年,文献[11]提出了一种深度注意力模型,大大提高了系统翻译的忠实度。综上,基于注意力机制的模型成为目前神经机器翻译领域的主流模型。

近几年,研究人员在基于注意力机制模型的基础上,利用不同的方法进行了汉藏藏汉神经机器翻译研究。2017年,文献[12]基于注意力机制和迁移学习方法,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。2018年,文献[13]将注意力机制模型应用于汉藏机器翻译任务中,实现了汉藏书面语料和口语语料的神经机器翻译。2019年,文献[14]在transformer模型上,运用百万句子单语数据大规模迭代式回译策略,实现了藏汉神经机器翻译模型,文献[15]也使用transformer实现了藏汉神经机器翻译模型,并将藏语单语语言模型融合到藏汉神经机器翻译中。

汉藏机器翻译中的命名实体处理一直是最难以突破的一个技术环节,为处理命名实体、同源词、外来词和形态复杂的词,本文在模型训练时,改进藏文字节对编码算法,优化了基于注意力机制的翻译模型,使得汉藏神经机器翻译效果更加准确。

## 1 基于注意力机制和改进字节对编码模型

藏文是拼音文字,音节之间用分隔符隔开,词与词之间没有明确的分隔符,再者藏文有格助词、助动词等汉语语法不具有的语法单元,所以对应的翻译句子长度比汉语长。

### 1.1 注意力机制模型

为了解决长句的翻译问题,本文系统(阳光汉藏机器翻译系统V2)采用了基于注意力机制的神经机器翻译模型。

对于给定的平行句对 $x = \{x_1, x_2, \dots, x_{l_s}\}$ 和 $y = \{y_1^*,$

$y_2^*, \dots, y_n^*\}$ ,使用RNN对 $x$ 进行编码,如形式:

$$h_i = \begin{bmatrix} h_{i1} \\ h_{i2} \end{bmatrix} = \begin{bmatrix} \text{GRU}_1(x_i, h_{i-1}) \\ \text{GRU}_2(x_i, h_{i+1}) \end{bmatrix} \quad (1)$$

式中, $\text{GRU}_1$ 和 $\text{GRU}_2$ 是两个门控循环单元,从两个方向循环地对 $x$ 编码,然后拼接每个词的输出状态,即 $h = \{h_1, h_2, \dots, h_{l_s}\}$ 。

解码器依次生成藏文词 $y_j$ :

$$p = (y_j | \{y_1, y_2, \dots, y_{j-1}, x\}) = g(y_{j-1}, s_j, c_i) \quad (2)$$

通过最大化藏文词汇的似然来优化整个翻译模型:

$$p = (y_j | y_{j-1}^*, s_j, c_j) \propto \exp(o(y_{j-1}^*, s_j, \alpha_j)) \quad (3)$$

式中, $o(y_{j-1}^*, s_j, \alpha_j)$ 是非线性函数; $y_{j-1}^*$ 是第 $j-1$ 步的参考词语; $s_j$ 是 $j$ 时刻循环神经网络的隐藏状态:

$$s_j = \text{GRU}(y_{j-1}^*, s_{j-1}, \alpha_j) \quad (4)$$

式中,GRU是非线性函数; $\alpha_j$ 是每次解码动态更新的汉文上下文表示,依赖于汉文编码序列 $h = \{h_1, h_2, \dots, h_i\}$ 即每个汉文词语状态的加权和:

$$\alpha_j = \sum_{i=1}^{l_s} \alpha_{i,j} h_i \quad (5)$$

式中, $\alpha_{i,j}$ 是第 $i$ 个汉文词语与第 $j$ 个藏文词语之间的对齐概率:

$$\alpha_{i,j} = \exp(e_{i,j}) / \sum_{i=1}^{l_s} \exp(e_{i,j}) \quad (6)$$

$$e_{i,j} = f(s_{j-1}, h_i) \quad (7)$$

式中, $e_{i,j}$ 是对齐模型,通过前馈神经网络,计算出 $i$ 时刻生成的藏文词与第 $j$ 个汉文词的匹配程度。图1为基于注意力机制的神经网络翻译模型,其中, $y_j$ 表示模型在第 $j$ 步所预测词语的概率分布,极大似然估计(MLE)表示计算损失的方法<sup>[6]</sup>。

### 1.2 改进的藏文字节对编码算法

本文为了提高模型的稳定性和翻译的准确率,训练时使用了一种带字数阈值的藏文字节对编码算法,优化了翻译模型。首先简单描述一下原始字节对编码算法。

#### 1.2.1 字节对编码

字节对编码(byte pair encoding, BPE)指通过自动发现稀疏词,建立基于稀疏词的神经网络翻译模型<sup>[7]</sup>,即通过构造高频的字符片段,将稀疏词拆分为合适的子词,使得这些子词在语料中的出现次数

足够高, 从而进行训练, 得到最优的翻译模型, BPE 算法流程图如图 2 所示。

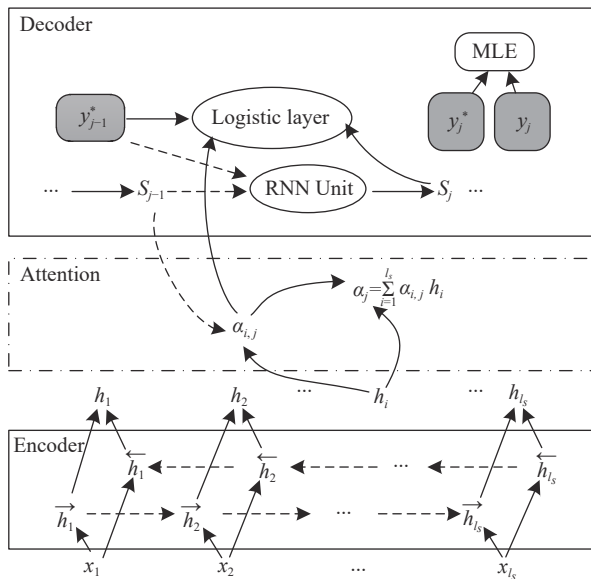


图 1 基于注意力机制的神经网络翻译模型

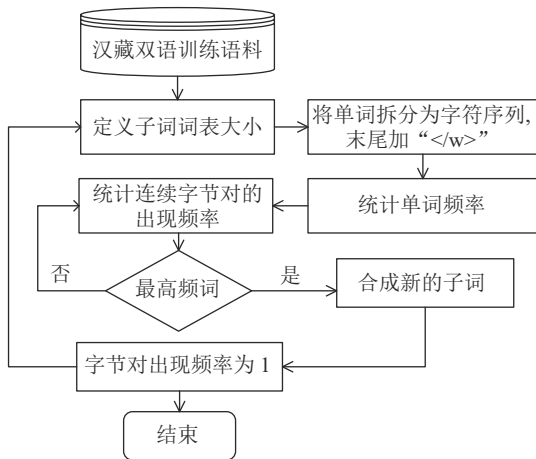


图 2 原始 BPE 算法流程图

### 1.2.2 改进的藏文字节对编码算法

BPE 算法不需要对语料中的词汇进行任何处理, 由于藏文的词与词之间没有空格, 所以形式上不存在单词边界的问题, 所以改进的 BPE 算法中不需要词头词尾标记符号 “</w>”。另外, 通过分析藏文和汉文使用原始 BPE 算法之后生成的子词长度分布情况, 发现藏文和汉文对应最大子词长度分别为 39 和 21, 在长度峰值上, 藏文子词明显大于汉文子词。基于此, 在分词粒度较大的藏文部分, 本文提出了一种改进的 BPE 算法, 使用长度阈值控制藏文子词, 合并出现频率略低但长度更适合的子词。其算法流程如图 3 所示。

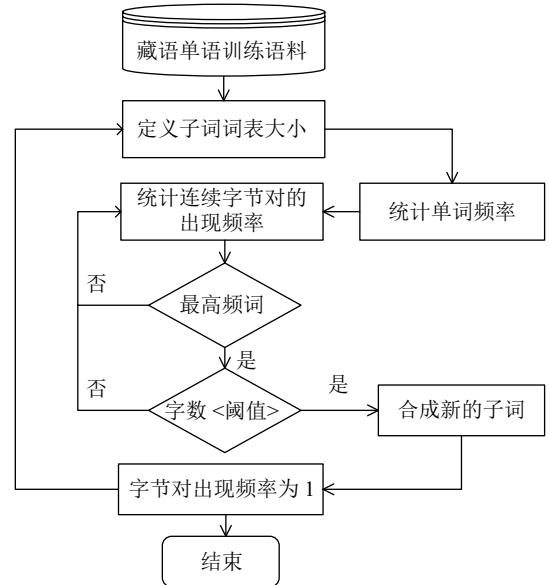


图 3 改进的 BPE 算法流程图

通过对比图 2 和图 3 的流程, 相对于原始的 BPE 算法, 改进的藏文 BPE 算法设置子词字数阈值, 不提取字数超过该阈值的子词。实验表明, 当设置的字数阈值为 23 时, 效果提升最为明显。

## 2 实验结果对比与分析

### 2.1 数据集

本文使用西藏大学和青海师范大学建设的 107 万汉藏句对, 经纠正错峰句对、检查拼写错误、校对断句错字和过滤重复句子, 最终将 100 万句对作为训练语料, 其中 80 万句对为新闻和法律题材, 20 万句对为其他领域语料; 测试集和验证集各 1000 句对。此外, 为了提高翻译效果, 建立了 15 万词条的汉藏地名词典和 5 万条的汉藏人名词典作为辅助工具。

### 2.2 参数设置

模型训练中, 通过反复调整参数, 最终获得最优的模型参数, 具体描述如下。

所有的模型参数都使用随机梯度下降算法进行优化<sup>[18]</sup>, 学习率使用 Adadelta 算法<sup>[19]</sup>进行自动调节; 训练语料汉文端和藏文端保留的最大句长为 50 词; 汉文和藏文词向量维度为 512; 编码器和解码器中循环神经单元的隐状态和输出状态均设为 512; 模型最终输出层采用 dropout 策略, dropout 设置为 0.5。

测试时本文使用束搜索算法进行解码<sup>[20]</sup>, 搜索过程中束大小设置为 10。汉文端和目藏文端的词表大小设置为 15 万, 覆盖率 100%。语料训练的轮

数最大为 50 轮。模型的超参数需要根据训练数据量来设置,通常数据量越大,模型的超参数如 `enc_hid_size`、`dec_hid_size` 可以适当放大。

### 2.3 结果与分析

本文实现了基于 PyTorch 框架的注意力机制汉藏神经网络机器翻译系统,循环单元采用门控循环单元,评测指标使用了 BLEU4,采用基于字的评测方法。

汉藏双语语料库进行分词时,汉文分词使用了感知机、Hanlp 和 BPE。藏文分词使用了西藏大学开发的基于 Perceptron+CRF 模型的藏文分词系统<sup>[21]</sup>和改进的藏文 BPE 算法。

本文在测试集、验证集一致的前提下,做了 6 组实验。训练模型时,除实验 1 未使用 BPE 算法

外,其他实验的译文和原文都使用了 BPE 算法和改进的藏文 BPE 算法。6 组实验的验证集和测试集情况如下表 1 所示。

表 1 实验数据集

实验数据	句子数	词语总数	人名地名数
验证集	1000	27863	3174
测试集	1000	28333	3243

为了体现命名实体的翻译效果,验证集和测试集的每个句子中至少包含了一个地名或人名,因此通过该测试集得到的结果应当具有一定的通用性且能够说明模型具有的泛化能力。

实验数据、方法、结果如表 2 所示。

表 2 实验数据、方法、结果

实验数据	实验1	实验2	实验3	实验4	实验5	实验6
训练集/万	80	80	80	80	100	100
题材	新闻、法律	新闻、法律	新闻、法律	新闻、法律	新闻、法律、其他	新闻、法律、其他
汉文分词	感知机	感知机	Hanlp	BPE	BPE	BPE
藏文分词	Perceptron+CRF	BPE	BPE	BPE	BPE	改进BPE
人名、地名/万	0	0	0	20	20	20
验证集BLEU值	30.46	31.56	31.84	33.87	34.09	36.84

在实验中,第一组实验使用了基于平均感知机的汉文分词方法<sup>[22]</sup>,目标端没有使用 BPE 操作,BLEU 值最低,只有 30.46。由于该汉文分词器的分词粒度较大,而藏文分词器的分词粒度较小,分词粒度的差异导致翻译效果不够好,另外由于汉文分词器的算法问题,导致前端的查询速度很慢。

第二组实验的汉文分词方法和第一组实验方法一样,但是对藏文端使用了 BPE 算法。虽然分词粒度的差异导致翻译效果不够好,但是解决了一些稀疏词和低频词翻译问题,BLEU 值提高了 1.10。

为了解决汉藏分词粒度的耦合性,第三组实验使用了 Hanlp 汉文分词器<sup>[23]</sup>,因为其分词粒度较小,也提供了更友好的 Python 访问接口;同时还提供了词性标注等更高级的功能,为后续对模型的优化提供了更好的扩展性,另外查询速度也得到提升。使用 BPE 操作,提供了更细的分词粒度,致使汉文分词和藏文分词粒度契合,BLEU 值达到了 31.84。

为了进一步提高翻译效果,第四组实验的汉文

分词使用了 BPE 算法,并增加了 20 万条的人名地名词典作为训练辅助工具,解决了命名实体翻译问题,BLEU 值提高了 2.03,达到了 33.87。

第五组实验在第四组实验的基础上,增加了 20 万条的人物传记、小说和口语等双语语料,由于语料题材的多样性,BLEU 值仅提高了 0.22,达到 34.09。

上述五组实验结果得知,如果仅仅使用原始 BPE 算法,BLEU 值提高 1.06,如果增加命名实体词典,BLEU 值也仅仅提高 2.03。为了体现改进的藏文 BPE 算法的性能,第六组实验在第五组实验的基础上,在藏文端使用了本文提出的 BPE 改进算法,这时 BLEU 值提高了 2.75,达到了 36.84。

本文根据模型在验证集上的 BLEU 值分数,当 BLEU 值出现震荡或者模型过拟合,BLEU 值开始出现明显下降时,停止训练。

### 2.4 与已商用化的汉藏翻译对比

在已有的汉藏神经机器翻译研究<sup>[13-15]</sup>中,每个模型采用的训练语料和测试集不同,且目前汉藏机



应时间。经简单测试,加载神经翻译模型和分词器只需要 50 ms。通过后端设计,还实现了 Android

前端和网页前端。汉藏神经机器翻译系统化总体架构如图 5 所示。

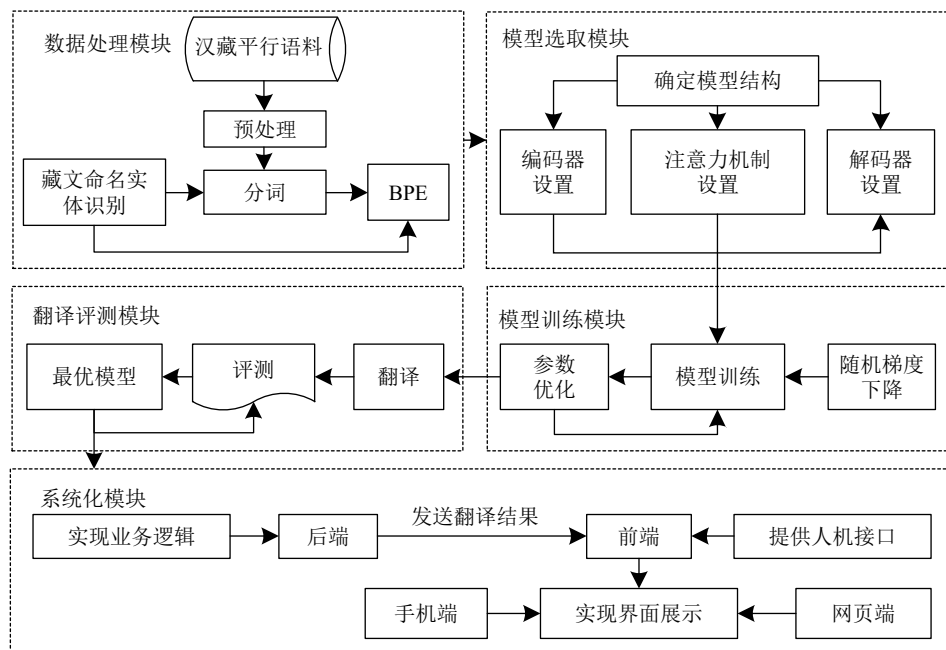


图 4 汉藏神经机器翻译模型总体流程

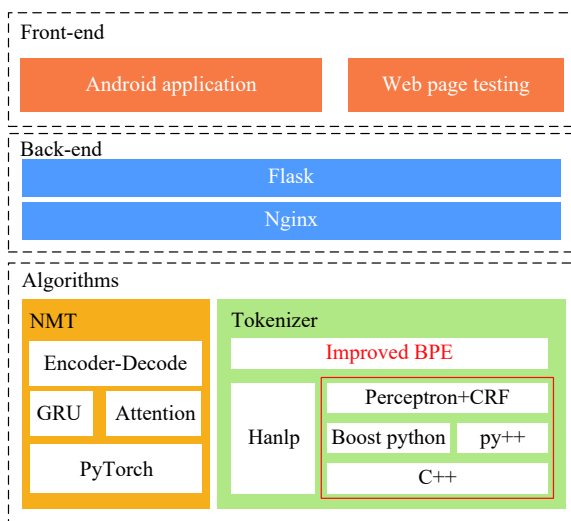


图 5 汉藏神经机器翻译系统总体架构

## 4 结束语

本文利用 100 万汉藏句对和 20 万汉藏人名地名词条,进行了基于注意力机制的神经机器翻译实验,并提出了一种改进的 BPE 算法,用以协调原始 BPE 得到的藏文粒度大于汉文粒度的情况,将 BLEU 提升了 2.75%,减少了过度翻译、翻译不充分的问题,提升了命名实体翻译效果。设计实现了基于注意力机制和改进字节对编码的汉藏神经机器翻译模型,部署在阳光汉藏机器翻译网站,实现了该汉藏神经机器翻译系统的应用推广。本文汉藏机

器翻译系统的模型具有语言无关性,完全可以应用到藏汉神经机器翻译研究中。

由于汉藏神经机器翻译目前缺乏大规模双语数据,而藏语单语语料比较充足,所以下一步将利用格到序列、半监督和无监督方法提升翻译效果。

共享和开放是计算语言学(自然语言处理)研究的发展趋势,该工作在汉英机器翻译技术领域获得了很好的进展,免费开放了一些汉英英汉双语平行语料,使得汉英机器翻译技术具有可比性和竞争性。汉藏机器翻译研究语言资源较少,没有公开的语料,而且资源问题一直是困扰神经机器翻译研究和产业化的首要问题<sup>[29]</sup>。为此,我们开放了部分实验数据和藏文地名词典(获取地址:<https://github.com/toudancairang/Tibetan-Computational-linguistics/tree/master>),希望吸引更多的人参与其中,建立藏文资源开放平台,推动藏语计算语言学(藏语自然语言处理)研究,促进中文信息处理技术的整体发展。

致谢:该模型的构建得到了中科院计算所自然语言处理实验室同仁的大力支持。

## 参 考 文 献

[1] 李沐,刘树杰,张冬冬,等.机器翻译[M].北京:高等教育

- 出版社, 2018.
- LI Mu, LIU Shu-jie, ZHANG Dong-dong, et al. Machine translation[M]. Beijing: Higher Education Press, 2018.
- [2] 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.
- LIU Yang. Recent advances in neural machine translation[J]. *Journal of Computer Research Development*, 2017, 54(6): 1144-1149.
- [3] KALCHBRENNER N, BLUNSOM P. Recurrent continuous translation models[C]//Proceedings of Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, PA: ACL, 2013: 1700-1709.
- [4] CHO K, VAN M B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). [S. l.]: ACM, 2014: 1724-1734.
- [5] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[EB/OL]. [2019-11-01]. <http://de.arxiv.org/pdf/1409.0473>.
- [6] JUNCZYS-DOWMUNT M, DWOJAK T, HOANG H. Is neural machine translation ready for deployment? A case study on30translation directions[EB/OL]. [2019-11-01]. <https://arxiv.org/pdf/1610.01108.pdf>.
- [7] TAN Zhi-xing, SU Jin-song, WANG Bo-li, et al. Lattice-to-sequence attentional neural machine translation models[J]. *Neurocomputing*, 2018, 284: 138-147.
- [8] 谭知行. 格到序列神经机器翻译研究[D]. 厦门: 厦门大学, 2018.
- TAN Zhi-xing. Research on lattice-to-sequence neural machine translation[D]. Xiamen: Xiamen University, 2018.
- [9] 张文, 冯洋, 刘群. 基于简单循环单元的深层神经网络机器翻译模型[J]. *中文信息学报*, 2018, 32(10): 36-44.
- ZHANG Wen, FENG Yang, LIU Qun. Deep neural machine translation model based on simple recurrent units[J]. *Journal of Chinese Information Processing*, 2018, 32(10): 36-44.
- [10] ZHANG J J, ZHAO Y, LI H R, et al. Attention with sparsity regularization for neural machine translation and summarization[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2019, DOI: 10.1109/TASLP.2018.2883740.
- [11] ZHANG B, XIONG D Y, SU J. Neural machine translation with deep attention[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, DOI: 10.1109/TPAMI.2018.2876404.
- [12] 李亚超, 熊德意, 张民, 等. 藏汉神经网络机器翻译研究[J]. *中文信息学报*, 2017, 31(6): 103-109.
- LI Ya-chao, XIONG De-yi, ZHANG Min, et al. Research on Tibetan-Chinese neural network machine translation[J]. *Journal of Chinese Information Processing*, 2017, 31(6): 103-109.
- [13] 李博涵, 刘汇丹, 龙从军. 深度学习在汉藏机器翻译中的应用研究[J]. *民族翻译*, 2018(3): 51-58.
- LI Bo-han, LIU Hui-dan, LONG Cong-jun. Application of deep learning in Chinese-Tibetan machine translation[J]. *Ethnic Translation*, 2018(3): 51-58.
- [14] 桑杰端珠. 稀疏资源条件下的藏汉机器翻译研究[D]. 西宁: 青海师范大学, 2019.
- SANGJIE Duanzhu. Research on Tibetan-Chinese machine translation under sparse resources[D]. Xining: Qinghai Normal University, 2019.
- [15] 慈颖嘉措, 桑杰端珠, 孙茂松, 等. 融合单语语言模型的藏汉机器翻译方法研究[J]. *中文信息学报*, 2019, 33(12): 61-66.
- CIZHEN Jiacao, SANGJIE Duanzhu, SUN Mao-song, et al. Tibetan-Chinese machine translation based on tibetan language model enhanced transform[J]. *Journal of Chinese Information Processing*, 2019, 33(12): 61-66.
- [16] RANZATO M, CHOPRA S, AULI M, et al. Sequence level training with recurrent neural networks[EB/OL]. [2019-10-12]. <http://de.arxiv.org/pdf/1511.06732>.
- [17] SENNRICH R, HADDOW B, BIRCH A. Neural machine translation of rare words with subword units[C]//Proc of the 54th ACL. Stroudsburg, PA: ACL, 2016: 1715-1725.
- [18] 史加荣, 王丹, 尚凡华, 等. 随机梯度下降算法研究进展[J/OL]. *自动化学报*, [2020-03-11]. <https://doi.org/10.16383/j.aas.c190260>.
- SHI Jia-rong, WANG Dan, SHANG Fan-hua, et al. Research advance on stochastic gradient descent algorithms[J/OL]. *ACTA Automation SINICA*, [2020-03-11]. <https://doi.org/10.16383/j.aas.c190260>.
- [19] ZEILER M D. ADADELTA: an adaptive learning rate method[EB/OL]. [2019-11-10]. <http://export.arxiv.org/pdf/1212.5701>.
- [20] WISEMAN S, RUSH A. Sequence-to-sequence learning as beam-search optimization[C]//Proc of EMNLP. Stroudsburg, PA: ACL, 2016: 1296-1306.
- [21] 头旦才让, 仁青东主, 尼玛扎西. 基于CRF的藏文地名识别技术研究[J]. *计算机工程与应用*, 2019, 55(18): 111-115.
- THUPTEN Tsering, RINCHEN Dhondub, NYIMA Tashi. Research on Tibetan location name recognition technology under CRF[J]. *Computer Engineering and Applications*, 2019, 55(18): 111-115.
- [22] 仁青东主, 头旦才让, 尼玛扎西. 汉藏机器翻译研究综述[J]. *中国藏学*, 2019(4): 222-226.
- RINCHEN Dhondub, THUPTEN Tsering, NYIMA Tashi. A summary of research on Chinese-Tibetan machine translation[J]. *China Tibetology*, 2019(4): 222-226.
- [23] 汤连杰. HanLP2.0[J]. *软件和集成电路*, 2019(8): 95.
- YANG Lian-jie. HanLP2.0[J]. *Software and Integrated Circuit*, 2019(8): 95.
- [24] LUONG M, SUTSKEVER I, LE Q, et al. Addressing the rare word problem in neural machine translation[C]//Proc of the 53rd ACL. Stroudsburg, PA: ACL, 2015: 11-19.
- [25] JEAN S, CHO K, MEMISEVIC R, et al. On using very large target vocabulary for neural machine translation[C]//Proc of the 53rd ACL. Stroudsburg, PA: ACL, 2015: 1-10.