

# 基于多语言联合训练的汉-英-缅神经机器翻译方法

满志博<sup>1,2</sup>, 毛存礼<sup>\*1,2</sup>, 余正涛<sup>1,2</sup>, 李训宇<sup>1,2</sup>, 高盛祥<sup>1,2</sup>, 朱俊国<sup>1,2</sup>

1. 昆明理工大学, 信息工程与自动化学院, 昆明, 650500

2. 昆明理工大学, 云南省人工智能重点实验室, 昆明, 650500

270004294@qq.com, maocunli@163.com, ztyu@hotmail.com

1242041057@qq.com, gaoshengxiang.yn@foxmail.com, jgzhu@mtlab.hit.edu.cn

## 摘要

多语言神经机器翻译是解决低资源神经机器翻译的有效方法, 现有方法通常依靠共享词表的方式解决英语、法语以及德语相似语言之间的多语言翻译问题。缅甸语属于一种典型的低资源语言, 汉语、英语以及缅甸语之间的语言结构差异性较大, 为了缓解由于差异性引起的共享词表大小受限制的问题, 提出一种基于多语言联合训练的汉英缅神经机器翻译方法。在Transformer框架下将丰富的汉英平行语料与汉缅、英缅的语料进行联合训练, 模型训练过程中分别在编码端和解码端将汉英缅映射在同一语义空间降低汉英缅语言结构差异性对共享词表的影响, 通过共享汉英语料训练参数来弥补汉缅数据缺失的问题。实验表明在一对多、多对多的翻译场景下, 提出方法相比基线模型的汉-英、英-缅以及汉-缅的BLEU值有明显的提升。

**关键词:** 汉语-英语-缅甸语; 低资源语言; 多语言神经机器翻译; 联合训练; 语义空间映射; 共享参数

## Chinese-English-Burmese Neural Machine Translation Method Based on Multilingual Joint Training

Zhibo Man<sup>1,2</sup>, Cunli Mao<sup>\*1,2</sup>, Zhengtao Yu<sup>1,2</sup>, Xunyu Li<sup>1,2</sup>, Shengxiang Gao<sup>1,2</sup>, Junguo Zhu<sup>1,2</sup>

1. Faculty of Information Engineering and Automation, Kunming University of Science and Technology  
Kunming 650500, China

2. Yunnan Key Laboratory of Artificial Intelligence, Kunming University of Science and Technology  
Kunming 650500, China

270004294@qq.com, maocunli@163.com, ztyu@hotmail.com

1242041057@qq.com, gaoshengxiang.yn@foxmail.com, jgzhu@mtlab.hit.edu.cn

## Abstract

Multilingual Neural Machine Translation is an effective method to solve low-resource Neural Machine Translation. Existing methods usually rely on a shared vocabulary to solve the problem of multilingual translation between similar languages in English, French, and German. Burmese language is a typical low-resource language. The language structure between Chinese, English and Burmese is quite different. In order to alleviate the problem of the limited size of the shared vocabulary caused by the difference, a multilingual combination is proposed. Trained Chinese-English-Burmese neural machine translation method. The rich Chinese-English parallel corpus and the Chinese-Burmese and English-Burmese corpus are jointly trained under the Transformer framework. During the model training process, the Chinese and English-Burmese maps are mapped to the same semantic space on the encoding side and the decoding side to

\*毛存礼(通信作者):maocunli@163.com

国家自然科学基金重点项目(61732005); 国家自然科学基金(61662041, 61761026, 61866019, 61972186); 云南省自然科学基金重点项目(2019FA023); 云南省中青年学术和技术带头人后备人才项目(2019HB006)

reduce the difference in Chinese-English-Burmese language structure. The influence of sex on the shared vocabulary is to compensate for the lack of Chinese-Burmese data by sharing Chinese and English material training parameters. Experiments show that in one-to-many and many-to-many translation scenarios, the proposed method has a significant improvement over the baseline models of Chinese-English, English-Burmese, and Chinese-Burmese BLEU.

**Keywords:** Chinese-English-Burmese , Low Resource Language , Multilingual Neural Machine Translation , Joint training , Semantic space mapping , Shared parameters

## 1 引言

目前，多语言神经机器翻译（Multilingual Neural Machine Translation, MNMT）(Aharoni et al., 2019; Wang et al., 2018; Lee et al., 2017; Wang et al., 2019)在低资源机器翻译方面取得了较好的效果，与标准双语翻译的模型相比，其通过构建多种语言之间的联合训练(Caruana, 1997)模型，能够共享资源丰富语言的模型参数来提升资源稀缺语言机器翻译性能(Lignos et al., 2019)，但是，目前的方法主要利用在相似的语言之间，例如，英语、德语、法语等，这些语言之间有大量的“同源词”或者相同的子词词根，在进行多语言词汇共享词表时，这些语言的词表会有很多相同的词语可以得到共享。

基于统计的缅甸语机器翻译方法(Nwet et al., 2011a; Nwet et al., 2011b)依赖于大规模的汉缅平行语料，缅甸语是一种典型的资源稀缺型语言，多语言机器翻译可以有效解决汉缅数据缺乏的问题，但是，汉语、缅甸语以及英语三种语言的结构差异较大，没有相同的词语或者词根，共享词表时由于受到词表大小的限制会造成许多词语无法在词表中出现，导致三种语言的语义空间无法对齐，如图1，汉语“他”对应两个缅甸语单词，英语单词“very much”对应一个缅甸语单词，利用汉英缅数据进行多语言模型训练时，会出现无法准确的在编码过程中将三种语言的词语对齐的问题。



Figure 1: 汉语、英语、缅甸语互译句子示例

针对以上问题，提出的基于多语言联合训练的汉英缅神经机器翻译方法是在Johnson等人(2017)提出模型的基础上进行改进，在Transformer框架下，利用丰富的汉英平行语料与汉缅、英缅的语料进行联合训练，在编码、解码过程中，将汉语、英语以及缅甸语三种语言进行语义映射，缩小三种语言之间的语义距离，同时，丰富三种语言共享词表中的词语。在Transformer的框架中，每个归一化和残差连接层中，将所有语言对的训练参数进行共享学习，提升汉缅神经机器翻译性能。

本文的贡献如下：

(1)通过在Transformer的编码-解码端的汉语、英语以及缅甸语三种语言词嵌入映射共享语义空间解决汉英缅由于差异性较大引起的词表共享受限制的问题。

(2)将汉英缅三种语言对的训练参数进行共享，利用汉英高资源语言对的训练参数弥补汉缅低资源语言对训练不充分的问题。

本文的第2节介绍了针对低资源语言以及缅甸语的机器翻译相关工作；第3节介绍了多语言神经机器翻译的研究背景；第4节描述了基于多语言联合训练的神经机器翻译方法；第5节通过

在一对多和多对一两种翻译场景下进行实验对比证明本文方法的优势；第6节对全文进行总结并指出进一步的研究工作。

## 2 相关工作

本文将相关工作的分为两类，分别是缅甸语的神经机器翻译以及低资源语言的神经机器翻译方法研究：

针对于缅甸语的机器翻译的研究工作：目前，由于缅甸语与其他语言的双语资源较少，针对于缅甸语的机器翻译的研究工作较少，Nwet等人(Nwet et al., 2011b)提出一种通过缅甸语-英语词对齐的英缅统计机器翻译方法，由于这种方法在一定程度上受到词表大小的限制，Nwet等人(Nwet et al., 2011a)进一步提出通过扩展英缅双语的平行语料的机器翻译方法。以上针对于缅甸语的机器翻译研究都是基于统计的方式，基于统计的方式需要大规模的双语词典或者是双语平行语料，缅甸语是一种典型的资源稀缺型语言，利用统计的方式不能完全适用于缅甸语。

缅甸语是一种资源稀缺型语言，解决低资源语言的神经机器翻译的方法主要包括：

(1)基于枢轴的神经机器翻译方法(Kim et al., 2019)：借助枢轴语言构建低资源的神经机器翻译模型，提高神经机器翻译性能。英语-缅甸语以及缅甸语-汉语本身的语料就缺乏并且翻译性能不佳，使用枢轴的方式对缅甸语机器翻译性能提高并不明显。

(2)基于迁移学习的神经机器翻译方法(Lakew et al., 2018; Dabre et al., 2019; Firat et al., 2016b; Sachan and Neubig, 2018)：借助预训练的思想，将资源丰富语言对训练模型或参数迁移到低资源语言对的训练过程中。缅甸语和其他语言之间的语法差异性极大，直接利用迁移学习的思想将模型迁移到缅甸语上的效果性能不佳。

(3)近些年来，利用多语言神经机器翻译联合训练逐渐成为解决低资源机器翻译的主流方法。例如，对所有源语言使用相同编码，目标语言使用不同的解码器。Dong等人(2015)在一对多的翻译场景下，提出将多语言翻译过程中的源语言的编码器共享，为每个目标语言分配不同的解码器的方法。Lee等人(2017)在多对一的翻译场景下，提出在编码器端采用字符级输入并将多种源语言共享编码器的方法。对所有源语言到目标语言的语言对使用不同的编码器、解码器，分别为每种语言对训练翻译模型。例如，Firat等人(2016a)提出一种基于共享注意力机制的多路、多对多的神经机器翻译方法。Zoph等人(2016)提出将注意力机制进行联合的多到一的神经机器翻译方法。以上的两类方法一方面需要每种语言专用的编码器或解码器，从而限制了翻译模型的泛化性。另一方面，训练不同的语言对的翻译模型在一定程度上也加大了模型训练的成本。对所有源语言、目标语言均使用相同的编码器、解码器。例如，Ha等人(2016)和Johnson等人(2017)训练了用于多语言翻译的单个NMT模型，使用目标语言符号作为翻译方向的指导。这种方法将不同的语言合并为一个联合表示空间，但是忽略了语言的多样性。Zhang等人(2019)提出了一种具有语言敏感机制的多语言神经机器翻译方法。在多语言神经机器翻译的训练过程中增加敏感机制的表示，达到共享的同时又不丢失语言本身的多样性。

由于汉英缅三种语言的差异性较大，不具备相同的词语或者词根。在多语言神经机器翻译框架下，仅仅利用共享词表的方法会在一定程度上限制三种语言的词汇表征能力，以上方法还不能完全应用于缅甸语的机器翻译问题。为此，本文在(Johnson et al., 2017; Wang et al., 2019)基础上，提出将汉英缅三种语言进行语义空间映射降低语言差异性，解决三种语言由于语言结构差异导致的词表受限制的问题，在公共语义空间中共享多语言联合训练模型参数，来提升缅汉机器翻译的性能。

## 3 多语言神经机器翻译研究背景

### 3.1 编码器-解码器

本文的主要工作是基于Transformer(Vaswani et al., 2017)的架构基础上进行多语言的神经机器翻译。在本文中选取Johnson等人(2017)以及Ha等人(2016)提出的方法作为实验中的基准模型。在多语言的神经机器翻译框架(Johnson et al., 2017)下，给定一句包含 $n$ 个单词的源语言的句子为 $x = (x_1, x_2, \dots, x_{|n|})$ ，其包含目标语言的参考译文为 $y = (y_1, y_2, \dots, y_{|m|})$ ，以及对应的目标语言的标签为 $T$ ，例如，2EN,2ZH以及2MY。在整个编码-解码的结构中主要包括：源语言、目标语言词嵌入、编码器、解码器和输出层。在词嵌入层每个语言和目标语言的多语言神经机器翻译的单词都会映射为一个向量矩阵 $W_E$ 来表示。编码器-解码器(Sutskever et al., 2014)的表示

如下:

$$H_{enc} = \text{Encoder}([T, x]) \tag{1}$$

$$S_{dec} = \text{Decoder}([y, H_{enc}]) \tag{2}$$

其中,  $H_{enc} \in R^{|x| \times d}$ ,  $S_{dec} \in R^{|y| \times d}$ 分别表示编码器和解码器输出,  $d$ 表示模型的尺寸。

与NMT模型相比, 多语言模型一定程度上是将多个语言对进行联合训练, 并将所有源语言和目标语言应用统一的神经机器翻译框架。针对统一的编码器-解码器的框架, 在整个翻译的迭代过程中的目标函数:

$$L_{m-T}(D; \theta) = \sum_{l=1}^L \sum_{d=1}^{|D_l|} \sum_{T=1}^M \log P(y_t^l | x^l, y_{<t}^l; H_{enc}, S_{dec}, \theta_{attention}) \tag{3}$$

其中,  $L$ 表示联合训练语言的句子对个数,  $M$ 表示目标语言句子的长度,  $P(y_t^l | x^l, y_{<t}^l)$ 表示第 $l$ 个翻译对中第 $d$ 个句子的第 $t$ 个单词的翻译概率。 $\theta_{attention}$ 表示训练过程中的注意力机制参数。

### 3.2 Transformer

Transformer是一种基于编码器-解码器的框架, 由多个网络层堆叠而成。其中, 编码器是6个相同的堆栈层组成, 每个层包含一个自注意力机制层和基于词语位置的前馈子层, 利用位置信息可以较好的将句子中每个词语的位置显式的加入到了神经网络中。解码器也遵循类似的结构。除了以上的自注意力机制层以及基于词语以及基于位置的前馈神经网络层以外, 在解码器自注意力机制后是多头交叉的注意力机制网络。在编码器中,  $e_{i,t}$ 是源语言向量表征与目标语言向量表征的相似度分数。

$$e_{i,t} = \frac{1}{\sqrt{d}} q_i k_t^T \tag{4}$$

$$q_i = W_q \cdot H_{enc}^i, k_t = W_k \cdot S_{dec}^t \tag{5}$$

在机器翻译中, Transformer将编码器中的隐状态视为一组键 (Key) 值 (Vaule) 对的集合。 $W_q$ 和 $W_k$ 表示的是交叉注意力机制的训练参数,  $d$ 表示模型的尺寸。

## 4 基于多语言联合训练的汉-英-缅神经机器翻译模型

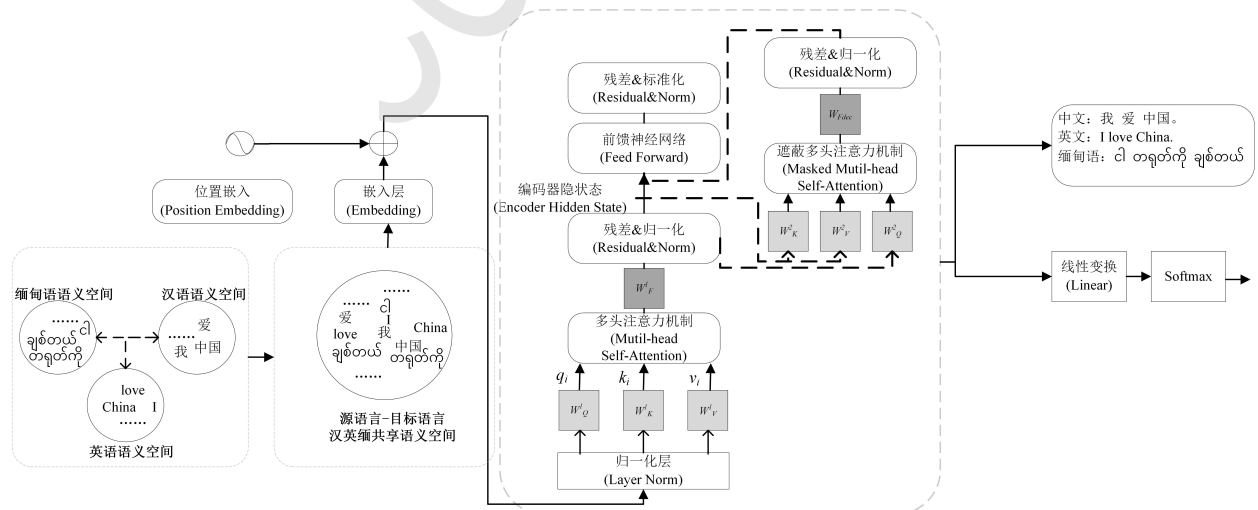


Figure 2: 基于多语言联合学习的汉英缅神经机器翻译模型框架

针对于汉英缅三种语言各自语义空间不同以及翻译训练过程中参数共享导致的翻译模型的性能不佳的问题。本文在Transformer的架构下进行汉英缅多语言神经机器翻译, 本文的具体模型架构如图2所示。

#### 4.1 汉英缅多语言语义空间映射

传统的多语言机器翻译方法都是采用共享词表。但是，当多种差异性较大的语言进行共享词表时，在词表中每种语言可用的词表内容会变小，因此，每种语言的词汇表征能力会相应地降低，这就是当相似语言进行共享此表示效果会比差异性较大的语言共享词表效果更好的原因。如图2所示，汉语、英语以及缅甸语三种语言之间的差异性较大。汉语、英语都属于SVO语序的语言，其语句特点是有中心语的后置和前置。缅甸语则属于SOV语序（和其他语序为宾语先于谓语的语序）语言倾向于将形容词放在名词前面，此类顺序称之为中心语后置。这就意味着，在缅甸语和英语的句法构成中存在很大不同，而且缅甸语的虚词和助词在英语中却不存在对应的词。因此会出现不同语言的相同含义的词语之间无法对应，另外还存在一种语言中的单词可能被翻译成另一种语言的一个或多个单词等问题，以上的语言特点都会导致三种语言的语义空间较为独立(Win, 2011)。

针对以上的问题，在汉英缅多语言联合训练的神经机器翻译任务中，我们在三种语言共享词表之前，首先将三种语言词嵌入进行共享，这样可以保证我们得到的是三种语言共享词嵌入的语义空间而不是三种语言的词汇相似程度。具体如下：

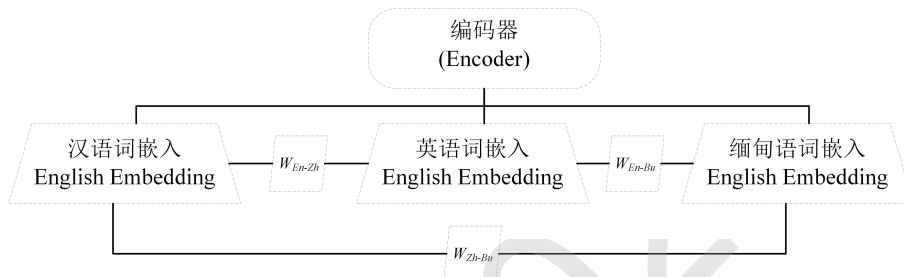


Figure 3: 汉英缅多语言词嵌入语义映射图

(1)如图3，分别训练汉英缅三种语言的单语词嵌入，我们利用Mikolov等人(Mikolov et al., 2013)提出的skip-gram算法得到汉英缅三种语言的单语词嵌入  $E_{Bu}$ ,  $E_{Zh}$ 。

(2)给定一句包含  $n$  个单词的源语言英语句子为  $x_{En} = (x_1, x_2, \dots, x_{|n|})$  我们将编码器端生成的英语词向量表示  $E_{En}$  进行提取。任意一个英语单词的经过lookup在  $R_{En}^{D \times V^{En}}$  操作后的词向量表示为  $E_{En}(x_{En})$ 。其中， $D$  表示词嵌入的维度，在本文中，我们设置为512维， $V^{En}$  表示为英语的词表。

(3)给定一句包含  $n$  个单词的源语言的中文句子、缅甸语句子为  $x_{Zh} = (x_1, x_2, \dots, x_{|n|})$ ,  $x_{Bu} = (x_1, x_2, \dots, x_{|n|})$ 。分别将汉语和缅甸语与英语的词向量学习跨语言线性映射矩阵  $W_{En-Zh} \in R_{En-Zh}^{D \times D}$ ,  $W_{En-Bu} \in R_{En-Bu}^{D \times D}$  最小化表示：

$$\sum_{(x_{Zh}, x_{En}) \in D_{Zh-En}} \|W_{En-Zh} E_{Zh}(x_{Zh}) - E_{En}(x_{En})\|^2 \quad (6)$$

$$\sum_{(x_{Bu}, x_{En}) \in D_{Bu-En}} \|W_{En-Bu} E_{Bu}(x_{Bu}) - E_{En}(x_{En})\|^2 \quad (7)$$

(4)将(3)中得到的  $W_{En-Bu} E_{Bu}(x_{Bu})$  和  $W_{En-Zh} E_{Zh}(x_{Zh})$  在进行跨语言线性最小化的表示：

$$\sum_{(x_{Zh}, x_{Bu}) \in D_{Zh-Bu}} \|W_{En-Bu} E_{Bu}(x_{Bu}) - W_{En-Zh} E_{Zh}(x_{Zh})\|^2 \quad (8)$$

由于缅甸语属于低资源语言，几乎没有开源的英缅以及汉缅的双语词典，其中，步骤3、步骤4中的词典  $D_{Bu-En}$ ,  $D_{Zh-En}$ ,  $D_{Zh-Bu}$  是通过将汉语、英语以及缅甸语三种语言的单语数据利用GIZA++<sup>-1</sup>词对齐的方式得到。通过词对齐的方式我们可以获得较高质量的英语-缅甸语，缅甸语-汉语以及汉语-英语的双语词典数据。

<sup>-1</sup><https://github.com/moses-smt/giza-pp>

(5)最后, 利用学习到的缅甸语词嵌入矩阵对源语言的词嵌入矩阵进行替换, 替换后的词嵌入矩阵具有汉语、英语以及缅甸语三种语言的语义信息, 更进一步的增强了三种语言对齐的语义空间。

通过步骤(1)-步骤(5), 我们得到了汉英缅三种语言的共同的语义表征方法。我们将汉英缅三种语言的词语进行了词嵌入转换, 通过在数学空间中进行匹配转换不考虑特定语言的字母, 缓解了嵌入级别的汉英缅三种语言词汇不匹配的问题。

## 4.2 汉英缅多语言参数共享

如图2, 在神经机器翻译的结构中, 当编码器在对汉语、英语以及缅甸语进行编码训练的过程中, 每个子层都会产生训练过程中的参数(Sachan and Neubig, 2018), 为了将汉英缅三种语言训练过程中的语义距离减小, 我们将所有子层生成的参数进行共享。除了3.1节中的共享词嵌入映射在相同的语义空间之外, 具体参与共享的参数还有: (1)自注意力机制中的线性变化产生的参数:  $W_K^1, W_Q^1, W_V^1, W_F^1$ (2)编码器-解码器的注意力机制:  $W_K^2, W_Q^2, W_V^2, W_F^2$ 。在汉英缅多语言神经机器翻译中, 为了将三种语言对之间的训练参数最大化共享。本文将以上的所有训练参数进行共享, 减小语义差距。具体共享过程如下:

Step1: 共享前馈神经网络中的子层的参数( $\Theta = \{W_E, H_{enc}, W_{L1}, W_{L2}\}$ ), 其中,  $W_E$ 表示经过语义映射后的词嵌入表示,  $H_{enc}$ 表示编码过程中的隐状态表示,  $W_{L1}, W_{L2}$ 表示输入共享语义空间的向量以及前馈神经网络中的参数。

Step2: 共享自注意力子层中的特征权重( $\Theta = \{W_E, H_{enc}, W_K^1, W_Q^1, W_V^1, W_F^1\}$ ), 其中,  $W_K^1, W_Q^1, W_V^1$ 表示注意力机制层生成的参数,  $W_F^1$ 表示前馈神经网络生成的训练参数。

Step3: 共享编码器-解码器注意力机制的子层( $\Theta = \{W_E, H_{enc}, W_K^2, W_Q^2, W_V^2, W_F^2\}$ ), 其中,  $W_K^2, W_Q^2, W_V^2$ 表示注意力机制层生成的参数,  $W_F^2$ 表示前馈神经网络生成的训练参数。

## 5 实验

### 5.1 实验数据

**汉英缅语义空间映射实验数据:** 在本文中, 我们利用GIZA++的方式对汉英、英缅以及汉缅的双语词语进行对齐, 我们使用构建的10万汉缅、英缅的双语句子以及开源的9.0M英汉双语句子进行词对齐的训练, 获得英汉、英缅以及汉缅的双语词典。另外, 在实验过程中, 我们将训练汉英缅多语言映射的实验数据进行训练集、测试集以及验证集的划分比例为: 8:1:1。

**汉英缅翻译实验数据:** 本文在一对多和多对多的两个翻译场景下进行汉英缅的多语言神经机器翻译。在本文中, 一对多的翻译场景是指源语言端是汉语, 目标语言端是英语和缅甸语, 源语言端是英语, 目标语言端是汉语和缅甸语以及源语言端是缅甸语, 目标语言端是英语和汉语。多对多的翻译场景是指源语言端是汉语、英语以及缅甸语, 目标语言端也是汉语、英语以及缅甸语。针对于汉语和英语的双语语料, 我们利用来自2018国际机器翻译大会(WMT-18)<sup>0</sup>的中英数据集。目前, 公开的英缅, 汉缅数据较少。本文利用构建了10万句英缅以及10万句汉缅的双语语料。具体的语料的信息如表1所示。

数据集	训练集	验证集	测试集
汉英	9.0M	2000	2000
英缅	98K	2000	2000
汉缅	98K	2000	2000

Table 1: 实验数据集表

### 5.2 实验设置

**多汉英缅多语言语义空间映射实验设置:** 在实验过程中, 我们将词向量的维度设置为512维, 训练过程中的迭代次数设置为10。

<sup>0</sup><http://www.statmt.org/wmt18/translation-task.html>

**翻译实验设置:** 在实验中, 使用比特对编码(Byte Pair Encoding, BPE)<sup>1</sup>(Sennrich et al., 2015)对汉语、英语以及缅甸语的单词进行了亚词切分。词表大小为35K。使用NIST的BLEU脚本<sup>2</sup>对翻译结果进行评测。实验环境为Ubuntu16.04, Linux系统, Tensorflow版本为tensorflow-gpu的1.13.2, 编译语言为Python3.7。我们选择Transformer\_Base作为我们实验的基础框架设置。在实验过程中, 我们设置Transformer的编码器和解码器均为6层, 在编码器和解码器中的词向量的维度以及注意力机制的单元为512。我们使用Adam优化器对训练过程中的学习率进行调整。另外, 为了防止过拟合的问题, 我们设置Dropout参数为0.1, 在训练过程中我们设置每个batch包含的源语言单词为2048, 每3000个batch对验证集做一次解码, 如果连续10次的验证集中的BLEU不再提高, 则提前终止训练, 防止模型过拟合。

### 5.3 实验结果及对比分析

#### 实验一: 一对一及一对多翻译场景下不同模型实验结果对比分析

在实验一中, 我们设置翻译场景为一对多, 一种源语言对应多种目标语言。具体如下表2所示, 分别选择汉语、英语以及缅甸语作为源语言对应另外两种不同的目标语言。我们设置对比实验如下:

(1)谷歌多语言神经机器翻译模型(MNMT): Johnson等人(2017)提出了一种基于双向LSTM的多语言神经机器翻译模型。

(2)Dong等人(2015)在一对多的翻译场景下, 提出将多语言翻译过程中的源语言的编码器共享, 为每个目标语言分配不同的解码器的方法。

(3)Transformer: 我们将比较在仅使用一对一的翻译场景下的源语言到目标语言的实验结果。

(4)Baseline: 基线模型是指在Transformer的框架下, 不使用共享语义空间以及共享参数的思路进行的翻译实验。

翻译场景	方法	源语言-目标语言					
		英汉	英缅	汉英	汉缅	缅英	缅汉
一对一	Transformer	26.02	14.55	25.08	14.09	16.77	16.00
	MNMT	25.56	13.09	23.45	13.78	15.40	16.14
一对多	Dong等人	24.77	14.87	23.60	13.55	15.30	16.40
	Baseline	26.04	15.06	24.89	14.35	16.80	16.30
	本文方法	<b>26.80</b>	<b>16.42</b>	<b>26.20</b>	<b>15.77</b>	<b>17.68</b>	<b>19.30</b>

Table 2: 一对一及一对多翻译场景下的实验结果表

如表2所示, 在一对多的翻译场景下, 本文提出的方法在汉语-缅甸语的翻译方向上BLEU值达到了15.77。相比较于谷歌提出的多语言神经机器翻译模型有明显的提升, 提升了1.99个BLEU值, 这说明基于Transformer的多语言神经机器翻译框架中的遮蔽注意力机制可以更好地对目标语言的进行翻译; Dong等人提出的方法的翻译效果在一对多的翻译场景下和谷歌提出的多语言神经机器翻译模型方法相当, 在翻译方向相同的情况下, 本文提出的方法对于汉语-缅甸语以及缅甸语-汉语的效果都要更好。另外, 在相同的翻译方向, 例如, 汉语-缅甸语, 本文提出的方法在一对多的情况下相比较于一对一的实验结果提升了1.68个BLEU值, 这说明利用大规模的汉语-英语的双语平行语料和较少的汉缅平行语料进行多语言联合神经机器翻译可以充分的弥补汉语-缅甸语之间数据缺乏导致模型效果不佳的问题。

#### 实验二: 一对一及多对多翻译场景下实验结果对比分析

在多到多的翻译场景下, 我们将源语言端、目标语言端设置为汉语、英语以及缅甸语三种语言, 我们将对比其他低资源神经机器翻译方法, 具体如下:

<sup>1</sup><https://github.com/bheinzerling/bpemb>

<sup>2</sup><https://www.nist.gov/itl/iad/mig/tools>

(1)谷歌多语言神经机器翻译模型: Johnson等人(2017)提出了一种基于双向的LSTM的多语言神经机器翻译模型。

(2)基于RNNSearch(RS)的多语言神经机器翻译方法: 在编码和解码端均使用RNN的网络结构(Bahdanau et al., 2014)。

(3)基于枢轴的神经机器翻译方法: kim等人(2019)提出一种基于枢轴的神经机器翻译方法, 本文在保持其他设置不变的情况下, 利用基于枢轴的方法对汉英缅进行神经机器翻译, 在实验中仅使用汉英、英缅数据获得汉缅的机器翻译模型。

翻译场景	方法	源语言-目标语言					
		英汉	汉英	英缅	缅英	汉缅	缅汉
一对一	Transformer	26.02	25.08	14.55	16.77	14.09	16.00
	MNMT	24.01	24.30	13.90	14.66	13.09	15.05
	RS	25.67	23.43	13.76	15.22	14.56	15.67
多对多	枢轴	-	-	-	-	14.05	17.09
	Baseline	25.44	24.80	14.22	14.50	15.44	15.86
	本文方法	<b>25.89</b>	<b>24.86</b>	<b>17.06</b>	<b>17.33</b>	<b>16.82</b>	<b>18.91</b>

Table 3: 一对一及多对多翻译场景下的实验结果表

如表3所示, 在汉语-缅甸语的翻译方向, 本文提出的方法相比较于基于枢轴的方法提高了2.77个BLEU值。在缅甸语-汉语的翻译方向, 本文提出的方法相比较于基于枢轴的方法提高了1.82个BLEU值, 这说明利用多语言的联合学习的方式可以有效的通过高资源语言对弥补低资源语言对的数据稀缺的问题。在相同的翻译方向, 本文方法都比RNNSearch的BLEU效果明显, 利用Transformer可以更好地将翻译信息融合, 利用参数共享的思想可以将语义之间的距离缩小。

同时, 对比表3和表4我们可以发现当翻译方向相同时, 多对多的翻译场景的效果好于一对多的翻译场景。例如, 汉-缅和缅-汉的多对多的翻译效果均优于一对多的翻译场景。以上的实验现象说明, 利用共享编码器的思想可以较好地三种语言之间进行映射, 减小语言之间的差异性。

在英语-汉语翻译方向, 本文的方法BLEU值为25.89, 一对一的Transformer模型的BLEU值为26.02, 本文方法下降了0.13个BLEU值, 因为, 在多对多的翻译场景下, 汉语-缅甸语的数据在汉语-英语的数据中增加了部分的噪声, 导致模型性能下降。

### 实验三: 多对多翻译场景下不同的词嵌入表征方式对汉英缅翻译效果的影响

在多对多的翻译场景下, 我们将讨论利用不同的词嵌入表征方式对翻译效果的影响, 具体实验结果如表4所示:

(1)单独表示: 对汉英缅每种语言都进行单独的初始化, 不共享词表及词嵌入的表示。

(2)Weight Tying: press等人(2017)提出一种将机器翻译中所有的输入和输出词向量同时共享的方法。

(3)本文方法: 将汉英缅三种语言的词嵌入共享, 编码及解码都进行相同语义空间的映射表示。

如表4所示, 本文方法的BLEU值高于其他两种词嵌入的表示方法, 说明在实验过程中, 将三种语言进行映射利用相同的表征会缩小语义空间的距离, 提升翻译模型的性能。同时, 也验证了本文方法的有效性。

实验结果显示了本文方法对于汉缅、英缅的显著提升, 使用不同的向量空间的表示方法对实验结果也有一定的影响, 汉语、英语以及缅甸语的语义空间相对较为独立, 使用单独的表示方法将三种语言的语义空间独立出来, 缅甸语相对英语和汉语之间的差异性极大, 这样的方法没办法将三种语言的语义空间统一缩小。利用Weight Tying的方法虽然可以较好的缩小三种语



翻译场景	词嵌入方法	源语言-目标语言					
		英汉	汉英	英缅	缅英	汉缅	缅汉
多对多	单独表示	24.62	23.50	15.40	16.61	14.22	16.56
	Weight Tying	24.67	23.55	15.76	16.72	14.56	16.42
	本文方法	<b>25.89</b>	<b>24.86</b>	<b>17.06</b>	<b>17.33</b>	<b>16.82</b>	<b>18.91</b>

Table 4: 不同的词嵌入表征方式对汉英缅翻译效果的影响

言的语义空间，减小语义差异，但是针对于缅甸语这种低资源语言，会出现汉缅、英缅之间较多的一对多或者多对一的情况，本文方法更好的缩小了这种语言之间的差异性，使三种语言更好的表示在同一个语义空间。

#### 实验四：不同的词汇表设置对汉英缅翻译效果的影响

在实验过程中，我们设置汉语、英语以及缅甸语的词汇表为独立的35K大小的词汇表，我们对比分析基线系统中独立使用词汇表对实验结果的影响。具体如下表5所示。

翻译场景	方法	源语言-目标语言					
		英汉	汉英	英缅	缅英	汉缅	缅汉
一对多	Baseline(独立词表)	25.55	23.20	13.72	14.04	13.17	15.57
	本文方法	<b>26.80</b>	<b>26.20</b>	<b>16.42</b>	<b>17.68</b>	<b>15.77</b>	<b>19.30</b>
多对多	Baseline(独立词表)	24.45	23.98	14.10	13.44	14.97	13.26
	本文方法	<b>25.89</b>	<b>24.86</b>	<b>17.06</b>	<b>17.33</b>	<b>16.82</b>	<b>18.91</b>

Table 5: 不同的词汇表设置对汉英缅翻译效果的影响

在独立词表的情况下，将本文方法与Baseline的方法进行对比，在一对多和一对多两个场景下，本文方法的BLEU都明显优于Baseline的方法。设置独立词表对比基线系统和本文提出的方法，在缅甸语、汉语以及英语的词表大小相同的情况下，本文提出的方法可以更好地将实验过程中的参数以及训练策略共享，通过高资源语言对的训练信息弥补低资源语言优于缺少数据导致翻译性能不佳的问题。

#### 5.4 翻译示例分析

如图4、图5所示，我们将翻译方向定位汉语-缅甸语，当对输入的源语言汉语句子中的“中国”进行翻译时，基线模型的翻译输出误判为两个缅甸语单词，本文方法输出的缅甸语句子正确的将两种语言的词语进行对齐，将汉缅句子中的单词在不同的语义空间中进行了校正，另外，对于基线模型输出的英语句子中“Today”时漏翻译以及错翻译，翻译示例证明了本文方法不但能将缅甸语、英语以及汉语三语语义空间不齐的问题进行了校正，也能够将互译的英缅、汉缅以及英汉单词进行校正也验证了本文方法的有效性。

针对于汉缅、英缅翻译效果的显著提升，在文中我们通过减小三者的语义空间缩小不同语言之间的距离，通过构建三语的词向量映射，构建词典将低资源语言缅甸语和英语、汉语进行语义映射，得到更好的双语表示。因此，通过实例证明本文方法得到了较好的翻译效果。

## 6 结论

针对汉英缅结构差异大导致多语言翻译时共享词表能力受限的问题，本文利用汉语、英语

1. 源语言句子 (汉语)	今天, 我们将从缅甸的仰光市回到中国。
	目标语言句子 (缅甸语)
基线系统	ယနေ့ကုန်ပုံတို့သည်မြန်မာနိုင်ငံမှရန်ကုန်မှပြည်တရုတ်ပြန်သွားပါမည်။
本文方法	ယနေ့ငါတို့ မာနိမြန်မာနိုင်ငံကရန်ကုန်ကနေ တရုတ်ပြည်ပြန်တော့မယ်။
参考译文	ဒီနေ့ငါတို့ မြန်မာနိုင်ငံကရန်ကုန်ကနေ တရုတ်ပြည်ပြန်တော့မယ်။
	目标语言句子 (英语)
基线系统	We <b>are going to</b> return to China from Yangon, Myanmar.
本文方法	We will return to China from Yangon, Myanmar, <b>today</b> .
参考译文	<b>Today</b> , we will <b>return to</b> China from Yangon, Myanmar.

Figure 4: 汉缅, 汉英翻译示例

2. 源语言句子 (缅甸语)	မြန်မာ - တရုတ်ဆက်ဆံရေးသည်ရှေးအချိန်ကတည်းကအလွန်ဖော်ရွေခဲ့သည်။
	目标语言句子 (汉语)
基线系统	缅甸语言一直和中国的关系很融洽。
本文方法	从很早的时候, 缅甸和中国的关系就很友好。
参考译文	缅甸与中国的关系自古以来就非常友好。
	目标语言句子 (英语)
基线系统	<b>Fast</b> - China has been <b>very popular since the beginning</b> .
本文方法	<b>Myanmar</b> -China relations have been <b>better since the beginning</b> .
参考译文	<b>Myanmar</b> -China relations have been very <b>friendly since ancient times</b> .

Figure 5: 汉缅, 汉英翻译示例

以及缅甸语三种语言进行联合语义的表征来提升缅汉机器翻译模型的性能。实验结果表明提出的方法在一对多的翻译场景下, 汉-缅的翻译方向上达到了15.77的BLEU值, 在多对多的翻译场景下, 汉-缅的翻译方向上达到了16.82的BLEU值, 相比较于基线模型均有明显的提升。在下一步的工作中, 我们将在多语言翻译框架下探索不同的参数共享方式对翻译效果的影响, 从而提升缅甸语的机器翻译性能。

### 参考文献

Roei Aharoni, Melvin Johnson, and Orhan Firat. 2019. Massively multilingual neural machine translation. pages 3874–3884.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2017. Learning bilingual word embeddings with (almost) no bilingual data. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 451–462.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv: Computation and Language*.

R Caruana. 1997. Multitask learning. *Machine Learning*, 28(1):41–75.

Raj Dabre, Atsushi Fujita, and Chenhui Chu. 2019. Exploiting multilingualism through multistage fine-tuning for low-resource neural machine translation. pages 1410–1416.

Daxiang Dong, Wu Hua, He Wei, Dianhai Yu, and Haifeng Wang. 2015. Multi-task learning for multiple language translation. In *Meeting of the Association for Computational Linguistics International Joint Conference on Natural Language Processing*.

Orhan Firat, Kyunghyun Cho, and Yoshua Bengio. 2016a. Multi-way, multilingual neural machine translation with a shared attention mechanism.

- Orhan Firat, Baskaran Sankaran, Yaser Alonaizan, Fatos T Yarman Vural, and Kyunghyun Cho. 2016b. Zero-resource translation with multi-lingual neural machine translation. *arXiv: Computation and Language*.
- Thanh Le Ha, Jan Niehues, and Alexander Waibel. 2016. Toward multilingual neural machine translation with universal encoder and decoder.
- Melvin Johnson, Mike Schuster, Quoc V Le, Maxim Krikun, Yonghui Wu, Zhibo Chen, Nikhil Thorat, Fernanda B Viegas, Martin Wattenberg, Greg S Corrado, et al. 2017. Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5(1):339–351.
- Yunsu Kim, Petre Petrov, Pavel Petrushkov, Shahram Khadivi, and Hermann Ney. 2019. Pivot-based transfer learning for neural machine translation between non-english languages. pages 866–876.
- Surafel Melaku Lakew, Aliia Erofeeva, Matteo Negri, Marcello Federico, and Marco Turchi. 2018. Transfer learning in multilingual neural machine translation with dynamic vocabulary. *arXiv: Computation and Language*.
- Jason Lee, Kyunghyun Cho, and Thomas Hofmann. 2017. Fully character-level neural machine translation without explicit segmentation. *Transactions of the Association for Computational Linguistics*, 5:365–378.
- Constantine Lignos, Daniel Cohen, Yenchieh Lien, Pratik Mehta, W Bruce Croft, and Scott Miller. 2019. The challenges of optimizing machine translation for low resource cross-language information retrieval. pages 3495–3500.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *Computer Science*.
- Khin Thandar Nwet, Khin Mar Soe, and Lar Thein Ni. 2011a. Building bilingual corpus based on hybrid approach for myanmar-english machine translation. *International Journal of Scientific Engineering Research*, Volume 2(Issue 8).
- Khin Thandar Nwet, Khin Mar Soe, and Lar Thein Ni. 2011b. Developing word-aligned myanmar-english parallel corpus based on the ibm models. *International Journal of Computer Applications*, 27(8).
- Ofir Press and Lior Wolf. 2017. Using the output embedding to improve language models. 2:157–163.
- Devendra Singh Sachan and Graham Neubig. 2018. Parameter sharing methods for multilingual self-attentional translation models. *arXiv: Computation and Language*.
- Rico Sennrich, Barry Haddow, and Alexandra Birch. 2015. Neural machine translation of rare words with subword units. *Computer Science*.
- Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. *arXiv: Computation and Language*.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. pages 5998–6008.
- Yining Wang, Jiajun Zhang, Feifei Zhai, and Jingfang Xu. 2018. Three strategies to improve one-to-many multilingual translation. In *Conference on Empirical Methods in Natural Language Processing*.
- Yining Wang, Long Zhou, Jiajun Zhang, Feifei Zhai, and Chengqing Zong. 2019. A compact and language-sensitive multilingual translation method. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*.
- Aye Thida Win. 2011. Words to phrase reordering machine translation system in myanmar-english using english grammar rules. 3:50–53.
- Barret Zoph and Kevin Knight. 2016. Multi-source neural translation.