# 融入音素特征的英-泰-老多语言神经机器翻译方法

沈政[1,2]，毛存礼[*1,2]，余正涛[1,2]，高盛祥[1,2]，王琳钦[1,2]，黄于欣[1,2]
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
1591744723@qq.com, maocunli@163.com, ztyu@hotmail.com
gaoshengxiang.yn@foxmail.com, 2424172505@qq.com, huangyuxin2004@163.com

## 摘要

多语言神经机器翻译是提升低资源语言翻译质量的有效手段。由于不同语言之间字符差异较大，现有方法难以得到统一的词表征形式。泰语和老挝语属于具有音素相似性的低资源语言，考虑到利用语言相似性能够拉近语义距离，提出一种融入音素特征的多语言词表征学习方法：（1）设计音素特征表示模块和泰老文本表示模块，基于交叉注意力机制得到融合音素特征后的泰老文本表示，拉近泰老之间的语义距离；（2）在微调阶段，基于参数分化得到不同语言对特定的训练参数，缓解联合训练造成模型过度泛化的问题。实验结果表明在ALT数据集上，提出方法在泰-英和老-英两个翻译方向上，相比基线模型提升0.97和0.99个BLEU值。

**关键词：** 多语言神经机器翻译；泰语；老挝语；低资源语言；音素；参数分化

# English-Thai-Lao multilingual neural machine translation fused with phonemic features

**Zheng Shen[1,2], Cunli Mao[*1,2], Zhengtao Yu[1,2], Shengxiang Gao[1,2], Linqin Wang[1,2], Yuxin Huang[1,2]**
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
1591744723@qq.com, maocunli@163.com, ztyu@hotmail.com
gaoshengxiang.yn@foxmail.com, 2424172505@qq.com, huangyuxin2004@163.com

## Abstract

Multilingual neural machine translation is an effective methods to improve the performance of low-resource language translation. However, characters in different languages are significantly different, and existing methods are difficult to obtain a unified word representation form. Thai and Lao are low-resource languages with phonemic similarity. Considering that the use of language similarity can shorten the semantic distance, in this article, a multilingual word representation learning method incorporating phonemic features is proposed: (1) Design the phoneme feature representation module and the Thai-Lao text representation module, and then obtain the Thai-Lao text representation after fused phoneme features based on the cross-attention mechanism to shorten the semantic distance between the Thai-Lao and Laotian; (2) In the fine-tuning stage, specific training parameters for different language pairs are obtained based on parameter differentiation, which alleviates the problem of over-generalization of the model

*毛存礼(通信作者):maocunli@163.com
国家自然科学基金重点项目（61732005,U21B2027）；国家自然科学基金（62166023，61866019）；云南省自然科学基金重点项目( 2019FA023);云南省重大科技专项计划项目(202103AA080015，202002AD080001)

第二十一届中国计算语言学大会论文集，第305页-第316页，南昌，中国，2022年10月14日至16日。
(c) 2022 中国中文信息学会计算语言学专业委员会

caused by joint training. Experimental results on ALT datasets show that the proposed method improves the BLEU values by 0.97 and 0.99 in Thai-English and Lao-English translation tasks, respectively, compared with the benchmark model.

**Keywords:** multilingual neural machine translation , Thai , Lao , low-resource languages , phoneme , parameter differentiation

## 1 引言

近年来，神经机器翻译(NMT)(Sutskever et al., 2014; Bahdanau et al., 2014; Wang et al., 2022)因其优越的性能引起了广泛的关注，成为机器翻译领域的主流方法。随之而来，基于统一的翻译模型实现多种语言对联合训练的框架成为了研究热点(Xu et al., 2021)，目前，MNMT(Wang et al., 2019; Bapna et al., 2022)在低资源语言(Man et al., 2020)翻译上取得了较好的效果，相比单独训练双语翻译模型，MNMT能够通过共享跨语言知识来提升资源稀缺语言的机器翻译性能。然而，在如何利用语言之间特有的知识上仍有较大的研究空间。

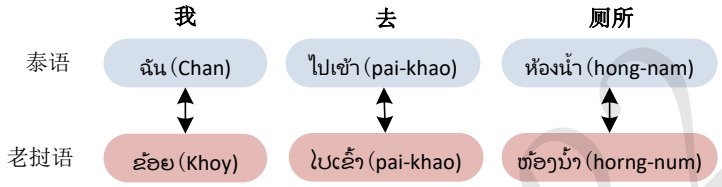


Figure 1: 泰语-老挝语音素相似性示例

现有方法在进行多语言词表征时，由于不同语言之间字符差异性较大难以得到统一的词表征形式，例如，泰语和老挝语属于孤立型语言，不具备天然分词，在机器翻译模型训练的过程中，泰语、老挝语和英语之间的语言差异性极大，仅仅通过联合训练或参数共享的方式无法得到准确的语义表征。泰语和老挝语都属于汉藏语系壮侗语族的壮傣语支，在构词特点、词语音素以及句法结构上都有相同或相似的地方，特别是在音素层面上，大部分具有相同含义的泰语、老挝语音素相同(Ding et al., 2016; Yu et al., 2020)。如图1所示，泰语和老挝语句法结构基本一致，都属于主语-谓语-宾语(Subject-Verb-Object, SVO)的结构，音素也有较高的相似度，如汉语“去”对应的老挝语音素 ***pai-khao*** 和泰语音素 ***pai-khao*** 相同，并且，汉语“我”和“厕所”对应的泰语、老挝语音素也具备一定的相似性，这说明泰语、老挝语两种语言在音素层面上存在大量的一致性。Tan et al. (2019)已经证明，相似性高的语言进行多语言联合训练时，该特性有助于提高翻译模型性能，这是因为模型在训练过程中能自动学习到语言在句法、词法等层面上的相似特征。

因此，鉴于泰语、老挝语之间的语言相似性，利用多语言机器翻译模型能够有效地学习到泰语和老挝语的相似特征，提升模型翻译性能。然而我们观察到泰语和老挝语的相似性并不体现在字符层面上，单纯利用联合训练的方式无法学习到音素层面的相似性特征。针对以上问题，本文提出了融入音素特征的多语言词表征学习方法，在Transformer的框架下，将音素和文本分别进行词向量表示，并基于交叉注意力机制将二者融合，最后，基于参数分化策略对模型进行微调。

本文的贡献主要有以下三点：

（1）为了进一步拉近泰语、老挝语之间的语义表征距离，提出联合泰语、老挝语音素特征和文本表示方法，基于交叉注意力机制进一步学习融合音素特征后的文本表示。

（2）在我们的工作中，由于泰-老之间具有语言相似性，为了更有效的对泰-老语言特定参数进行分化，我们在参数分化方法的基础上对模型进行微调，以此缓解联合训练造成的模型过度泛化问题。

（3）在公开数据集ALT上，实验结果表明所提方法优于多个基线模型，在泰-英翻译方向上BLEU值达到17.99，在老-英翻译方向上的BLEU值达到15.40，表明翻译性能提升得益于提出的融入音素特征的多语言词表征学习方法。

第二十一届中国计算语言学大会论文集，第305页-第316页，南昌，中国，2022年10月14日至16日。

## 2 相关工作

泰语和老挝语是典型的低资源语言，其双语平行语料稀缺，所以相关机器翻译研究较少。早期主要利用基于统计规则的机器翻译方法实现对泰语、老挝语的翻译(Phitakwinai et al., 2008; Asawavichienjinda et al., 2005)，但是统计机器翻译需要大量双语语料，用于泰语和老挝语效果不佳。随着机器翻译技术的发展，神经机器翻译方法逐渐被应用到泰语和老挝语的翻译任务上(Saengthongpattana et al., 2019; Poncelas et al., 2020)，但是当前研究多集中在现有方法的简单应用，没有有效利用泰语和老挝语的语言特点，因此翻译效果不佳。

MNMT可以通过不同语种之间的知识迁移提升资源稀缺语言的翻译效果，目前已经成为解决低资源机器翻译的主流方法之一。近年来，研究人员对MNMT模型结构进行了很多探索。主要有三类：(1)对所有源语言使用相同的编码器，对目标语言使用不同的解码器。Dong et al. (2015)在一对多的翻译场景下，提出所有源语言共享编码器，为每个目标语言分配不同解码器的方法。(2)对所有源语言和目标语言都使用不同的编码器和解码器。Zoph and Knight (2016)提出多种语言联合训练注意力机制的多对一多语言机器翻译方法。上述方法由于都要为每种语言单独训练编码器或解码器，极大地限制了模型在语言数量上的可扩展性。(3)对所有的源语言和目标语言均使用相同的编码器的解码器。Johnson et al. (2017)提出了一种多语言联合训练的单编码器-单解码器模型，并在源语言首字符前增加目标语言标志位以指导目标语言的生成。该方法虽然实现了利用单一模型翻译多个语种，但是忽略了语种之间的差异性。最新的多语言机器翻译方法多倾向于在单一模型上设计语言特定的子模块。Wang and Zhang (2021)训练过程中利用梯度差异逐步分离语言特定参数。Xie et al. (2021)根据各个神经元在该语言对上的重要性划分子网络。Khusainova et al. (2021)通过显示不同语言之间相关程度的语言树控制共享参数的数量。Zhu et al. (2021)引入一个轻量级的适配器，通过该适配器学习各个语种的特有信息。Zhang et al. (2020a)利用门控机制训练多语言机器翻译模型，实现模型动态选择是否共享参数。Zhang et al. (2020b)通过语言感知的归一化层将不同语言映射到不同的高斯空间中，并利用语言感知的线性层对不同语言之间的关系进行建模。

上述方法为本文提供了较好的思路，本文在Johnson et al. (2017)提出方法的基础上进行改进，针对现有方法难以得到统一的泰老词表征形式问题，提出融入音素特征的英-泰-老多语言神经机器翻译方法，利用泰语和老挝语的音素相似性，拉近其语义距离。Wang and Zhang (2021)首次提出利用参数分化的思想训练多语言神经机器翻译模型，取得了不错的效果，而本文主要研究泰语-老挝语这种具有语言相似性的语种，这样的语种更适合在模型充分学到到语言相似特征后再进行参数分化，因此，我们在微调阶段再利用参数分化学习语言特定参数。

## 3 研究背景

在本节中，我们主要介绍基于注意力机制的Transformer框架(Vaswani et al., 2017)和泰语、老挝语的语言相似性。

### 3.1 基于Transformer的神经机器翻译模型

Transformer是基于序列到序列框架(Sutskever et al., 2014)实现的，由多层编码器和多层解码器堆叠而成，每层编码器包含一个多头自注意力层和一个前馈神经网络层，每层解码器除了上述两个模块，在多头自注意力层后是一个多头交叉注意力层，每个模块由残差连接和归一化进行关联。对于给定源语言句子$x = (x_1, x_2, ..., x_n)$，首先通过编码器将其编码为一个稠密的隐向量表示，然后利用解码器将其解码成目标语言句子$y = (y_1, y_2, \ldots, y_z)$。

多头注意力机制是Transformer中的一个重要模块,可以表示为：

$$MultiHead(Q, K, V) = Concat(head_1, head_2, ..., head_M)W \tag{1}$$

其中，$Q(query)$，$K(key)$，$V(value)$是输入句子的隐向量表示，$W$为参数矩阵，$M$为多头注意力机制头数，每个头计算如下：

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \tag{2}$$

$$= softmax\left(\frac{(QW_i^Q)(KW_i^K)^T}{\sqrt{d_k}}\right)(VW_i^V) \tag{3}$$

其中，$W_i^Q$，$W_i^K$，$W_i^V$是参数矩阵，$d_k$是隐藏层的维度。

多头自注意力机制能建立句子中词与词之间的相互连接，得到融合上下文信息的隐向量表示，多头交叉注意力机制主要用于连接源语言与目标语言向量表示。由于Transformer网络不使用递归方式编码，因此在模型中使用位置嵌入来利用序列位置信息。

### 3.2 多语言神经机器翻译模型

与NMT模型相比，MNMT对多个语言对进行联合训练，实现了多任务参数共享。我们选择Johnson et al. (2017)提出的方法作为实验的基准模型，该方法训练了一个单编码器-单解码器的MNMT模型，并通过在源语言句子首位增加目标语言标志位指导目标语言的生成，其目标函数为所有语言对下每个词生成概率的乘积:

$$\mathcal{L}(D;\theta) = \sum_{l=1}^{L}\sum_{d=1}^{|D_l|}\sum_{t=1}^{N} \log P(y_t^l|x^l; y_{<t}^l; \theta_{enc}; \theta_{dec}; \theta_{attn}) \tag{4}$$

其中，$D$是训练语料中所有平行句对的集合，$\theta$是模型中所有参数的集合，$L$表示模型联合训练的语言对总数，$N$表示目标语言句子长度，$|D_l|$表示训练语料中属于第$l$个语言对的平行句对数量，$P(y_t^l|x^l; y_{<t}^l)$表示第$l$语言对中第$d$个句子的第$t$个单词的翻译概率，$\theta_{enc}$表示模型中编码器的参数，$\theta_{dec}$表示模型中解码器的参数，$\theta_{attn}$表示模型中注意力机制的参数。

### 3.3 泰语、老挝语语言相似性

本小节我们主要回答以下两个问题:

（1）泰语、老挝语语言相似性如何体现?

（2）语言相似性如何影响机器翻译性能?

#### 3.3.1 泰语、老挝语语言相似性如何体现

泰语和老挝语都属于汉藏语系壮侗语族的壮傣语支，两国文字的发音方面较为相似(Ding et al., 2016)，泰老两种语言的音素都是由元音、辅音、尾辅音、声调符号等构成的，并且大多数情况下一一对应。为了更准确地利用泰老的音素相似特征，我们利用编辑距离对语料中泰老平行句对的音素相似性进行统计分析。

| 音素相似度 | <0.6 | 0.6-0.7 | 0.7-0.8 | >0.8 |
|---|---|---|---|---|
| 语料占比(%) | 1.58 | 61.93 | 26.08 | 10.41 |

Table 1: 泰-英、老-英数据集统计信息

由表1可得，语料中音素相似度在0.6以上的平行句对占比达到98.42%，这说明泰老在音素层面有较高的相似性。

#### 3.3.2 语言相似性如何影响机器翻译性能

受人文地理等的影响，现有的很多语言都是由同一古语言发展而来，从而在发音、书写、句法等方面有较高的相似性。NMT属于跨语言任务，其关键在于如何根据源语言的语义表征解码出对应目标语言，当两种语言属于相似语言时，其语义空间更加接近，翻译效果更佳，例如，两种语言可通过共享词表共享其同源词向量表征，WMT相似语言翻译任务对此已进行了大量的研究(Ojha et al., 2019; Baquero-Arnal et al., 2019)，证明了利用语言相似性可以有效提升翻译模型性能。由3.3.1的分析可以得出，泰语和老挝语具有较高的音素相似性，但传统NMT模型无法充分利用该特性。因此本文考虑基于交叉注意力机制通过泰老音素相似性拉近其语义距离，提升翻译模型性能。

## 4 融合音素特征的多语言神经机器翻译模型

本文的目标是构建一个源端为泰语和老挝语，目标端为英语的MNMT模型。总体框架如图2所示，主要包括音素特征表示模块、泰老文本表示模块、基于交叉注意力机制的音素-文本表示模块、目标语言解码器。

### 4.1 泰老文本表示模块

给定一个泰语或老挝语句子为$x = (x_1, x_2, ..., x_n)$，其中$n$为文本$x$的序列长度，文本序列通过带有位置嵌入的传统嵌入层得到其词向量表征$E_t$，计算如下：

$$E_t = Emb_t(x) + PE_t(x) \tag{5}$$

其中，$Emb_t$为文本序列词嵌入层，$PE_t$为文本位置嵌入层，$E_t \in R^{n \times d_k}$。

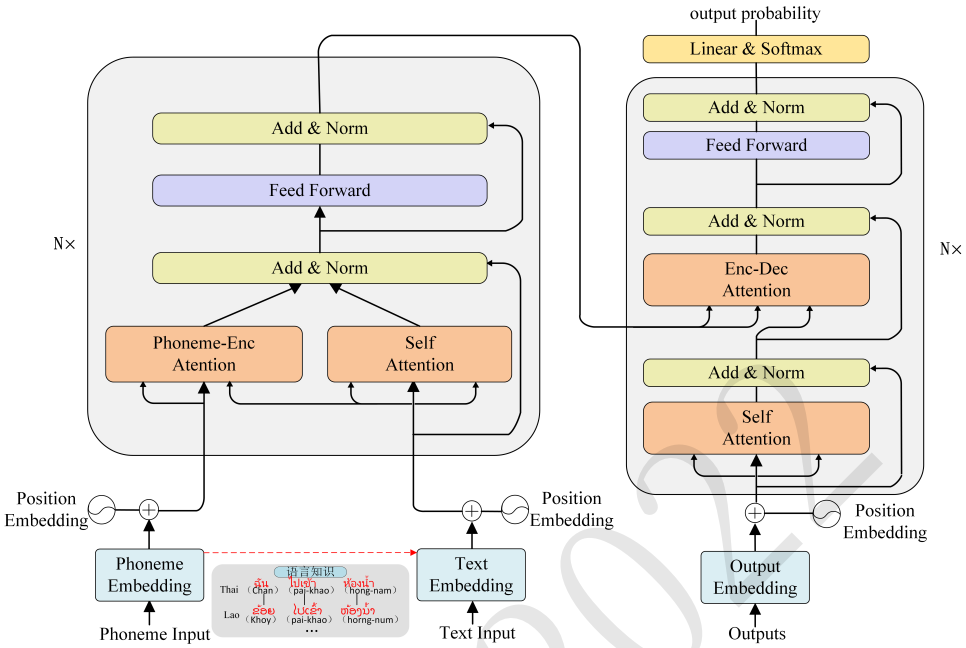


Figure 2: 融合音素特征的多语言神经机器翻译模型框架

### 4.2 音素特征表示模块

对于文本序列$x$，通过G2P（字符转音素）[0]工具将其转化成对应的音素序列$x_p = (x_{p1}, x_{p2}, ..., x_{pm})$，其中$m$为音素$x_p$的序列长度，音素序列通过带有位置嵌入的传统嵌入层得到其词向量表征$E_p$，计算如下：

$$E_p = Emb_p(x_p) + PE_p(x_p) \tag{6}$$

其中，$Emb_p$为音素序列词嵌入层，$PE_p$为音素位置嵌入层，$E_p \in R^{m \times d_k}$。

### 4.3 基于交叉注意力机制的音素-文本表示模块

为了拉近老挝语和泰语的语义距离，本文通过交叉注意力机制将音素特征融入泰老文本表示，如图3所示。首先，文本词向量表征$E_t$经过自注意力层计算源语言序列上下文向量$H_t$：

$$H_t = MultiHead(E_t, E_t, E_t) \tag{7}$$

同时，文本词向量表征$E_t$为查询向量，音素词向量表征$E_p$为键向量和值向量，经过音素-文本交叉注意力机制得到融入音素特征的文本表示$H_p$：

$$H_p = MultiHead(E_t, E_p, E_p) \tag{8}$$

然后，采用加权的方式将$H_t$和$H_p$进行融合：

$$H = \alpha * H_t + (1 - \alpha) * H_p \tag{9}$$

[0]https://github.com/dmort27/epitran

其中$\alpha$是超参数。

最后，使用位置前馈网络（FFN）更新序列每个位置的状态，得到$H_{enc}$：

$$H_{enc} = FFN(H) \tag{10}$$

经过多层编码后，将编码器的输出$H_{enc}$输入解码器进行解码。

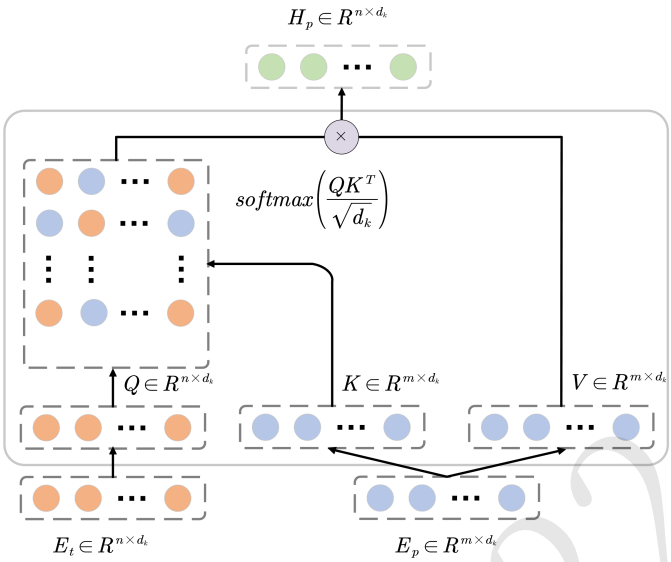


Figure 3: 基于交叉注意力机制的音素-文本表示模块

### 4.4 目标语言解码器

与泰老文本表示模块类似，首先将泰语或老挝语句子$x$对应的英语句子$y = (y_1, y_2, \ldots, y_z)$进行词向量表征得到$E_y$，其中$z$为目标语言序列长度。如图2右侧所示，本文解码器采用传统的Transformer框架，每层解码器由多头自注意力层、多头交叉注意力层、前馈神经网络层三个子层组成。

首先利用多头自注意力机制提取目标句子特征：

$$H_y = MultiHead(E_y, E_y, E_y) \tag{11}$$

然后使用多头交叉注意力机制实现融合音素特征的源语言上下文向量$H_{enc}$和目标句子特征$H_y$的交互：

$$H_c = MultiHead(E_y, E_{enc}, E_{enc}) \tag{12}$$

然后，使用FFN更新序列每个位置的状态，得到$H_{dec}$：

$$H_{dec} = FFN(H_c) \tag{13}$$

最后将解码器最后一层的输出作为softmax层的输入，并预测目标句子的概率分布：

$$P = Softmax(W_p H_{dec} + b) \tag{14}$$

其中$W_p$和b是模型参数。

### 4.5 微调

在微调阶段，考虑到不同语言之间的参数干扰问题，我们基于参数分化思想(Wang and Zhang, 2021)，与之不同的是我们没有在训练阶段使用该思想，主要是因为训练阶段分离参数会导致模型无法充分学习到语言相似特征。因此，我们基于该思想对模型进行微调，即针对训练好的模型，分别利用泰语-英语和老挝语-英语的验证集获取两个语言对在各个参数上的梯度，并依此计算各个参数上两个语言对梯度的余弦相似度：

$$sim(\theta_i) = \frac{g_i^{t_1} \cdot g_i^{t_2}}{||g_i^{t_1}|| \cdot ||g_i^{t_2}||} \tag{15}$$

第二十一届中国计算语言学大会论文集，第305页-第316页，南昌，中国，2022年10月14日至16日。

其中，$\theta_i$是模型第i个参数，$t_1$指老挝语到英语的翻译任务，$t_2$指泰语到英语的翻译任务，$g_i^{t_1}$是任务$t_1$在$_i$上的梯度。

模型每微调一定步数计算一次梯度，并对$t_1$和$t_2$梯度相似度较低的参数进行分离，即$t_1$和$t_2$的该参数不再共享，两个任务分别针对该参数微调，直到模型再次收敛。

## 5 实验

为了在实验中公平的比较模型性能，并且同时验证所提泰-老音素相似性特征的有效性，在本文中我们的主要研究对象是泰语和老挝语这两种音素相似的语言，具体实验如下。

### 5.1 实验数据

本文的泰-英、老-英的语料直接来源于公共数据集亚洲语言树库（ALT）[1]，泰-英和老-英分别有20106条平行语料。由于该数据集没有划分训练集、验证集和测试集，本文选取泰-英和老-英数据各1000条作为验证集，取1106条作为测试集，剩余18000条作为训练集，如表2所示：

| 数据集 | 训练集（句对） | 验证集（句对） | 测试集（句对） |
|---|---|---|---|
| 泰-英 | 18000 | 1000 | 1106 |
| 老-英 | 18000 | 1000 | 1106 |

Table 2: 泰-英、老-英数据集统计信息

### 5.2 实验环境及配置

本文实验的神经网络模型是基于Torch1.8实现的，编译语言为Python 3.8，在单个NVIDIA Tesla T4 GPU上进行实验。在实验中，我们使用BPE对所有源语言和目标语言进行联合子词切分，词表大小为4k。本文选择Transformer模型作为基础模型，模型的编解码器分别设置为3层。在编码器和解码器中的词向量和隐藏层的维度设置为128维。优化器选择参数设置为$\beta_1$ = 0.9，$\beta_2$ = 0.98的Adam优化器优化模型。我们参照Vaswani et al. (2017)使用warm_steps = 4000的warm-up策略来调整学习率，每个批次包含大约4096个词。我们训练模型直到连续10次验证集的BLEU值没有提升，则认为模型收敛并停止训练，该方法可以有效防止模型过拟合。在解码过程中，beam search设置为5,并采用BLEU(Papineni et al., 2002)指标来评估模型性能。

### 5.3 实验结果及分析

本文初步探索了泰老音素级别的相似性，并在编码端利用该特征参与模型训练，解码端统一设置为英语，这是因为在自回归的框架下逐词解码并转音素再参与模型解码会导致翻译错误的累积。因此，我们主要考虑在泰-英和老-英方向上，探究音素相似性对实验结果的影响。

#### 5.3.1 一对一及多对一的翻译场景下不同模型实验结果对比分析

在实验中，我们基于fairseq[2]框架与其他模型对比，并按照原论文参数复现，在达到最好结果时进行比较分析。设置对比试验如下：

(1)Bilingual：Vaswani et al. (2017)为每个语言对分别训练一个Transformer神经机器翻译模型，其参数设置与本文提出的方法一致。

(2)Multi-Source：Zoph and Knight (2016)在多对一的翻译场景下为每个源语言分配不同的编码器，目标语言共享解码器。

(3)Adapter：Bapna et al. (2019)在Transformer每一层的顶端为每一个语言对增加一个额外的适配器，每个适配器分别学习各个语言地特定知识。

(4)PD：Wang and Zhang (2021)提出在训练阶段利用不同语言对参数梯度的相关性分离语言特定参数。

(5)LaSS：Lin et al. (2021)通过判断神经元重要性为各语言对裁剪冗余神经元，以此使对所有语言对都重要的神经元学习通用知识，对单个语言对重要的神经元学习语言特定的知识。

[1]http://www2.nict.go.jp/astrec-att/member/mutiyama/ALT/index.html
[2]https://github.com/facebookresearch/fairseq

(6)基线模型（Baseline）：基线模型是指基于Transformer框架，不使用音素特征和参数分化策略下进行的翻译实验。

| 翻译场景 | 方法 | 老-英 | △ | 泰-英 | △ |
|---|---|---|---|---|---|
| 一对一 | Transformer | 9.72 | - | 14.70 | - |
| 多对一 | Multi-Source | 12.75 | +3.03 | 16.13 | +1.43 |
| | Adapter | 14.53 | +4.81 | 16.79 | +2.09 |
| | PD | 14.04 | +4.32 | 16.36 | +1.66 |
| | LaSS | 12.54 | +2.82 | 15.24 | +0.54 |
| | Baseline | 14.43 | +4.71 | 17.00 | +2.30 |
| | 本文方法（Transformer-Base） | 7.12 | -2.60 | 8.44 | -6.26 |
| | **本文方法** | **15.40** | **+5.68** | **17.99** | **+3.29** |

Table 3: 一对一及多对一翻译场景下的实验结果

如表3所示，在一对一的翻译场景下，基于Transformer框架在老-英和泰-英翻译方向上BLEU值分别达到了9.72和14.70。在多对一的翻译场景下，所有模型相比一对一场景下的BLEU值均有明显提升，其中，本文提出的方法在老-英和泰-英翻译方向上BLEU值分别达到了15.40和17.99，取得了最高水平，在老-英和泰-英翻译方向上BLEU值分别提升了5.68和3.29，这说明利用MNMT方法将老挝语-英语和泰语-英语联合训练，可以通过知识迁移有效缓解老挝语和泰语数据稀缺导致的模型翻译性能不佳的问题。

此外，本文方法相比Multi-Source在老-英和泰-英翻译方向上BLEU值分别提升了2.65和1.86，这说明共享编码器可以有效利用泰老语言相似性提升模型翻译效果。相比Adapter，本文方法在老-英和泰-英翻译方向上BLEU值分别提升了0.87和1.20，这说明低资源情况下单独训练额外参数效果不佳。相比PD，本文方法在老-英和泰-英翻译方向上BLEU值分别提升了1.36和1.63，这说明该方法会过早分离模型参数从而导致模型知识迁移不充分。相比Lass，本文方法在老-英和泰-英翻译方向上BLEU值分别提升了2.86和2.75，这说明该方法依赖大规模的模型参数和训练数据，在低资源情况下会出现过度裁剪而丢失部分共有参数的问题。相比Baseline，本文方法在老-英和泰-英翻译方向上BLEU值分别提升了0.97和0.99，说明本文方法可以有效拉近泰老之间的语义距离并缓解联合训练造成的模型过度泛化的问题，提升翻译模型性能。为了进一步证明本文参数设置的合理性，本文设置了在Transformer-Base参数下的对比实验，该参数下的的BLEU值远小于本文参数下的BLEU值，这说明参数过大会导致模型过拟合，从而使测试集上的翻译效果较差。

### 5.3.2 消融实验

为了探究融入音素特征和基于参数分化的微调策略的有效性，本文设置了消融实验，如表4所示。

| 方法 | 老-英 | △ | 泰-英 | △ |
|---|---|---|---|---|
| Baseline | 14.43 | - | 17.00 | - |
| Baseline+音素 | 15.13 | +0.70 | 17.74 | +0.74 |
| Baseline+音素（拼接） | 6.50 | -7.93 | 9.77 | -7.23 |
| Baseline+参数分化 | 14.64 | +0.21 | 17.23 | +0.23 |
| **Baseline+音素+参数分化** | **15.40** | **+0.97** | **17.99** | **+0.99** |

Table 4: 消融实验

实验结果表明，融入音素特征使模型在老-英和泰-英翻译方向上BLEU值分别提升了0.70和0.74，说明该方法可以有效拉近泰老之间的语义距离，缓解泰老字符差异较大导致的词表征形式不统一的问题。基于参数分化思想的微调策略使模型在老-英和泰-英翻译方向上BLEU值分别提升了0.21和0.23，说明该方法可以学习到语言特定知识，缓解联合训练造成模

型过度泛化的问题。基线模型+音素的方式相比基线模型+参数分化的方式，在老-英和泰-英翻译方向上的BLEU值提升更为明显，说明本文提出的方法对翻译性能带来的提升更依赖于泰语和老挝语之间的音素相似性。两种方法可同时使用，此时模型效果达到最佳，在老-英和泰-英翻译方向上BLEU值分别提升了0.97和0.99。为了进一步证明本文方法的有效性，本文利用Koehn (2004)提出的重采样方法进行了显著性检验($p < 0.05$)。

为了探讨音素融合方式对实验结果的影响，我们设计了利用拼接方式融合音素特征的实验，即模型输入为拼接了音素的文本，实验结果表明，拼接方式会使模型BLEU值远低于基线模型，这说明直接拼接会使输入序列过长，使得模型学习困难较大，从而导致模型性能下降。

### 5.3.3 音素特征融合层数对翻译效果的影响

为了探究编码器不同层数融合音素特征对实验结果的影响，我们设计对比试验，对比三层，结果如表5所示。

| 融合层数 | 老-英 | 泰-英 |
|---|---|---|
| 第一层 | 14.05 | 16.89 |
| 第二层 | 14.74 | 16.95 |
| 第三层 | 14.64 | 16.78 |
| 第一层+第二层 | 15.00 | 17.38 |
| 第一层+第三层 | 14.90 | 16.95 |
| 第二层+第三层 | 14.52 | 16.93 |
| **第一层+第二层+第三层** | **15.13** | **17.74** |

Table 5: 音素特征融合层数对翻译效果的影响

表5实验结果表明，音素特征融合层数对模型翻译效果有着重要的影响，当只进行单层融合时，第二层效果最佳，当进行两层融合时，第一层+第二层效果最佳，当三层都融入音素特征时模型效果达到最优。当只进行单层融合时，模型无法从音素特征中学习到有效的信息，甚至会引入噪声信息，从而影响模型性能。当模型进行两层融合时，音素特征低层融合比高层融合效果更佳，这表明，模型低层更容易学习到泰语和老挝语的音素相似特征。当模型进行三层融合时，模型能充分利用泰老音素相似性拉近语义距离，提升翻译效果。

### 5.3.4 音素特征和文本特征融合比例对翻译效果的影响

本文利用超参数$\alpha$对模型学到的文本特征和音素特征比例进行平衡，并针对该超参数的设置进行了探究，讨论$\alpha$取值对实验结果的影响。

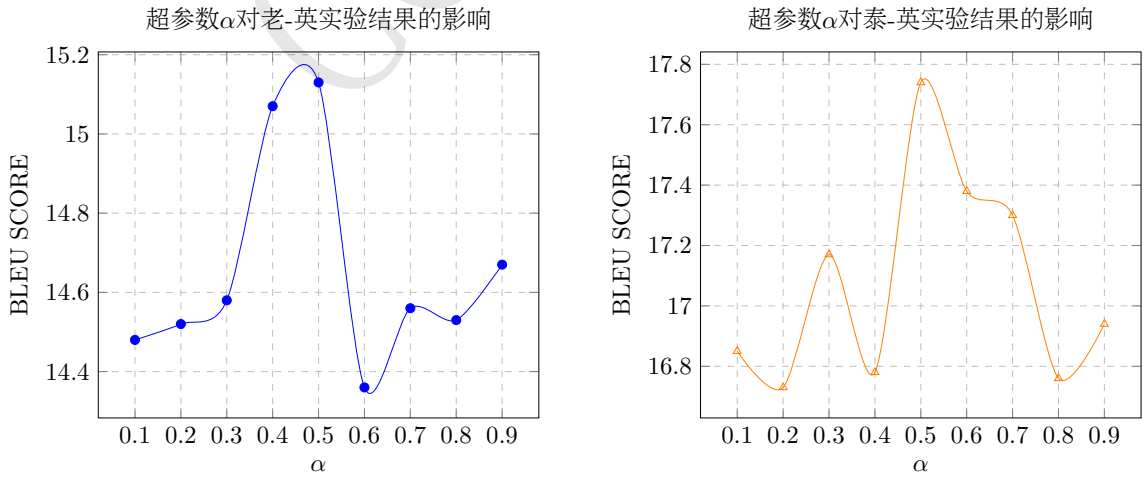


Figure 4: 超参数$\alpha$对老-英、泰-英实验结果的影响

如图4左图所示，在老-英翻译方向上，当文本占比小于音素占比时，$\alpha$从0.1变化到0.4，实验性能逐步上升，在0.5时逐渐达到最优，此时文本特征和音素特征比例达到平衡，当$\alpha$达

到0.6时，意味着文本占比大于音素占比，实验性能降到最低，这说明音素特征在占比0.4时对文本特征干扰最大，当$\alpha$达到0.9时，此时，文本占比达到最高，性能相较$\alpha$在0.6到0.8有所提升，说明此时的文本特征相较音素特征对翻译性能影响更大。

如图4右图所示，在泰-英翻译方向上，当音素占比小于0.5时，此时的文本占比大于0.5，音素特征占比较小在一定程度上为模型引入了噪声，因此模型翻译效果不佳。当$\alpha$达到0.9时，音素特征占比极小，对文本特征的干扰也达到最小，性能相较$\alpha$在0.6到0.8有所提升。

我们得到结论，在泰-英和老-英翻译方向上，模型对于$\alpha$的取值较为敏感，当音素特征和文本特征占比趋近0.5时文本特征和音素特征比例达到平衡状态，可使模型更好地学习到相似特征，从而性能逐渐达到最优。

### 5.3.5 翻译实例分析

本文分别以泰语-英语和老挝语-英语的翻译结果为例，分析融入音素特征并基于参数分化微调对模型生成译文质量的影响。

| 老-英翻译示例 | |
|---|---|
| 源语言 | **ລັດຖະບານທະຫານ**ບໍ່ຍອມຮັບ**ສຽງສ່ວນໃຫຍ່ຂອງNLD** ໃນ**ການເລືອກຕັ້ງ**ຄັ້ງນັ້ນ. |
| Baseline | The **NLD government** has not accepted **the majority of the polls**. |
| 本文方法 | **military government** does not accept **the majority of the NLD** in the **election**. |
| 参考译文 | The **military junta** did not accept **NLD 's majority** in that **election**. |
| **泰-英翻译示例** | |
| 源语言 | Hans ผู้เป็น**สามี**เคยเป็น**นักโทษ**สมัยสงครามโลกครั้งที่สอง และไปทำงานเป็นชาวนาให้ครอบครัวของ Josie |
| Baseline | Hans was the second World War II and to work as a procedure for Josie family. |
| 本文方法 | Hans, a **husband**, who had been the second World War **prisoner**, and worked as Josie 's family. |
| 参考译文 | **Husband** Hans was a **prisoner** of war in World War II, and went to work as a farmer for Josie 's family. |

Table 6: 老-英、泰-英翻译示例

如图6所示，基线模型会出现部分关键词漏译、错译的现象，例如，在老挝语-英语的翻译中，基线模型输出的英语句子中漏译了“election”，错译了“military junta”和“NLD 's majority”；泰语-英语翻译结果类似，基线模型输出的英语句子也漏译了“husband”和“prisoner”，这是由于基线模型泰老词表征没有统一，知识迁移效果不佳。而本文方法有效缓解了基线模型中的错译、漏译问题，这充分证明了在模型中融入额外音素相似特征可以有效拉近泰老之间的语义距离，同时通过基于参数分化的微调策略也可以有效缓解模型过度泛化的问题。

## 6 结论

针对现有多语言神经机器翻译方法难以得到统一词表征形式的问题，本文利用泰语和老挝语音素相似性拉近泰老之间的语义距离，并在微调阶段基于参数分化缓解联合训练造成模型过度泛化的问题。实验结果证明了本文方法的有效性和优越性，在老挝语-英语和泰语-英语翻译任务的BLEU值分别达到了15.40和17.99，比基线模型均有明显提升。我们的工作不仅利用多语言模型提升了低资源语言泰语、老挝语到英语的翻译性能，我们还探究了泰语、老挝语之间的音素相似性特征，并将其融合到模型中，有效地改善了在多语言翻译过程中低资源语言由于数据稀缺以及语言特性导致的共享语言知识困难的问题。此外，我们的方法进一步为更多低资源的相似性语言机器翻译任务提供了较好的思路。

第二十一届中国计算语言学大会论文集，第305页-第316页，南昌，中国，2022年10月14日至16日。

## 参考文献

Thanin Asawavichienjinda, Kammant Phanthumchinda, Chitr Sitthi-Amorn, and Edgar J Love. 2005. The thai version of the quality-of-life in epilepsy inventory (qolie-31-thai version): translation, validity and reliability. *JOURNAL-MEDICAL ASSOCIATION OF THAILAND*, 88(12):1782.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Ankur Bapna, Naveen Arivazhagan, and Orhan Firat. 2019. Simple, scalable adaptation for neural machine translation. *arXiv preprint arXiv:1909.08478*.

Ankur Bapna, Orhan Firat, Pidong Wang, Wolfgang Macherey, Yong Cheng, and Yuan Cao. 2022. Multilingual mix: Example interpolation improves multilingual neural machine translation.

Pau Baquero-Arnal, Javier Iranzo-Sánchez, Jorge Civera, and Alfons Juan. 2019. The mllp-upv spanish-portuguese and portuguese-spanish machine translation systems for wmt19 similar language translation task. In *Proceedings of the Fourth Conference on Machine Translation (Volume 3: Shared Task Papers, Day 2)*, pages 179–184.

Chenchen Ding, Masao Utiyama, and Eiichiro Sumita. 2016. Similar southeast asian languages: Corpus-based case study on thai-laotian and malay-indonesian. In *Proceedings of the 3rd Workshop on Asian Translation (WAT2016)*, pages 149–156.

Daxiang Dong, Hua Wu, Wei He, Dianhai Yu, and Haifeng Wang. 2015. Multi-task learning for multiple language translation. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1723–1732.

Melvin Johnson, Mike Schuster, Quoc V Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, et al. 2017. Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5:339–351.

Albina Khusainova, Adil Khan, Adín Ramírez Rivera, and Vitaly Romanov. 2021. Hierarchical transformer for multilingual machine translation. *arXiv preprint arXiv:2103.03589*.

Philipp Koehn. 2004. Statistical significance tests for machine translation evaluation. In *Proceedings of the 2004 conference on empirical methods in natural language processing*, pages 388–395.

Zehui Lin, Liwei Wu, Mingxuan Wang, and Lei Li. 2021. Learning language specific sub-network for multilingual machine translation. *arXiv preprint arXiv:2105.09259*.

Zhibo Man, Cunli Mao, Zhengtao Yu, Xunyu Li, Shengxiang Gao, and Junguo Zhu. 2020. 基于多语言联合训练的汉-英-缅神经机器翻译方法(chinese-english-burmese neural machine translation method based on multilingual joint training). In *Proceedings of the 19th Chinese National Conference on Computational Linguistics*, pages 446–456.

Atul Kr Ojha, Ritesh Kumar, Akanksha Bansal, and Priya Rani. 2019. Panlingua-kmi mt system for similar language translation task at wmt 2019. In *Proceedings of the Fourth Conference on Machine Translation (Volume 3: Shared Task Papers, Day 2)*, pages 213–218.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Suwannee Phitakwinai, Sansanee Auephanwiriyakul, and Nipon Theera-Umpon. 2008. Thai sign language translation using fuzzy c-means and scale invariant feature transform. In *International Conference on Computational Science and Its Applications*, pages 1107–1119. Springer.

Alberto Poncelas, Wichaya Pidchamook, Chao-Hong Liu, James Hadley, and Andy Way. 2020. Multiple segmentations of thai sentences for neural machine translation. *arXiv preprint arXiv:2004.11472*.

Kanchana Saengthongpattana, Kanyanut Kriengket, Peerachet Porkaew, and Thepchai Supnithi. 2019. Thai-english and english-thai translation performance of transformer machine translation. In *2019 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*, pages 1–5. IEEE.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, 27.

Xu Tan, Jiale Chen, Di He, Yingce Xia, Tao Qin, and Tie-Yan Liu. 2019. Multilingual neural machine translation with language clustering. *arXiv preprint arXiv:1908.09324*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Qian Wang and Jiajun Zhang. 2021. Parameter differentiation based multilingual neural machine translation. *arXiv preprint arXiv:2112.13619*.

Yining Wang, Long Zhou, Jiajun Zhang, Feifei Zhai, Jingfang Xu, 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*, pages 1213–1223.

Shuo Wang, Zhixing Tan, and Yang Liu. 2022. Integrating vectorized lexical constraints for neural machine translation. *arXiv preprint arXiv:2203.12210*.

Wanying Xie, Yang Feng, Shuhao Gu, and Dong Yu. 2021. Importance-based neuron allocation for multilingual neural machine translation. *arXiv preprint arXiv:2107.06569*.

Hongfei Xu, Qiuhui Liu, Josef van Genabith, and Deyi Xiong. 2021. Modeling task-aware mimo cardinality for efficient multilingual neural machine translation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 361–367.

Zhiqiang Yu, Zhengtao Yu, Yuxin Huang, Junjun Guo, Zhenhan Wang, and Zhibo Man. 2020. Transfer learning for chinese-lao neural machine translation with linguistic similarity. In *China Conference on Machine Translation*, pages 1–10. Springer.

Biao Zhang, Ankur Bapna, Rico Sennrich, and Orhan Firat. 2020a. Share or not? learning to schedule language-specific capacity for multilingual translation. In *International Conference on Learning Representations*.

Biao Zhang, Philip Williams, Ivan Titov, and Rico Sennrich. 2020b. Improving massively multilingual neural machine translation and zero-shot translation. *arXiv preprint arXiv:2004.11867*.

Yaoming Zhu, Jiangtao Feng, Chengqi Zhao, Mingxuan Wang, and Lei Li. 2021. Counter-interference adapter for multilingual machine translation. *arXiv preprint arXiv:2104.08154*.

Barret Zoph and Kevin Knight. 2016. Multi-source neural translation. *arXiv preprint arXiv:1601.00710*.