

基于枢轴语言系统融合的词汇混淆神经网络机器翻译

赵小兵^{1,2} 金波^{1,2} 孙媛^{1,2,*}

¹中央民族大学 信息工程学院, 北京 100081

²国家语言资源监测与研究少数民族语言中心

*Corresponding author: Yuan Sun

nmzxb_cn@163.com, 20301807@muc.edu.cn, tracy.yuan.sun@qq.com

摘要

神经机器翻译在低资源语言的翻译任务中存在翻译难度大、译文质量不佳的问题。本文针对低资源语言与汉语之间没有双语平行语料的情况,采用正反向枢轴翻译的方法,生成了三种低资源语言到汉语的平行句对,采用词汇级的系统融合技术,将Transformer模型和对偶学习模型翻译生成的目标语言译文进行融合,然后通过混淆神经网络进行词汇选择,生成了更为优质的目标语言译文。实验证明,本文提出的多模型融合方法在爱沙尼亚语-汉语、拉脱维亚语-汉语、罗马尼亚语-汉语这三种低资源语言翻译任务中均优于独立模型的翻译效果,进一步提升了低资源语言神经机器翻译的译文质量。

关键词: 神经机器翻译; 枢轴语言; 系统融合; 混淆网络; 低资源语言

Neural Machine Translation for Vocabulary Confusion Network Based on Pivotal Language System Fusion

Xiaobing Zhao^{1,2} Bo Jin^{1,2} Yuan Sun^{1,2,*}

¹ School of Information Engineering, Minzu University of China

² Natural Language Resource Monitoring and Research Center of Minority Languages

*Corresponding author: Yuan Sun

nmzxb_cn@163.com, 20301807@muc.edu.cn, tracy.yuan.sun@qq.com

Abstract

Neural machine translation (NMT) has many problems in the translation of low resource languages, such as difficulty in translation and poor quality of translation. Aiming at the situation that there is no bilingual parallel corpus between low resource language and Chinese, this paper adopts the forward and backward pivot translation method to generate three parallel sentence pairs from low resource language to Chinese, and integrates the target language translations generated by Transformer model and dual learning model with the lexical system fusion technology. Then the confusion neural network is used to select words and produce a better target language translation. Experiments show that the multi-model fusion method proposed in this paper is superior to the independent model in Estonian-Chinese, Latvian-Chinese and Romanian Chinese-Chinese translations of the three low-resource languages, and further improves the translation quality of neural machine translation of low-resource languages.

Keywords: Neural machine translation, Pivot language, System combination, Confusion network, Low-resource language

1 引言

机器翻译 (Machine Translation, MT) 是指通过计算机将源语言句子翻译到与之等价的目标语言句子的过程, 是自然语言处理领域的重要研究方向之一 (冯洋 et al., 2020)。2013年基于深度学习模型的神经机器翻译 (Neural Machine Translation, NMT) 被提出后 (Kalchbrenner et al., 2013), 发展极其迅速, 从最初的RNN结构 (Sutskever et al., 2014), 到基于注意力机制的模型 (Bahdanau et al., 2014), 再到目前最流行的Transformer架构 (Vaswani et al., 2017), NMT的翻译技术和质量不断得到改善和提高。

NMT的成功与数据资源、算法模型和算力资源密不可分, 尤其依赖于大规模、高质量的双语平行语料。但是, 世界上很多语言的语料资源十分匮乏, 甚至缺失, 想要获取高质量的双语平行语料往往需要很高的代价。目前, 除了汉语、英语等少量几种资源丰富的语言外, 大多数语言都无法找到大规模的双语平行语料。受限于数据因素的制约, NMT在低资源语言中的翻译效果仍然并不理想 (Koehn et al., 2017; 哈里旦木·阿布都克里木 et al., 2017)。

低资源语言的机器翻译通常有两种情况, 一种是源语言和目标语言之间只有小规模的双语平行语料, 这种情况的双语翻译, 可以通过词替换 (Fadaee et al., 2017)、反向翻译 (Sennrich et al., 2015) 等方法来构造伪平行语料以达到扩充语料的目的; 另一种是源语言和目标语言之间没有平行语料, 这种情况下, 目前采用较为广泛的方法是基于枢轴语言的翻译。本文针对第二种情况, 以英语为枢轴语言, 采用源语言 \rightarrow 枢轴语言 \rightarrow 目标语言 (记作: SPT模型)、目标语言 \rightarrow 枢轴语言 \rightarrow 源语言 (记作: TPS模型)、源语言 \leftarrow 枢轴语言 \rightarrow 目标语言 (记作: PST模型) 等多种模型的翻译, 构造了爱沙尼亚语-汉语(et-zh)、拉脱维亚语-汉语(lv-zh)、罗马尼亚语-汉语(ro-zh)的双语平行句对, 实现了三种低资源语言到汉语的翻译。翻译过程中, 我们在对偶学习模型和经典transformer模型的基础上, 采用词汇级的系统融合技术, 对最终的翻译质量进行提升。实验结果表明, 本文提出的方法在et-zh、lv-zh、ro-zh的机器翻译任务中均优于单个模型的翻译效果, 尤其在ro-zh的翻译任务中, 最多可以提升3.58个BLEU值。

2 相关工作

2.1 基于枢轴语言的机器翻译

通过枢轴语料扩充法生成伪平行数据的方式, 可以有效缓解低资源语言双语平行语料匮乏的问题 (贾承勋 et al., 2021)。这种方法在之前盛行的统计机器翻译当中有一定应用 (Kauers et al., 2002; 贾承勋 et al., 2021; De Gispert et al., 2006; Utiyama et al., 2007)。随着NMT的发展, 基于枢轴语言翻译的方法逐渐应用于低资源语言的神经机器翻译当中。2016年Johnson等人 (Johnson et al., 2017) 将基于枢轴思想的方法运用到Google NMT当中, 构建了多语言翻译系统, 证实了这种方法明显优于没有增量训练的通用模型, 同时实现了国际语义表征。

基于枢轴语言扩充语料的方法, 由于从源语言到枢轴语言和从枢轴语言到源语言的翻译模型是独立训练的, 所以Cheng等人 (Cheng et al., 2019) 提出了一种基于枢轴的神经机器翻译联合训练算法。通过生成并共享枢轴语言的词向量, 以及对小规模源语言-目标语言双语语料进行最大似然估计的方法, 将源语言-目标语言的翻译模型联系起来, 进行联合训练。

尽管基于枢轴语言的神经机器翻译缓解了源语言与目标语言之间平行句对缺少的问题, 但是, 在训练过程中, 由于源语言、枢轴语言、目标语言之间的关系不是非常紧密, 甚至毫无关系等原因, 从源语言到枢轴语言的翻译模型中产生的错误会传递到枢轴语言到目标语言的翻译模型中, 容易产生错误传播等问题 (Zheng et al., 2017), 这一点在我们的实验中也得到了证实; 除此之外, 两种翻译模型单独训练导致的解码时间加长等问题, 也影响着神经机器翻译的性能。而后, Chen等人 (Chen et al., 2017) 针对零资源机器翻译提出了Teacher-Student框架, 其主要思想是为了源语言-目标语言的低资源翻译模型 (“学生”), 可以利用枢轴语言-目标语言的翻译模型 (“老师”) 为指导, 在源语言-枢轴语言的平行语料库上训练出一个源语言-目标语言的NMT模型。该框架提出了句子级和词汇级的两种“老师”模型, 允许翻译模型直接进行参数估计, 而无需将解码过程分解为两个步骤, 因此既提升了效率, 又避免了错误传递。

2.2 多系统融合技术

对于同一机器翻译任务, 不同模型的翻译结果可能不同, 系统融合技术就是充分利用各模型翻译结果中的有用信息, 在特定规则的基础上, 重新生成一个翻译译文的技术。李茂

西 (李茂西 et al., 2010)将系统融合技术应用到机器翻译的方法分为句子级系统融合 (Rosti et al., 2007)、短语级系统融合 (Huang et al., 2007)、词汇级系统融合 (Rosti et al., 2007)三类。

这三类方法在文本融合过程中筛选的粒度上有所不同,因此在融合算法和翻译结果上有一定的差异。句子级系统融合技术将经过多个翻译系统翻译后产生的多个翻译假设,利用贝叶斯风险解码或者重打分的方法,从中选择一个最优的假设作为最后的输出。这种方法不会产生新的翻译,只是在现有的翻译中挑选出最优的一个,比较依赖于系统翻译效果。短语级系统融合技术通过合并所有系统的短语表,从中抽取一个新的源语言到目标语言的短语表,从新的短语表和语言模型中去重新解码源语言句子,这种方法在源语言和目标语言之间没有平行句对的情况下,难以生成源语言到目标语言的短语表。为了得到更好的译文效果,同时能够充分利用低资源语言平行句对,我们将目前流行的transformer模型和对偶学习模型 (Sestorain et al., 2018)分别产生的翻译结果进行了合并,并采用词汇级系统融合技术对目标语言的译文进行优化,最终通过实验结果证实了这一点。词汇级系统融合技术在下文中着重介绍。

3 总体思路

本文总体思路主要包括四个方面:语料生成、训练模型、汇总译文和模型融合。总体结构如图1所示。

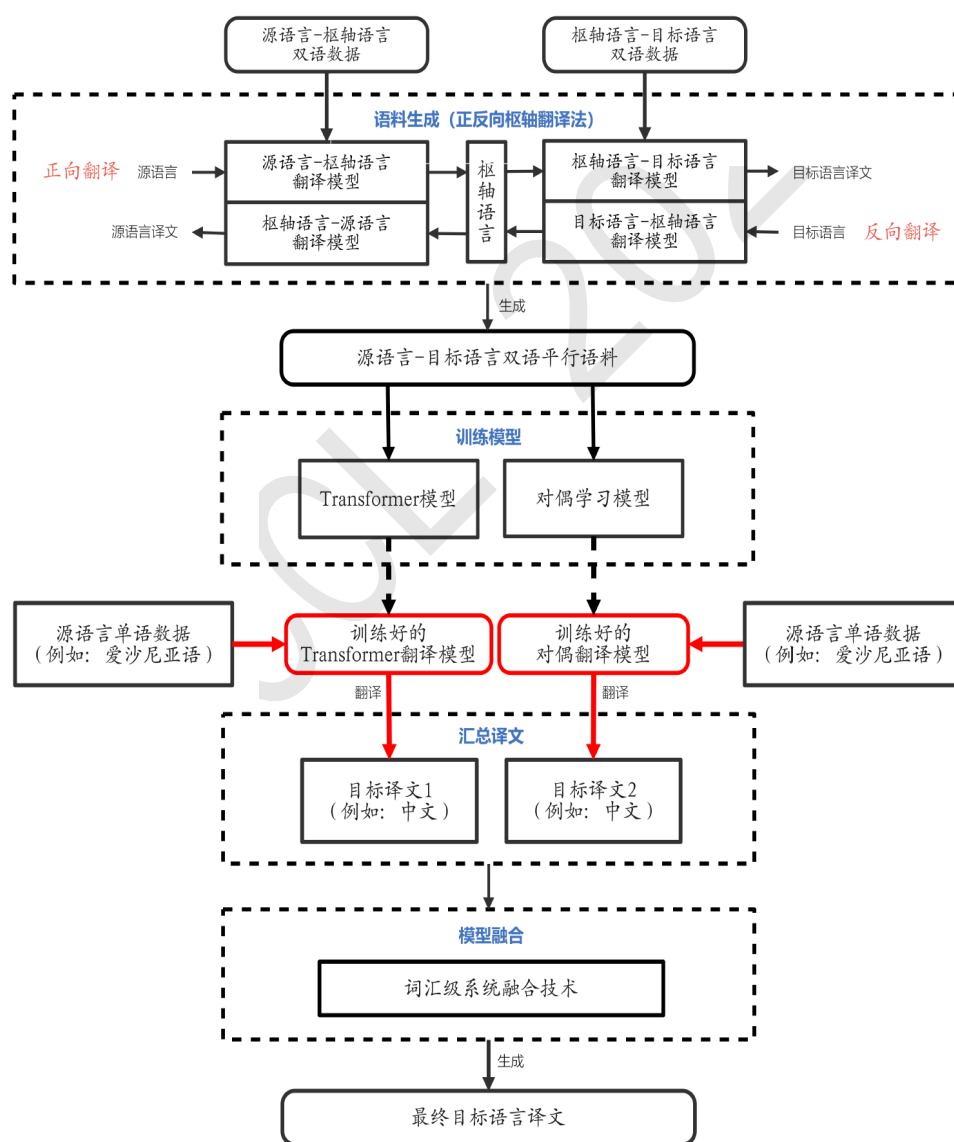


图 1. 总体结构

语料生成 从语料库中获取源语言-枢轴语言（记作：S-P）、枢轴语言-目标语言（记作：P-T）的双语平行语料。通过获取的语料分别双向训练Transformer模型，得到以下四个不同方向的基础翻译模型：源语言-枢轴语言翻译模型（记作：S-P模型）、枢轴语言-源语言翻译模型（记作：P-S模型）、枢轴语言-目标语言翻译模型（记作：P-T模型）、目标语言-枢轴语言翻译模型（记作：T-P模型）。在四个基础的翻译模型上，利用英语作为枢轴语言，通过正反向枢轴翻译法，生成源语言-目标语言（记作：S-T）的双语平行语料。

训练模型 利用生成的S-T双语平行语料，分别训练Transformer模型和对偶学习模型。

汇总译文 将相同的源语言单语数据，分别通过训练好的Transformer模型和对偶翻译模型进行翻译，生成各自的目标语言译文，然后汇总目标语言译文。

模型融合 通过词汇级别的系统融合技术，将汇总后的目标语言译文进行融合，然后经过混淆神经网络进行词汇选择，生成最终的更优质的目标语言译文，进而提升低资源语言的机器翻译效果。

4 模型架构

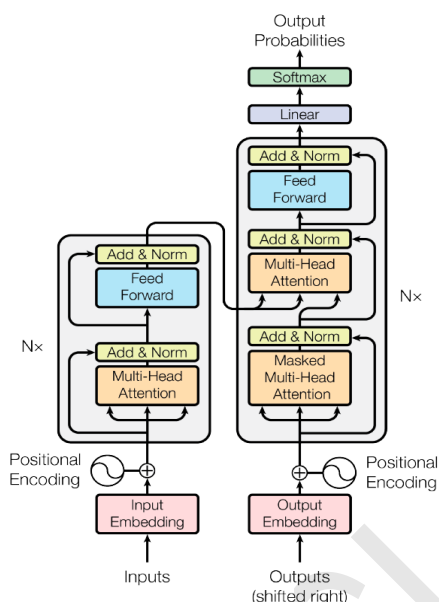


图 2. Transformer模型

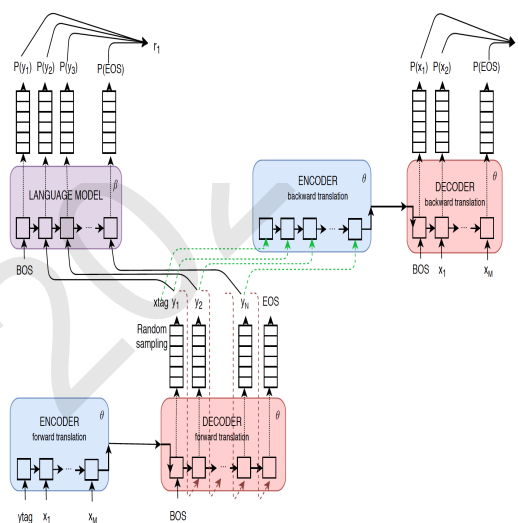


图 3. 对偶学习模型

4.1 Transformer模型

由Vaswani等人 (Vaswani et al., 2017)提出的一个完全基于注意力机制的Transformer模型，该模型创新性的使用了自注意力机制来对序列进行编码，其编码器和解码器均由注意力模块和前向神经网络构成。目前，Transformer已成为NMT研究中的主流模型，且在自然语言处理的其他领域中也有广泛应用。其模型架构如图2所示。

在给定双语平行句对的情况下，NMT模型训练的目标函数如公式(1)所示，一般表示为最大化训练语料的翻译概率，即最大似然估计(Maximum Likelihood Estimation, MLE)。

$$L(\theta) = \sum_{n=1}^N \log P(y^{(n)} | x^{(n)}; \theta) \quad (1)$$

其中， $p(y|x; \theta)$ 为翻译模型， θ 为模型参数。给定一个源语言句子序列 $X = \{x_1, x_2, \dots, x_n\}$ ，NMT的最终目标就是从解码出的候选翻译 $H(x)$ 中找到一个目标语言句子 \hat{y} ，使得公式(2)的最大似然估计最大。

$$\hat{y} = \underset{y \in H(x)}{\operatorname{arg\,max}} \log P(y | x; \theta) \quad (2)$$

Transformer模型采用一种特殊的被称为缩放点积(Scaled Dot-Product Attention)的注意力机制。基本思想为：假设输入的查询Q和键K的维度为 d_k ，值V的维度为 d_v ，则注意力的计算过程可以分为三个步骤：首先，计算查询Q和键K的点乘，除以 $\sqrt{d_k}$ ；然后，通过softmax函数计算权重进行归一化；最后，乘以值V进行加权求和，得到最终的注意力向量。如公式(3)所示。

$$\operatorname{Attention}(Q, K, V) = \operatorname{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3)$$

4.2 对偶学习模型

由Sestorain等人 (Sestorain et al., 2018)提出的对偶机器翻译模型。模型框架如图3所示。

该模型经过两次调用后进行优化，一次用于前向翻译，另一次用于后向翻译。在前向转换的过程中，编码器取一个句子 $x = y_{tag}, x_1, \dots, x_M$ ，其中 y_{tag} 是目标语言的标记，并对翻译输出 $y = y_1, y_2, \dots, y_N$ 。在每一步中，解码器从当前分布中随机选择下一个单词，并将这个词作为下一个解码步骤的输入。

然后，通过预训练的目标语言模型RNN，计算句子 y 的流畅程度。注意，首先输入特殊的句首符号，这样可以得到第一个单词的概率。然后使用对数输出来计算样本语句的概率。第一个奖励计算如公式(4)所示。

$$r_1 = \log P_Y(y) = \sum_{i=1}^N \log P_Y(y_i | y_1, \dots, y_{i-1}) \quad (4)$$

此外，执行反向翻译，句子 y 加上开始时添加的源语言标记 x_{tag} ，被编码为一个隐藏状态向量序列 $h_0, h_1, h_2, \dots, h_N$ 传入编码器。通过多层的神经网络，这些输出被传递给解码器，解码器在每一步都通过原始输入单词 x_i 和对数输出来计算源句 x 经由 y 重构的概率，这个值作为第二个奖励，如公式(5)所示。

$$r_2 = \log P_\theta(x | y) = \sum_{i=1}^M \log P_\theta(x_i | x_1, \dots, x_{i-1}, h_0, h_1, \dots, h_N) \quad (5)$$

需要注意的是，整个对偶过程是对等的两个步骤，从 X 和 Y 两种语言的任何一种开始均可。接下来，结合 r_1 、 r_2 以及一个超参数 α ，对句子 y 计算一个总得分 R ，如公式(6)所示。

$$R = \alpha r_1 + (1 - \alpha) r_2 \quad (6)$$

最后，在多语言神经机器翻译模型 P_θ 的基础下优化总得分，从而计算出不断优化更新的梯度，梯度是一个条件期望，对句子 y 进行抽样，得到对期望梯度的无偏估计，如公式(7)所示，以更好的确保 R 得分为最优。

$$\begin{aligned} \nabla_\theta E_{y|x}[R] &= \nabla_\theta \sum_y P_\theta(y|x) R(y) \\ &= \sum_y \nabla_\theta P_\theta(y|x) R(y) + P_\theta(y|x) \nabla_\theta R(y) \\ &= E_{y|x}[R(y) \nabla_\theta \log P_\theta(y|x) + (1 - \alpha) \nabla_\theta \log P_\theta(x|y)] \\ &\approx \frac{1}{K} \sum_i^k R(y_i) \nabla_\theta \log P_\theta(y_i|x) + (1 - \alpha) \nabla_\theta \log P_\theta(x|y_i) \end{aligned} \quad (7)$$

4.3 模型融合

针对同一个源语言句子，多模型融合就是充分利用不同模型生成的译文，分别抽取其中的有用信息，归纳得到最终的目标语言译文。本文采用词汇级的系统融合技术。主要步骤为：首先，从参与融合的译文选择一个对齐参考，将其他的译文对齐到该对齐参考上，通过不同译文间的词对齐信息建立混淆网络；然后，对混淆网络中每两个节点间弧线上的候选词进行置信度估计；最后，将候选词的置信度结合语言模型、长度惩罚、插入惩罚等特征进行混淆网络解码，找到一个最优的路径，并输出路径上的每个词，生成优化后的译文 (李茂西 et al., 2010)。

4.3.1 构建混淆网络

混淆网络是一个包括多个节点的有向图，每个节点表示一个状态，节点之间有严格的偏序关系，相邻节点之间存在一个或多个有向边，每条边都对应一个词，从起始节点到达终点的每一条有向路径都表示一个句子 (朱俊国 et al., 2019)。在混淆网络的构建中，我们采用多混淆网络的方法 (Rosti et al., 2007)。该方法的核心思想是：在参与融合的译文中，轮流将每一个译文作为对齐参考，其余译文与对齐参考进行对齐，得到多个单混淆网络，再通过一个新的起始节点将这些单个的混淆网络连接在一起，构成一个多混淆网络。对齐的过程我们采用一种基于TERp (Snover et al., 2009)的词对齐方法 (Barrault et al., 2010)]，实现对齐参考与其他译文的融合。相对于早期TER (Snover et al., 2006)的5种基本操作：插入、删除、匹配、替换、移位，TERp方法中增加了词干匹配、同义匹配以及复数匹配 (朱俊国 et al., 2019)。下面以一个例子来说明混淆网络的构建，给定三个参与融合的译文句子 (以英文为例)，如表1所示。

表 1. 译文句子

I did not see such a beautiful scene.
I have never watch anything like this before.
I have never enjoy any seductive view.

从参与融合的译文句子中选择第一个译文作为对齐参考，并将其他两个译文对齐到该对齐参考上。对齐结果如表2所示。对齐过程中，对齐结构的更新需要满足如下条件：第一，在对齐结构中存在对齐的词 (对应匹配、替换、移位操作)，则在相应的位置上增加一条边；第二，在对齐结构中不存在对齐的词 (对应插入操作)，则在对齐结构前一位置后插入一个空词，并使空词和新加入的词进行对齐；第三，最后在对齐结构中没有对齐的位置 (对应删除操作) 插入空词 (朱俊国 et al., 2019)。

表 2. 对齐结果

I	did	not	see	such	a	beautiful	scene
I	have	never	watch	anything	like	this	before
I	have	never	enjoy	any	seductive	null	view

根据上述对齐结果，对同一位置的相同的词进行合并，通过不同译文间的词对齐信息建立混淆网络，其结构如图4所示。

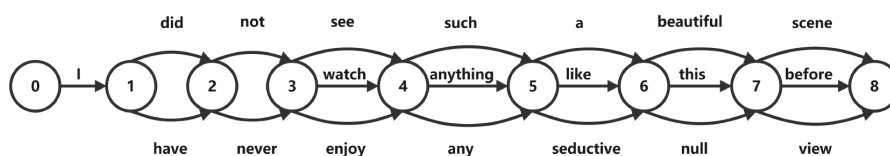


图 4. 根据词对齐生成的一个单混淆网络

在构建好的混淆网络中，每两个节点之间的弧线上的词表示它们是最初融合结果中在相应位置的候选词。对于混淆网络节点*i*和*i + 1*弧线上的候选词中，第*j*个候选词的置信得分，如公式 (8) 所示。

$$\omega_{i,j} = \mu \sum_{\mu=1}^N \sum_{v=1}^N \tau_{\mu} \tau_v C_{\omega} \quad (8)$$

上式给出了在有N个模型，每个模型提供N个译文参与融合时，词的置信度计算公式。其中， τ_u 是系统 u 对应的先验概率， τ_v 是词所在译文的权重，一般采用均匀权重，但是有时为了给排名靠前的译文中的词赋以更高的权重，也可以采用基于排名的权重，即出自第 v 个译文中的每一个词的概率都要乘上 $\frac{1}{1+v}$ ， C_{ω} 是第 u 个模型第 v 个译文中的词，如果在混淆网络节点 i 和 $i+1$ 之间的弧线上出现候选词，则该值取1，否则取0。 u 为归一化因子，它保证在节点 i 和 $i+1$ 之间出现的所有候选词的总置信度为1 (李茂西 et al., 2010)。

4.3.2 混淆网络解码

混淆网络采用的是一种投票策略。在合并同一个位置上的词时，需要计算每一个在该位置上的词的后验概率。一般地，在基于混淆网络的融合模型中该值受两个因素的影响，一个是融合译文的先验概率，另一个是词所在译文的权重 (朱俊国 et al., 2019)。假设给定一个源语言的句子 F ，混淆网络解码就是求满足下面式 (9) 中的目标语言句子。

$$E^* = \underset{E}{\operatorname{argmax}} (\alpha \log P_{AL} + \beta N_{nulls}(E) + \gamma \log P_{LM} + \delta N_{words}(E)) \quad (9)$$

其中 α 、 β 、 γ 、 δ 分别对应融合过程中产生翻译假设的词的置信度 P_{AL} 、插入惩罚 $N_{nulls}(E)$ 、语言模型得分 P_{LM} 、长度惩罚 $N_{words}(E)$ 的权重。

5 实验结果与分析

5.1 语料设置

本文的实验语料来源为CWMT2018¹和欧洲平行语料库²。训练集语料如表3所示：

表 3. 语料使用情况

双语平行语料	语料规模 (万句)	语料来源
爱沙尼亚语-英语(et-en)	65	欧洲平行语料库
拉脱维亚语-英语(lv-en)	63	
罗马尼亚语-英语(ro-en)	40	
英语-汉语(en-zh)	600	CWMT2018

对于源语言，如爱沙尼亚语(et)、拉脱维亚语(lv)和罗马尼亚语(ro)的单语数据，采用Moses脚本³对单词进行标记化和大小写转换。对于目标语言，如汉语(zh)，采用中国科学院开发的中文分词工具⁴进行分词处理。选用英语(en)作为枢轴语言。

因为et、lv和ro属于低资源语言，它们与zh之间的语料资源十分匮乏，甚至缺失。本文实验所用的测试集为从欧洲平行语料库中分别随机抽取2000句et、lv和ro的单语数据，通过Google Translator API⁵进行翻译，获得对应的zh译文，最终组成et-zh、lv-zh和ro-zh的双语平行语料。这与Unanue等人 (Unanue et al., 2018)的做法一致。

5.2 结果与分析

分别利用et-en、lv-en、ro-en、en-zh等语料，通过正向和反向两种方式训练Transformer模型，得到基于枢轴传播的四种基础翻译模型，即：S-P模型、P-S模型、P-T模型、T-P模型。其中，S为低资源语言（例如：et、lv、ro），P为英语，T为汉语。实验结果如表4所示：

¹<http://www.cipsc.org.cn/cwmt/2018/>

²<http://www.statmt.org/europarl/>

³<https://github.com/moses-smt/mosesdecoder.git>

⁴<http://ictclas.nlpir.org/>

⁵<https://github.com/matheuss/google-translate-api.git>

表 4. 枢轴语言翻译结果

翻译任务	BLUE
英语-爱沙尼亚语(en-et)	29.90
爱沙尼亚语-英语(et-en)	45.29
拉脱维亚语-英语(lv-en)	51.39
英语-拉脱维亚语(en-lv)	42.13
罗马尼亚语-英语(ro-en)	72.94
英语-罗马尼亚语(en-ro)	63.27
英语-汉语(en-zh)	32.55
汉语-英语(zh-en)	21.25

在训练好的四个基础翻译模型上，通过正反向枢轴翻译法和S-T的双语平行语料（et-zh、lv-zh、ro-zh）训练模型。为了对比不同S-T语料下Transformer模型的效果，我们通过S、P、T和四个基础翻译模型，构建了基于枢轴语言的四种模型，如表5所示。

表 5. 枢轴模型

模型名称	数据类型	模型说明
SPT模型	训练集	通过S-P模型、P-T模型，构建S-T的双语平行语料，训练Transformer模型。
PST模型	训练集	通过P-S模型、P-T模型，构建S-T的双语平行语料，训练Transformer模型。
TPS模型	训练集	通过T-P模型、P-S模型，构建S-T的双语平行语料，训练Transformer模型。
TSPT模型	测试集	通过S-P模型、P-T模型，构建S-T的双语平行语料，训练Transformer模型。

在爱沙尼亚语-汉语、拉脱维亚语-汉语、罗马尼亚语-汉语的双语平行语料上，四种模型的表现如表6所示：

表 6. 不同模型训练结果（BLEU值）

模型名称	爱沙尼亚语-汉语	拉托维亚语-汉语	罗马尼亚语-汉语
SPT模型	6.75	13.43	12.42
PST模型	2.94	4.16	0.84
TPS模型	6.14	12.27	12.09
TSPT模型	10.80	13.92	13.57

从表6可以看出，用不同模型生成的S-T双语平行语料，训练后的效果不一样。这是因为枢轴传播模型存在错误累积的问题，导致生成的S-T的语料不是特别准确。其中，PST模型的错误累积最明显。TSPT模型因其仅对测试集进行翻译转化，相对来说翻译性能较好，在三对语言翻译任务中BLEU值均超过了前面的三种模型。

接下来，分别在爱沙尼亚语-汉语、拉脱维亚语-汉语、罗马尼亚语-汉语的双语平行语料上训练对偶学习模型，实验结果如表7所示：

表 7. 对偶模型实验结果 (BLEU值)

模型名称	爱沙尼亚语-汉语	拉托维亚语-汉语	罗马尼亚语-汉语
SPT模型	6.75	13.43	12.42
PST模型	2.94	4.16	0.84
TPS模型	6.14	12.27	12.09
TSPT模型	10.80	13.92	13.57
对偶学习模型	11.25	14.83	14.24

从表7可以看出, 对偶模型相较于传统的枢轴四种模型进一步提升了翻译性能, 在三个翻译任务中, BLEU值均得到不同程度的提高, 尤其在拉脱维亚语-汉语任务中表现最好。为了进一步提升机器翻译效果, 我们采用词汇级别的系统融合技术, 分别在三对双语平行语料上将对偶学习模型和Transformer模型生成的译文进行融合, 然后训练模型, 实验结果如表8所示:

表 8. 模型融合实验结果 (BLEU值)

模型名称	爱沙尼亚语-汉语	拉托维亚语-汉语	罗马尼亚语-汉语
SPT模型	6.75	13.43	12.42
PST模型	2.94	4.16	0.84
TPS模型	6.14	12.27	12.09
TSPT模型	10.80	13.92	13.57
对偶学习模型	11.25	14.83	14.24
模型融合	13.37	17.28	17.15

从表8可以看出, 我们采用的基于词汇级的系统融合方法, 在三对语言的翻译任务中进一步提升了翻译性能, 尤其在罗马尼亚语-汉语翻译任务中, 相较于同语料下单独训练的Transformer模型和对偶学习模型, 最多可提升3.58个BLEU值。

最后, 以罗马尼亚语-汉语机器翻译任务为例, 表9展示了不同模型的翻译结果:

表 9. 罗马尼亚语-汉语翻译实例

源句子	Prioritățile politice în cadrul luptei împ-otriva imigrației clandestine a resorti-sanșilor din țările terțe
参考译文	在对第三国国民的非法移民的斗争策略优先级
SPT模型	打击第三国国民非法移民的优先策略事项
PST模型	在第三国国民非法移民的策略
TSP模型	打击第三国国民非法的政治事项
对偶模型	打击第三国国民非法移民的优先事项
模型融合	打击第三国国民非法移民的政治优先事项

由表9可以看出, SPT模型、PST模型、TSP模型和对偶学习模型只能正确翻译出部分词汇。我们提出的模型融合的方法, 综合了前面几种模型的优势。虽然存在一些误差, 但是它在语言表达上更加丰富, 并将其他模型未翻译出来的词汇进行了补充, 使得翻译结果更加接近正确译文, 进一步提升了低资源语言的翻译性能。

6 总结与展望

受限于数据因素的制约, 神经机器翻译在低资源语言的翻译任务中, 存在翻译难度大和译文质量不佳的问题。本文针对这种情况, 采用正反向枢轴翻译的方法, 生成了et-zh、lv-zh、ro-zh的双语平行句对, 缓解了低资源语言与汉语之间平行语料匮乏的问题。同时, 采用多模型融合的方法, 先将transformer模型和对偶学习模型生成的译文进行汇总, 然后通过词汇级别的系统融合技术将汇总后的译文进行融合, 生成了更优质的目标语言译文, 进而提升了低资源语言的神经机器翻译效果。实验表明, 本文提出的方法在et-zh、lv-zh、ro-zh的翻译任务中BLUE值均

有不同程度的提升。但是，本文主要针对的是源语言和目标语言之间没有平行句对的低资源语言机器翻译任务，所以在生成双语平行语料的过程中，不免存在翻译模型错误传播的问题。在接下来的工作中，我们将尝试进一步提升模型的可扩展性，引入更多模型，综合各模型的优势，研究如何减少或避免这种错误传播，同时，对译文中短语不连续或不符合语法结构的问题进行优化，以及对系统融合技术中融合性能不稳定等难题进行攻克。

致谢

感谢国家自然科学基金项目（61972436）的资助。

参考文献

- 冯洋. 邵晨泽. 2020. 神经机器翻译前沿综述. 中文信息学报. 34(11),1-18
- Nal Kalchbrenner. Phil Blunsom. 2013. Recurrent continuous translation models. *Proceedings of the 2013 conference on empirical methods in natural language processing*. 1700-1709
- I Sutskever. O Vinyals. QV Le. 2014. Sequence to sequence learning with neural networks. *arXiv preprint arXiv:1409.3215*
- Bahdanau D, Cho K, Bengio Y. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*
- Vaswani A. Shazeer N. Parmar N et al. 2017. Attention is all you need. *arXiv preprint arXiv:1706.03762*
- Koehn P. Knowles R. 2017. Six challenges for neural machine translation. *arXiv preprint arXiv:1706.03872*
- 哈里旦木·阿布都克里木, 刘洋, 孙茂松. 2017. 神经机器翻译系统在维吾尔语-汉语翻译中的性能对比. 清华大学学报(自然科学版). 57(08):878-883
- Fadaee M. Bisazza A. Monz C. 2017. Data augmentation for low-resource neural machine translation. *arXiv preprint arXiv:1705.00440*
- Sennrich R. Haddow B. Birch A. 2015. Improving neural machine translation models with monolingual data. *arXiv preprint arXiv:1511.06709*
- 贾承勋, 赖华, 余正涛, 文永华, 于志强. 2021. 基于枢轴语言的汉越神经机器翻译伪平行语料生成. 计算机工程与科学. 43(03):542-550
- Kauers M, Vogel S, Fügen C, et al. 2002. Interlingua based statistical machine translation. *Seventh International Conference on Spoken Language Processing*.
- De Gispert A, Mariño JB. 2006. Catalan-English statistical machine translation without parallel corpus: bridging through Spanish. *Proc. of 5th International Conference on Language Resources and Evaluation (LREC)*. 65-68.
- Utiyama M. Isahara H. 2007. A comparison of pivot methods for phrase-based statistical machine translation. *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*. 484-491
- Johnson M. Schuster M. Le Q V. et al. 2017. Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*. 5:339-351
- Yong Cheng. Qian Yang. Yang Liu et al. 2019. Joint training for pivot-based neural machine translation. *Joint Training for Neural Machine Translation*. Springer, Singapore. 41-54.
- Zheng H, Cheng Y, Liu Y. 2017. Maximum Expected Likelihood Estimation for Zero-resource Neural Machine Translation. *IJCAI*. 4251-4257.
- Chen Y. Liu Y. Cheng Y. et al. 2017. A teacher-student framework for zero-resource neural machine translation. *arXiv preprint arXiv:1705.00753*.
- Sestorain L, Ciaramita M, Buck C, et al. 2018. Zero-shot dual machine translation. *arXiv preprint arXiv:1805.10338*.

- Rosti A V, Ayan N F, Xiang B, et al. 2007. Combining outputs from multiple machine translation systems. *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*. 228-235.
- Huang F, Papineni K. 2007. Hierarchical system combination for machine translation. *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*. 277-286
- Rosti A V, Matsoukas S, Schwartz R. 2007. Improved word-level system combination for machine translation. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*. 312-319.
- 李茂西, 宗成庆. 2010. 机器翻译系统融合技术综述. *中文信息学报*. 24(4):74-85
- 朱俊国. 2019. 机器翻译的译后优化关键技术研究. 哈尔滨工业大学.
- Snover M G, Madnani N, Dorr B. et al. 2009. Ter-plus: paraphrase, semantic, and alignment enhancements to translation edit rate. *Machine Translation*. 23(2):117-127
- Barrault L. 2010. MANY: Open source MT system combination at WMT'10. *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and Metrics MATR*. 271-275
- Snover M, Dorr B, Schwartz R. et al. 2006. A study of translation edit rate with targeted human annotation. *Proceedings of association for machine translation in the Americas*. 200(6)
- Unanue I J, Arratibel L G, Borzeshi E Z. et al. 2018. English-Basque statistical and neural machine translation. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*.