doi:10.6043/j.issn.0438-0479.201811016

基于增量式自学习策略的多语言翻译模型

周张萍,黄荣城,王博立,胡金铭,史晓东*,陈毅东
(厦门大学信息科学与技术学院,福建 厦门 360001)

摘要:针对源语言到目标语言缺乏平行语料的情况,提出了一种基于增量式自学习策略的多语言翻译模型,即利用中介语双语语料训练源语言到目标语言的翻译模型.在 Transformer 架构下,相比于基于中介语和直接在伪平行语料上训练的普通双语翻译模型,使用该方法在第十四届全国机器翻译研讨会(CWMT 2018)多语言翻译评测数据集上的机器双语互译评估(BLEU)值提升了 0.98 个百分点.在此基础上,还对比了不同的预处理方法、训练策略以及多模型的平均和集成策略,其中多模型集成策略的 BLEU 值上可在多模型策略的基础上进一步提升 0.53 个百分点.

关键词:神经网络机器翻译;多语言机器翻译;增量式自学习
中图分类号:TP 391 **文献标志码**:A **文章编号**:0438-0479(2019)02-0170-06

第十四届全国机器翻译研讨会(CWMT 2018)机器翻译评测的英、日、汉多语言翻译任务中,仅提供专利领域的英汉、日汉双语平行数据以及汉语单语语料作为训练数据,目标是生成日英专利翻译.可见该任务为典型零资源的翻译问题.

数据驱动的机器翻译往往需要依赖大量的平行语料,为了缓解低资源语言和特定领域语料稀缺的问题,基于中介语的翻译方法和多语言翻译方法相继被提出.2001 年 Gollins 等[1]最先提出了基于中介语的翻译方法,通过中间语言桥接源语言与目标语言之间的关系.2003 年,Kishida 等[2]使用英语作为枢轴语言(pivot language)实现了从德语到意大利语的跨语言信息检索.2007 年,Wu 等[3]通过引入第三种语言实现基于短语的统计机器翻译,实验证明枢轴语言方法的机器双语互译评估(BLUE)值比法语-西班牙语翻译的直接训练模型的 BLEU 值提高了 0.06 个百分点.虽然该方法在统计机器翻译中应用较广,但错误传播会影响其翻译质量,即源语言到中介语的错译会传播给中介语到目标语言的翻译,导致源语言与目标语言的语义不一致[4].

多语言神经网络机器翻译方法的提出可以在缺乏源语言-目标语言平行语料的情况下学习源语言-目标语言之间的对应及转换关系,克服传统中介语翻译存在的错误传递问题.在早期研究中,Dong 等[5]提出了一种基于共享语义表示的多任务学习方法来扩展源语言,目的是增强翻译模型的泛化能力.该方法在源端使用单个编码器,并为每种目标任务单独使用注意力机制和解码器.Luong 等[6]针对多对多的翻译任务,分别使用单独的编码器和解码器来建模语言对.Sennrich 等[7]引入了一种跨语言共享注意机制的方法,提出了一种多对一的翻译机制,并使用生成伪平行语料的思想,使用中间语言来微调翻译模型.这些方法利用多种语言在共享语义空间内的隐含信息[8-9],缓解了多语言翻译的数据稀疏问题,提高了翻译模型性能,但是对于添加到系统中的每种语言都需要增加额外的编码器或解码器,这使得其网络结构的复杂度与系统支持的语言数量成线性关系,训练代价过高.

与上述改变模型结构的方法不同,Johnson 等[10]和 Ha 等[11]通过在输入端引入强制翻译的标签来训练多语言翻译模型.在 Johnson 等[10]的工作中,只为源端句子添加强制翻译标签以指定目标语言.作为改进,Ha 等[11]提出将语言特定标签应用于混合语言词汇表中,将不同语言的单词标注上该语言的标签并在源语

收稿日期:2018-11-10 **录用日期**:2019-02-18
基金项目:国家科技支撑计划项目(2012BAH14F03);国家自然科学基金(61573294);教育部博士点基金(20130121110040);国家语委委托项目(WT135-10);国家语委甲骨文专项(YWZ-J010)
* **通信作者**:mandel@xmu.edu.cn
引文格式:周张萍,黄荣城,王博立,等.基于增量式自学习策略的多语言翻译模型[J].厦门大学学报(自然科学版),2019,58(2):170-175.
Citation:ZHOU Z P, HUANG R C, WANG B L, et al. Multilanguage translation model based on incremental self-learning strategy[J]. J Xiamen Univ Nat Sci, 2019, 58(2): 170-175. (in Chinese)

http://jxmu.xmu.edu.cn

言句首添加一个强制翻译标签.这种在源语言中加入强制翻译标签的方法,可以在不添加额外编码器、解码器和注意力机制的情况下训练多语言翻译模型.

虽然多语言机器翻译能克服翻译的错误传递,缓解数据稀缺等问题,但是数据稀缺仍是翻译质量的一大瓶颈.因此在前人的基础上,本研究提出了基于增量式自学习方法的多语言神经网络机器翻译模型,利用其生成源语言到目标语言的伪平行语料,并将其应用于 CWMT 2018 日英翻译任务中,尝试在缺少日英平行句对的情况下实现日英翻译.

1　基于增量式自学习的多语言翻译模型

1.1　多语言翻译模型训练数据构建

受到 Ha 等[12]工作的启发,本研究在源语言的句首加上了目标语言标签,例如源语言为英语,目标语言为日语,标签则为＜2JA＞.同时,在句对分词后的每个单词前加上"＜源语言＞"格式的标签,以标识不同的语言.例如,英语句子:"I_love_you",指定目标语言为日语时该句被标注为"＜2JA＞_＜EN＞I_＜EN＞love_＜EN＞you",并将词与标签共同作为多语言翻译模型的输入.

1.2　增量式自学习

使用增量式自学习策略进行翻译模型优化的方法在统计机器翻译中已被采用.例如,Béchara 等[13]提出迭代地使用翻译的输出来改进后续训练任务.Bertoldi 等[14]应用反向翻译[3]机制改进统计机器翻译(SMT)基线模型.Luong 等[6]采用反向翻译改进神经网络机器翻译(NMT)基线模型.Xia 等[15]提出了一种对偶学习机制,其中两个相反翻译方向的 NMT 模型为彼此提供翻译知识,以实现从单语数据中学习翻译模型.Lakew 等[16]采用自学习方法利用强制翻译策略生成平行句对,并进行多次迭代训练.基于增量式自学习策略的多语言翻译模型迭代的具体算法如下所示(本文中的多语言翻译模型为 NMT 模型,选用 RNNSearch 和 Transformer 架构).

$D=(\mathrm{src},\mathrm{tgt})$/* 输入数据 D,src 是源语言语料,tgt 是目标语言语料 */
　用 D 初始化多语言翻译模型 M.
　repeat
　　for $n=1$ to $|D|$ do
　　　使用多语言翻译模型 M 从 src_n 生成 src_n^*,从 tgt_n 生成 tgt_n^*
　　end for
　　$D^*=(\mathrm{src},\mathrm{tgt})\cup(\mathrm{src},\mathrm{src}^*)\cup(\mathrm{tgt}^*,\mathrm{tgt})$.
　　使用 D^* 更新多语言翻译模型 M,
　until 多语言翻译模型 M 收敛.

这种自学习模型优化方法已成功用于 NMT 中,而本研究借鉴文献[16]的方法,与其增量式学习策略不同的有以下 3 个方面:1) 通过使用双语模型合成的源语言-目标语言语料进行基线系统的训练.因为有了伪平行语料,所以模型的训练更容易收敛,迭代的次数也会相应地减少,这样也加入了更多的信息以提升最终效果.2) 从现有的源语言-中间语言,中间语言-目标语言两对平行语料中抽取小部分源语言-目标语言的平行语料,并进行过采样,用于模型训练调优.3) 使用多语言模型本身生成的源语言-目标语言、目标语言-源语言两个方向的语料进行增量式训练,相互提升这两个翻译方向的翻译质量,进而实现模型的整体优化.

2　数据处理

数据是机器翻译的基础,而对数据的有效处理更是特别关键的一步.本节主要介绍数据的预处理和伪平行语料的生成方法.

2.1　数据预处理

在多语言翻译任务中,本研究采用了 CWMT 2018 发布的所有训练语料,包括日汉和英汉的训练数据各 300 万句对,开发集数据包括日汉、英汉、日英 3 个方向各 3 000 句对.对其进行换行符及组合字符的规范化、转义字符还原、全半角转换、乱码及控制字符过滤、长度及对齐 fast_align(https://github.com/clab/fast_align)过滤等预处理.对于汉语语料,本研究使用斯坦福分词工具进行分词,并采用基于规则的方法修复部分分词错误;对于英语语料,本研究使用 Moses(http://statmt.org/moses/)脚本进行词例化和大小写处理;对于日语语料,本研究使用 Mecab 工具进行分词.

2.2　亚词切分

为了处理低频词问题,本研究使用基于亚词切分(subword)[17]的翻译方法.在实验过程中,对比了字节对编码(BPE)与谷歌提供的开源的自然语言处理工具包 SentencePiece 的处理方式.为了平衡不同语言的语料规模,日、汉、英的词语数量按 1∶1∶1 的比例统计出共同的词表.

2.3　扩充伪语料

本研究将官方提供的日汉、汉英 300 万句对训练

http://jxmu.xmu.edu.cn

集分别记作 JAZH、ZHEN,中文单语语料记作 ZH. 再从预处理后的 JAZH、ZHEN 两份语料中通过匹配中文端自动抽取日英平行句对,共 18 035 句对,记作 JAEN.

本研究使用增量式自学习的方法利用已有的单语语料 ZH 和双语语料 JAZH、ZHEN 生成日英双语数据. 具体策略如下:

1) 在 JAZH、ZHEN 上先分别训练汉日、汉英单语向 NMT 模型,在合并两份语料后采用基于多语言标签的方法训练汉日、汉英多语言 NMT 模型.

2) 用汉日 NMT 模型将汉英训练集的中文端翻译成日文,得到日英语料 JAEN_ZHEN-ZH2JA.

3) 用汉英 NMT 模型将日汉训练集的中文端翻译成英文,得到日英语料 JAEN_JAZH-ZH2EN.

4) 用日汉、汉英多语言 NMT 模型将中文单语语料分别翻译成日文和英文,得到日英语料 JAEN_ZH-ZH2JA-ZH2EN.

5) 用多语言 NMT 模型将汉英训练集的中文端翻译成日文,得到日英语料 JAEN_MULTI-ZHEN-ZH2JA.

6) 用多语言 NMT 模型将日汉训练集的中文端翻译成英文,得到日英语料 JAEN_MULTI-JAZH-ZH2EN.

7) 用汉日 NMT、汉英 NMT 模型分别将汉英训练集的中文语料翻译成日文和英文,得到日英语料 JAEN_ZHEN-ZH2JA-ZH2EN.

8) 用多语言 NMT 模型将汉英训练集的中文端分别翻译成日文和英文,得到日英语料 JAEN_MULTI-ZHEN-ZH2JA-ZH2EN.

为了保证训练时显存不会溢出,本研究对合成的数据按句子长度过滤,长度限制为 256 个词. 为了避免影响译文质量,再去除包含未登录词的句对. 各数据规模如表 1 所示.

表 1 各数据集的数据规模

Tab. 1 The size of datasets

数据名称	数据规模/万句对
JAZH	300
ZHEN	300
JAEN	1.803 5
JAEN_ZHEN-ZH2JA	290
JAEN_JAZH-ZH2EN	243
JAEN_ZH-ZH2JA-ZH2EN	674
JAEN_MULTI-ZHEN-ZH2JA	290
JAEN_MULTI-JAZH-ZH2EN	243

3 实 验

本研究进行了以下 5 组实验:1) RNNSearch 与 Transformer 翻译模型;2) 使用斯坦福分词(https://nlp.stanford.edu/software/segmenter.shtml)、Mecab(http://taku910.github.io/mecab/)、Moses 预处理(http://statmt.org/moses/)、英语字节对编码(byte pair encoding, BPE, https://github.com/rsennrich/subword-nmt)[17]、句块(Sentencepiece, https://github.com/google/sentencepiece)等工具进行数据处理并进行对比;3) 增量式自学习的模型训练;4) 将基于中介语翻译、扩充伪训练数据与基于增量式自学习的方法进行对比;5) 单模型、模型平均(avgbest2)和模型集成(essemble4)的对比.

3.1 架构选择和超参设置

为了实现更好的翻译效果,在架构的选择上本研究对比了清华大学机器翻译演示系统(THUMT, https://github.com/thumt/THUMT)提供的 RNNSearch 和 Transformer 两个架构,在 JAZH 数据集上进行实验,采用相同的预处理方法,在日中开发集上检验模型的翻译性能,采用 BLEU[18] 作为评价指标,实验结果如表 2 所示,可见 Transformer 架构显著优于 RNNSearch 架构. 本研究分析认为:深层模型和自注意力对于捕获长距离信息更为有效,而该翻译任务是专利领域,句子较长,所以 Transformer 的优势较为明显. 因此,本研究选择 Transformer 模型进行后续的实验.

通过大量调参实验,本研究发现当 Transformer 的批处理(batch)设置为 24 000(6 000 词/显卡×4 显卡)、训练步数设为 20 万步、其他超参数与 Transformer 默认设置一致时,训练过程能够较稳定地收敛,模型性能较优. 所以在后续实验中均采用这一设置.

表 2 神经网络机器翻译架构对比实验结果

Tab. 2 Comparative experiments of neural network machine translation architectures

架 构	BLEU4-SBP/%
RNNSearch	42.10
Transformer	44.71

注:BLEU4-SBP 值为 4 元严格惩罚的 BLEU 值,下同.

http://jxmu.xmu.edu.cn

3.2　数据处理对比实验

首先,不同语言的数据处理方法不同,则翻译效果不同. 在某些特定情况下,英语和日语分别使用 BPE 分词和 SentencePiece 分词的效果会更好. 其次,对于共享字母表的语言,在两个或更多相关语言的串联上学习 BPE 可以提高分词的一致性,并减少在复制或音译专有名词时插入或删除字符的问题. 因此,两端分开处理和联合 BPE 处理数据两种方式对于模型训练效果也会有一定的影响. 除此之外,词表大小也会影响模型的训练效果.

在数据处理方式的选择上,本研究在 Transformer 模型架构下,对 JAEN_ZHEN-ZH2JA 和 JAZH 两个合并数据集上进行实验,采用相同的超参数设置,对比了日语 SentencePiece/英语 BPE 的两端分开处理与采用不同词表大小联合 BPE 处理,对翻译结果的影响. 实验结果如表 3 所示,可见在日英翻译任务上,词表大小为 2 万个时,联合 BPE 的效果最好.

表 3　亚词处理方式对比实验结果

Tab. 3　Comparative experiments of subword segmentation methods

数据处理方法	词表的词数量/万个	BLEU4-SBP/%
日语 SentencePiece/英语 BPE	2	34.65
联合 BPE	2	35.93
	3	33.63

3.3　增量式自学习训练实验

在增量式自学习过程中,本研究采用的策略:首先,使用原始数据即评测提供的数据进行模型的训练,将该模型作为基线系统;然后,在该基线系统的基础上尝试加入不同的数据进行训练;最后,以 BLEU 值作为参考,观察在开发集上日英翻译的性能. 如表 4,仅列举了增量式自学习方法部分迭代轮次的大致情况,展示了数据规模逐渐加大,以及在同一轮次加入不一样的数据,导致模型训练的结果不一样. 可以看出第一轮在基线的基础上加入 JAEN_ZHEN-ZH2JA,其翻译效果提升明显. 第二轮在加入 JAEN_MULTI-ZHEN-ZH2JA 后,对翻译也有帮助,但是第三轮实验 1 和 2 分别加入数据集 JAEN_JAZH-ZH2EN 和数据集 JAEN_MULTI-JAZH-ZH2EN 后,BLEU 值逐渐下降,说明在某种程度上这 2 个数据不能为翻译提供有用的信息. 第四轮时在 C_4 数据集上接着加入数据集 JAEN_MULTI-ZHEN-ZH2JA-ZH2EN,此时 BLEU 值与第二轮相同,所以该数据集可以做更多实验进行观察.

由此可知,实验数据的规模以及数据的有用性对于翻译质量非常重要,鉴于时间的原因,后续将继续进行更深层次的增量式自学习实验,挖掘更多有效信息.

表 4　增量式自学习部分迭代的 BLEU 值

Tab. 4　BLEU scores in iterations of incremental self-learning

轮次	训练数据	BLEU4-SBP/%
多语言基线	C_0	31.93
第一轮	$C_1=C_0+$JAEN_ZHEN-ZH2JA	35.33
第二轮	$C_2=C_1+$JAEN_MULTI-ZHEN-ZH2JA	36.57
第三轮实验 1	$C_3=C_2+$JAEN_JAZH-ZH2EN	下降
第三轮实验 2	$C_4=C_2+$JAEN_MULTI-JAZH-ZH2EN	下降
第四轮	$C_5=C_4+$JAEN_MULTI-ZHEN-ZH2JA-ZH2EN	36.57

注:C_0 代表 ENJA, JAEN, JAZH, ZHEN,ZHJA, ENZH 几个数据集总和.

3.4　不同数据训练策略的翻译结果对比

在相同数据预处理方法及相同超参数设置的情况下,本研究进一步对以下 3 种数据训练策略进行了对比:

1) pivot:训练日汉、汉英单语向翻译模型,采用中介语翻译方法利用二者进行组合解码;

2) direct:利用小规模日英真实语料和大规模日英伪语料直接训练日英 NMT 模型;

3) multilingual:采用增量式自学习方法训练多语翻译模型.

实验结果如表 5 所示,可见增量式自学习策略的性能显著优于中介语翻译和直接在伪平行语料上训练的普通双语翻译模型. 对比表 4 与表 5 的数据可以发现,虽然多语言基线模型的性能比直接在伪平行语料上训练的双语模型低,但采用增量式自学习方法进行迭代训练,Transformer 模型的翻译性能会不断提高. 说明训练中加入的小规模真实日英平行语料对于翻译效果有提升作用,通过不断迭代生成的伪平行语料能够促进模型训练更快达到收敛,减少迭代的次数. 在完成两轮迭代训练,增量式自学习方法比直接在伪平行语料上训练的普通双语模型在日英开发集上 BLEU 值高 0.98 个百分点.

http://jxmu.xmu.edu.cn

表 5 NMT 训练策略对比实验结果

Tab. 5 Comparative experiments of NMT training strategies

方法	训练数据	BLEU4-SBP/%
pivot	JAZH, ZHEN	35.93
direct	JAEN, JAEN_ZHEN-ZH2JA, JAEN_JAZH-ZH2JA	36.84
multilingual	JAEN, JAZH, ZHEN, JAEN_ZHEN-ZH2JA, JAEN_ JAZH-ZH2JA, JAEN_MULTI-ZHEN-ZH2JA, JAEN_MULTI-JAZH-ZH2JA	37.82

3.5 多模型的平均和集成策略对比

为了进一步提升翻译质量,本研究尝试了采用多模型平均和集成策略.

多模型平均为平均可训练参数,当模型接近收敛时,这些参数在单个模型的最后时间步长进行保存.由于使用随机梯度下降算法来优化模型,所以在每个步骤中仅使用一小批数据,导致参数可能过度适应一个小批量的数据,通过模型平均可以获得更强大的参数[19].在本实验中取同一个训练过程中验证集分数最高的前后 k 个模型进行平均,取 $k=2,3,4$ 进行实验,最终发现 k 取 2 时效果最好,该模型记为 avgbest2,但仍比单模型差(如表 6 所示).

多模型集成为在预测下一个目标单词之前整合多个模型的概率分布的方法,它已被证明在神经机器翻译中有效.本实验用不同的初始化方式在相同的架构上独立训练 N 个模型,将 N 个模型以不同的初始化方式组合为一个集合模型可以避免仅做局部优化,进而获得更好的结果[19].实验中取这 N 个模型中验证集分数最高的 k 个模型进行集成解码,取 $k=2,3,4$ 进行实验,最终发现 k 取 4 时效果最好,该模型记为 ensemble4,与单模型相比 BLEU 值提升 0.53 个百分点(如表 6 所示).

表 6 平均模型和集成模型实验结果

Tab. 6 Exerimental results of average model and ensemble model

模型	BLEU4-SBP/%
multilingual	37.82
multilingual+avgbest2	37.80
multilingual+ensemble4	38.35

4 结 论

实验证明,本研究提出的基于增量式自学习策略得到多语言语料库训练 Transformer 能有效提高翻译性能.虽然在增量式自学习策略的多语言基线上得到的翻译结果不如直接翻译的方法,但利用自学习方法生成伪数据进行增量式迭代优化,Transformer 翻译模型的性能明显提升.随着迭代次数的增加,每轮迭代的提升逐渐减小,直至接近收敛.将迭代过程中的多个模型进行集成,可以进一步提升翻译质量.

虽然增量式自学习方法能够有效提升多语言翻译效果,但自学习产生的伪数据可能存在噪声,该方法在不同语言不同数据集上的适用性也有待更多实验进行验证.因此,在后续的工作中,将进行以下三个方面的研究:1) 针对伪数据可能存在噪声的问题,将尝试利用双语语言模型和双语词对齐等方法进行去噪;2) 将在不同的数据集上进行对比实验,验证本研究所提出的方法在不同语言任务上的效果;3) 尝试加入更多语言对,生成多个翻译方向的语料,进行多次增量训练,验证本研究所提出的的方法是否能有效帮助模型学习多语言间共同的翻译知识和联系,提升翻译效果.

参考文献:

[1] GOLLINS T, SANDERSON M. Improving cross language retrieval with triangulated translation[C]// International ACM SIGIR Conference on Research and Development in Information Retrieval. New Orleans: ACM, 2001: 90-95.

[2] KISHIDA K, KANDO N. Two-stage refinement of query translation in a pivot language approach to cross-lingual information retrieval: an experiment at CLEF 2003[C]// Comparative Evaluation of Multilingual Information Access Systems, Workshop of the Cross-Language Evaluation Forum, CLEF 2003. Trondheim: DBLP, 2004: 253-262.

[3] WU H, WANG H. Pivot language approach for phrase-based statistical machine translation [J]. Machine Translation, 2007, 21(3): 165-181.

[4] CHENG Y, YANG Q, LIU Y, et al. Joint training for pivot-based neural machine translation [C]// Twenty-Sixth International Joint Conference on Artificial Intelligence. Melbourne: IJCAI, 2017: 3974-3980.

[5] DONG D, WU H, HE W, et al. Multi-task learning for multiple language translation [C]// Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language

http://jxmu.xmu.edu.cn

Processing. Beijing:ACL,2015:1723-1732.

[6] LUONG M T,LE Q V,SUTSKEVER I,et al. Multi-task sequence to sequence learning[EB/OL]. [2018-11-01]. http://arxiv.org/pdf/1511.06114.

[7] SENNRICH R,HADDOW B,BIRCH A. Improving neural machine translation models with monolingual data[EB/OL]. [2016-06-03]. http://arxiv.org/pdf/1511.06709.

[8] LIU D,ZHU C,ZHAO T,et al. Pivot-based semantic splicing for neural machine translation[C]// Communications in Computer and Information Science. Singapore: Springer, 2016:14-24. doi:10.1007/978-981-10-3635-4_2.

[9] ZOPH B,YURET D,MAY J,et al. Transfer learning for low-resource neural machine translation[EB/OL]. [2018-11-01]. http://arxiv.org/pdf/1604.02201.

[10] JOHNSON M,SCHUSTER M,LE Q V,et al. Google's multilingual neural machine translation system:enabling zero-shot translation[EB/OL]. [2018-11-01]. http://arxiv.org/pdf/1611.04558.

[11] HA T L,NIEHUES J,WAIBEL A. Toward multilingual neural machine translation with universal encoder and decoder[EB/OL]. [2018-11-01]. http://arxiv.org/pdf/1611.04798.

[12] HA T L,NIEHUES J,WAIBEL A. Effective strategies in zero-shot neural machine translation[EB/OL]. [2018-11-01]. http://arxiv.org/pdf/1711.07893.

[13] BECHARA H, MA Y,, GENABITH J V. Statistical post-editing for a statistical MT system[C]// Proceedings of the Thirteenth Machine Translation Summit (MT Summit XIII). Xiamen:AAMT,2011:308-315.

[14] BERTOLDI N, FEDERICO M. Domain adaptation for statistical machine translation with monolingual resources [C]// The Workshop on Statistical Machine Translation. Association for Computational Linguistics. Athens: ACL,2009:182-189.

[15] XIA Y, HE D, QIN T, et al. Dual learning for machinetranslation[EB/OL]. [2018-11-01]. http://arxiv.org/pdf/1611.00179.

[16] LAKEW S M,LOTITO Q F,NEGRI M,et al. Improving zero-shot translation of low-resource languages[EB/OL]. [2018-11-01]. http//arxiv.org/pdf/1811.0138901.

[17] SENNRICH R.,HADDOW B.,BIRCH A. Neural machine translation of rare words with subword units[C]// Proceedings of ACL. Berlin:ACL,2016:1715-1725.

[18] PAPINENI K,ROUKOS S,WARD T,et al. IBM research report bleu: a method for automatic evaluation of machine translation[J]. Proceedings of Annual Meeting of the Association for Computational Linguistics,2002,30(2):311-318.

[19] LIU Y,ZHOU L,WANG Y,et al. A comparable study on model averaging,ensembling and reranking in NMT [C] // CCF International Conference on Natural Language Processing and Chinese Computing. Cham: Springer,2018:299-308.

Multilanguage translation model based on incremental self-learning strategy

ZHOU Zhangping,HUANG Rongcheng,WANG Boli,
HU Jinming,SHI Xiaodong,CHEN Yidong

(School of Information Science and Engineering,Xiamen University,Xiamen 360001,China)

Abstract: Without parallel corpus from the source language to the target language,we train multilingual neural machine translation models on bilingual corpus of the pivot language and propose an incremental learning strategy to improve source-language to target-language translation. Experimental results under Transformer framework show that our multilingual iterative method can improve the BLEU score by 0.98 percent point on the China workshop on machine translation (CWMT) 2018 multi-language translation evaluation data set,compared to traditional pivot-based translation and the vanilla multilingual neural machine translation (NMT). In addition,we also compared different preprocessing methods,training strategies,multi-model average and ensemble,where multi-model ensemble can further increase the BLEU score by 0.53 percent point unpon common multi-model strategy.

Keywords: neural machine translation;multilingual machine translation;iterative method

http://jxmu.xmu.edu.cn