

稀缺资源语言神经网络机器翻译研究综述

李洪政^{1,2,3,4} 冯冲^{1,2} 黄河燕^{1,2}

摘要 作为目前主流翻译方法的神经网络机器翻译已经取得了很大突破, 在很多具有丰富数据资源的语言上的翻译质量也不断得到改善, 但对于稀缺资源语言的翻译效果却仍然并不理想. 稀缺资源语言机器翻译是目前机器翻译领域的重要研究热点之一, 近几年来吸引了国内外的广泛关注. 本文对稀缺资源语言机器翻译的研究进行比较全面的回顾, 首先简要介绍了与稀缺资源语言翻译相关的学术活动和数据集, 然后重点梳理了目前主要的研究方法和一些研究结论, 总结了每类方法的特点, 在此基础上总结了不同方法之间的关系并分析了目前的研究现状. 最后, 对稀缺资源语言机器翻译未来可能的研究趋势和发展方向进行了展望, 并给出了相关建议.

关键词 稀缺资源语言, 神经网络机器翻译, 迁移学习, 枢轴语言, 回译, 多语言翻译, 无监督翻译

引用格式 李洪政, 冯冲, 黄河燕. 稀缺资源语言神经网络机器翻译研究综述. 自动化学报, 2021, 47(6): 1217–1231

DOI 10.16383/j.aas.c200103

A Survey on Low-resource Neural Machine Translation

LI Hong-Zheng^{1,2,3,4} FENG Chong^{1,2} HUANG He-Yan^{1,2}

Abstract As the mainstream approach in the field of machine translation, neural machine translation (NMT) has achieved great improvements on many rich-source languages, but performance of NMT for low-resource languages are not very good yet. Low-resource NMT has been one of the most popular issues in MT and attracted wide attention around the world in recent years. This paper presents a survey on low-resource NMT research. We first introduce some related academic activities and feasible data sets for the translation, then categorize and summarize several types of approaches mainly used in low-resource NMT in detail, and present their features, as well as relations between them, and describe the current research status. Finally, we propose some advices on possible research trends and directions of this field in the future.

Key words Low-resource language, neural machine translation, transfer learning, pivot language, back translation, multilingual translation, unsupervised translation

Citation Li Hong-Zheng, Feng Chong, Huang He-Yan. A survey on low-resource neural machine translation. *Acta Automatica Sinica*, 2021, 47(6): 1217–1231

神经网络机器翻译 (Neural machine translation, NMT) 于 2013 年正式出现^[1]. 在短短几年的时间里, 从最初的循环神经网络 (Recurrent neural

network, RNN) encoder-decoder 结构^[2], 到基于注意力机制的 RNN search 模型^[3] 及其各种变体, 再到目前最流行的 Transformer 架构^[4] 以及随后多样的预训练模型, NMT 以其独特的优势迅速成为主流的翻译方法, 翻译技术取得了巨大突破, 翻译质量也不断得到改善和提高.

NMT 的成功与算力资源、算法模型和数据资源密不可分, 尤其依赖于海量的双语数据资源. 而获取高质量的双语资源往往需要很多高昂的代价, 另一方面, 世界上目前现存的很多语言在双语数据资源方面却十分匮乏甚至缺失. 在机器翻译领域的研究中, 这些语言一般称为“稀缺资源语言”, 也称为“低资源语言” (Low-resource languages). 本文接下来会交替使用这两种术语.

在数据因素的制约下, NMT 在低资源语言中的翻译效果仍然并不理想. 而低资源语言机器翻译一直具有很多实际的需求和应用场景, 因此引起了国内外学术界和业界的广泛关注, 已经成为当前机

收稿日期 2020-03-03 录用日期 2020-05-07

Manuscript received March 3, 2020; accepted May 7, 2020

中国博士后科学基金 (2018M640069), 国家自然科学基金 (61902024, 61732005), 国家重点研发计划 (2018YFC0832104) 资助
Supported by China Postdoctoral Science Foundation (2018M640069), National Natural Science Foundation of China (61902024, 61732005), and National Key R&D Program of China (2018YFC0832104)

本文责任编辑 张民

Recommended by Associate Editor ZHANG Min

1. 北京理工大学计算机学院 北京 100081 2. 北京市海量语言信息处理与云计算应用工程技术研究中心 北京 100081 3. 北京理工大学外国语学院 北京 102488 4. 语言工程与认知计算工信部重点实验室 北京 102488

1. School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081 2. Beijing Engineering Research Center of High-Volume Language Information Processing and Cloud Computing Applications, Beijing 100081 3. School of Foreign Languages, Beijing Institute of Technology, Beijing 102488 4. Key Laboratory of Language, Cognition and Computation Ministry of Industry and Information Technology, Beijing 102488

器翻译领域的重要研究热点之一,也出现了很多值得关注的研究成果。

Google、Facebook、卡内基·梅隆大学(CMU)和爱丁堡大学等在低资源语言机器翻译上做了很多研究。国内机器翻译领域也非常重视这方面的研究。中科院自动化所、清华大学、苏州大学、东北大学、昆明理工大学、北京理工大学等多个科研团队在承担低资源语言机器翻译国家级科研项目、自主研发实用翻译系统等方面都积极推动深入的技术交流与合作,同时在全国机器翻译大会等多种学术活动都有广泛、密切的研讨,推动了这个方向的研究进展^[5-8]。

随着国内外研究的发展和深入,我们认为很有必要对目前稀缺资源语言机器翻译的研究进展进行比较全面的回顾。本文期望能够为机器翻译和相关领域的研究者提供有益的参考,帮助他们更好地深入了解低资源语言机器翻译的研究动态和选择未来的研究方向。

本综述的剩余部分组织如下:第1节介绍了与低资源语言机器翻译相关的学术活动和公开的数据资源;第2节详细梳理归纳了目前比较重要和常用的低资源翻译方法和技术,并总结了它们各自的特点;第3节总结了这些方法之间的关系,第4节分析了当前研究现状的主要特点;最后对未来的研究趋势和发展方向提出了展望和建议。

1 稀缺资源语言机器翻译学术活动及数据资源

1.1 相关学术活动

为了进一步推动稀缺资源语言机器翻译的发展,加快研究步伐,加强技术交流,实现技术与真实应用场景结合和技术落地,近两年以来国际上积极开展各类有影响力的学术活动。

机器翻译领域最权威的国际比赛之一的机器翻译大会(Conference on Machine Translation, WMT)近几年来在新闻领域的翻译评测任务都会涉及英语-低资源语言的翻译,如古吉拉特语,哈萨克语等^[9-12]。

“针对紧急事件的低资源语言”(Low Resource Languages for Emergent Incidents, LORELEI)是美国国防高级研究计划局(DARPA)资助的项目,该计划的目标是显著提高计算语言学和人类语言技术的水平,以实现低资源语言的快速和低成本开发。为此,美国国家标准技术研究所(NIST)推出了相应的评测活动(LoReHLT)¹,评测任务包括机器翻译、实体发现和链接等。该评测活动从2016开

¹ <https://www.nist.gov/itl/iad/mig/lorehlt-evaluations>

始每年一次,至2019年已经连续举办了4届。

WAT²(Workshop on Asian Translation)是专门针对亚洲语言翻译的会议,到2019年已经连续举办了6届。该会议具有亚洲低资源语言和英语的翻译评测活动。

另外两个专门的学术活动是“低资源翻译技术研讨会”(Workshop on Technologies for MT of Low Resource Languages, LoResMT)^{3,4}和“低资源自然语言处理与深度学习研讨会”(Deep Learning for Low-resource NLP)^{5,6},这两个论坛已于2018年和2019年分别连续举办了2届。

1.2 相关数据资源

机器翻译与数据资源密不可分。表1整理了一些可以用于低资源语言翻译的数据资源:

表1 低资源语言翻译相关的数据资源
Table 1 Data for low-resource MT

数据集	描述
WMT data	WMT 提供的英语-低资源语言的数据集。这也是目前研究中使用最多的数据集。
IWSLT data	面向口语翻译的 IWSLT 比赛也提供了一些低资源翻译数据集。
WAT data	WAT 提供亚洲低资源语言的翻译数据集。
LORELEI data ⁷	由 DARPA 开发的低资源单语-英语双语数据集。
JW300 ^[9]	该语料库涵盖了超过 300 种语言的双语数据。
WikiMatrix ^[14]	该语料库由 Facebook 开发构建,包含 85 种语言的维基百科平行语料。
FLORES ⁸	由 Facebook 开发的英语-尼泊尔语和僧伽罗语的双语数据集。
Indian Language Corpora Initiative (ILCI) corpus ^[15]	该语料库包括 11 种印度语言与英语的平行语料。
Asian Language Treebank ^[16]	该亚洲语言树库项目包括印尼语、老挝语等 9 种东南亚语言与英语的平行语料。

2 稀缺资源语言机器翻译方法

本部分将重点介绍低资源语言的机器翻译方法。由于 zero-shot (即待翻译语言对之间未经过翻译模型训练)和 zero-resource (即待翻译语言对之间没有平行语料数据)场景也属于低资源翻译的特

² <https://lotus.kuee.kyoto-u.ac.jp/WAT/>

³ <http://www.conference.amtaweb.org/>

⁴ <https://sites.google.com/view/loresmt/>

⁵ <https://sites.google.com/view/deepl018/home>

⁶ <https://sites.google.com/view/deepl019>

⁷ <https://www.darpa.mil/program/low-resource-languages-for-emergent-incidents>

⁸ <https://github.com/facebookresearch/flores>

殊形式, 故也将其纳入本文的考察范围之内. 目前已有研究方法大致可以分为五大类: 第一类是利用第三方枢轴语言的翻译方法; 第二类是从丰富资源语言到稀缺资源语言的迁移学习方法; 第三类是利用单语数据, 实现数据增强的方法; 第四类是半监督和无监督方法; 第五类是多语言和多任务的翻译方法. 接下来将分别介绍每类方法, 并在每类方法的最后总结各自的优势和不足等系列特点.

需要说明的是, 本文希望考察“低资源”这一特殊场景的翻译方法研究, 在整理文献时重点根据“低资源、零资源”等关键词语进行筛选, 同时也会考虑文献中的数据集规模, 如果某类方法出现在一些明显属于丰富资源语言或者较大实验数据集(比如规模多达几十万甚至百万)的文献中, 那么我们倾向不将这些文献纳入本文的研究范围. 另外, 在有些研究中, 某类方法可能会与其他类型的方法同时被使用. 另有一些研究中使用的方法也不一定能够严格地归入文中介绍的其中一类.

2.1 基于枢轴语言的方法

基于枢轴语言 (Pivot-based) 的翻译方法试图为源语言和目标语言寻找一种(或几种)枢轴语言(如英语), 实现源语言-枢轴语言-目标语言的翻译过程. 一种代表性的方法是首先利用源语言-枢轴语言翻译模型将源语言翻译为枢轴语言, 然后利用枢轴语言-目标语言翻译模型将枢轴语言翻译为目标语言^[17]. 这种方法在统计机器翻译中具有广泛的应用, NMT 兴起并发展以后, 也开始出现在 NMT 和一些商业翻译系统中. 例如 Google 在 GNMT 中很早就采用了枢轴语言翻译方法, 实验结果明显优于没有增量训练 (Incremental training, 也即, 使模型在不丢失已经学习到的已有数据信息的基础上, 继续学习新加入的训练数据) 的通用模型^[18].

基于枢轴语言的翻译方法通常会存在错误累积问题. 由于源语言-枢轴语言与枢轴语言-目标语言双语之间的关系不是非常紧密甚至毫无关系等原因, 源语言-枢轴语言翻译模型中的错误会传递到枢轴语言-目标语言模型中. 而且两种翻译模型通常单独训练, 这进一步放大了错误传递.

为了解决这一问题, 清华大学刘洋老师的团队做了深入而有影响的研究^[19-21]. Cheng 等^[19]提出了一种联合训练的神经网络翻译模型, 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-枢轴语言的翻译模型和枢轴语言-目标语言的翻译模型联系起来进行联合训练. 实验结果显示 BLEU 值比 Baseline (独立训练两种翻译模型) 提升了

1~2 个百分点.

同样是为了解决错误传递问题, Zheng 等^[20]针对零资源机器翻译, 通过最大期望似然估计 (Maximum expected likelihood estimation, MELE), 直接训练源语言-目标语言的翻译模型.

如图 1 所示, MELE 方法的目的是, 在枢轴语言 Z -目标语言 Y 平行语料的基础上, 针对源语言 X -目标语言 Y 翻译模型, 对枢轴语言 Z -源语言 X 翻译模型获得最大期望.

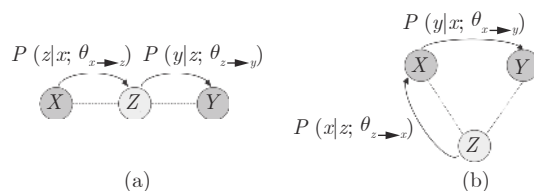


图 1 基于枢轴语言的方法 (a) 和 MELE 方法 (b)
Fig.1 Pivot-based method (a) and MELE method (b)

在 Cheng 和 Zheng 等的工作基础上, Chen 等^[21]针对零资源机器翻译进一步提出了“Teacher-Student”的框架. 该工作的主要思想是, 为了训练缺乏双语语料的源语言-目标语言的翻译模型(“学生”), 可以利用预训练的枢轴语言-目标语言的翻译模型(“老师”)指导基于源语言-枢轴语言双语数据训练的学生模型, 如图 2 所示.

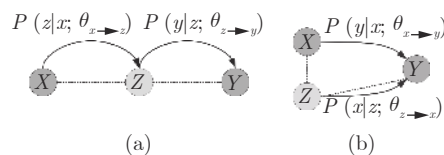


图 2 基于枢轴语言的方法 (a) 和“老师-学生”方法 (b)
Fig.2 Pivot-based method (a) and “Teacher-student” method (b)

该框架提出了句子级和词语级两种“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, 因此既有效率提升, 又可以避免错误传递. 与 Cheng 等的结果进行对比, 最好的 BLEU 分数在相同数据集上又提升了 3 个多百分点.

Ren 等^[22]假设源语言 X -枢轴语言 Z 属于丰富数据的语言对, 而枢轴语言 Z -目标语言 Y 属于稀缺语言对. 他们同样不针对 (X, Y) 进行直接建模, 但与前面研究不同的是, 他们将目标语言作为中间桥梁, 对目标语言-枢轴语言翻译模型进行建模, 把训练翻译模型 $P(Z|X)$ 分解为训练 $P(Y|X)$ 和 $P(Z|Y)$ 两个翻译模型, 然后利用期望最大化算法

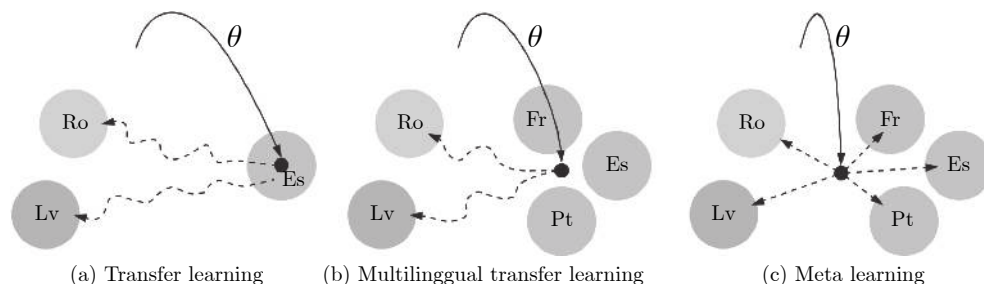


图 3 迁移学习, 多语言迁移学习与元学习

Fig. 3 Transfer learning, multilingual transfer learning and meta learning

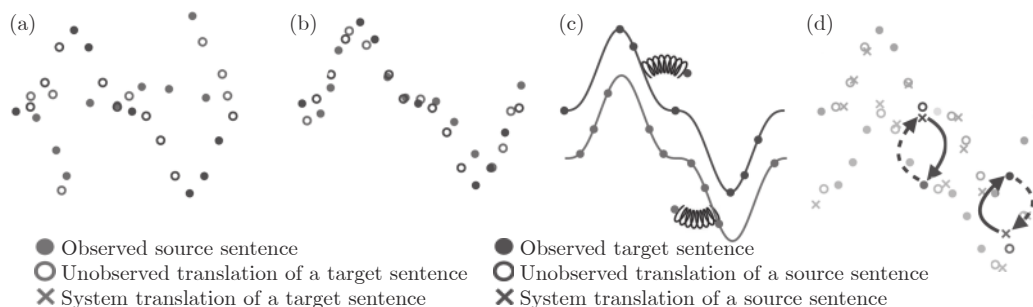


图 4 无监督翻译方法

Fig. 4 Unsupervised NMT

(Expectation-maximization, EM) 对模型进行训练.

Lakew 等^[23] 针对多语种的 zero-shot 翻译场景, 提出了一种以英语为枢轴语言的迭代式翻译模型, 在意大利语 - 罗马尼亚语数据上使 BLEU 分数提升了 8~10 个百分点.

除了文本信息, 图像等多模态信息也可以作为枢轴语言信息用于低资源语言翻译. 例如: Nakayama 和 Nishida^[24] 把图片作为枢轴信息, 认为对于改善零资源翻译效果同样有帮助. Chowdhury 等^[25] 利用 Flickr30k 图片数据集中的图像信息和图像描述信息实现了印地语-英语的多模态低资源语言翻译, 在图像信息的帮助下, BLEU 分数比单纯的文本翻译有了相应提升.

总结: 基于枢轴语言的翻译方法以第三方语言为中介, 以其简洁方便的特点能够适用于统计翻译和神经网络机器翻译等, 具有比较久的研究历史, 但该类方法由于不直接实现源语言和目标语言之间的翻译, 在训练过程的几个阶段容易产生错误累积等问题, 而且整个训练过程的解码时间相对较长.

2.2 迁移学习方法

迁移学习 (Transfer learning)^[26] 是机器学习的一种重要方法, 在自然语言处理领域同样有着广泛应用. Ruder 在其博士论文中有详细深入的论述^[27].

具体到低资源语言翻译, 迁移学习的一般做法是首先在数据资源丰富的语言上训练模型, 然后迁移到低资源语言上, 进行微调, 最终产生译文.

由于迁移学习涉及多种语言, 因此丰富资源和稀缺资源语言之间的语言相关性是一个经常研究的问题. 一般来说, 语言相似性越高, 越有利于迁移学习. 在“父”语言和“子”语言之间共享 BPE 得到的子词词汇表或者 BPE 向量表示可以利用语言之间的相似性.

Zoph 等^[28] 率先将迁移学习用在机器翻译中. 他们利用丰富资源的语言对训练了一个“父”模型, 然后对稀缺语言的“子”模型进行权重初始化, 最后在资源匮乏的语言对上训练最终的翻译模型.

Nguyen 和 Chiang^[29] 进一步探索了父模型的语言对也属于低资源语言, 但与子语言对之间有关系的场景, 通过使用 BPE 方法的共享词汇表和音译等改进了以前的方法. Dabre 等^[30] 也探索了语言相关性在迁移学习中对于翻译质量的影响效果.

Kocmi 和 Bojar^[31] 则验证了语言对之间不存在关联性的情景中, 利用 Transformer 在父、子模型之间共享词汇, 进行迁移学习翻译的效果. 模型以英语为中间语言, 分别在相关性较强的欧洲语言对之间和不相关的欧洲语言对之间进行了多组实验, 结果都有不同程度的明显提升. 但他们同时表明翻译效果可能还跟父语言对的规模有关.

Gu 等^[32]针对双语数据都非常稀缺的语言对, 提出了一种通用的翻译模型, 把多种源语言共享的词语级和句子级表示映射到一种目标语言中, 得到通用空间中的词向量表示. 这样做的优势是, 不同语言中语义相似的词语能够具有相似的表示并映射到同一个语义空间. 随后他们又提出了一种与模型无关的元学习框架(图 3)(Model-agnostic meta-learning, MAML)^[33], 在多个语种上的实验效果要远远优于迁移学习模型. Li 等^[34]也基于元学习方法利用多种领域的的数据解决低资源翻译中的领域迁移问题.

Kim 等^[35]利用跨语言词向量, 提出了一种不需要共享词汇表的迁移技术用于零资源翻译场景. 随后又提出了一种基于枢轴语言的迁移学习方法^[36], 首先预训练源语言 - 枢轴语言和枢轴语言 - 目标语言的翻译模型, 然后将二者迁移到源语言 - 目标语言翻译模型上进行微调. 苏州大学张民老师团队最近针对 zero-shot 场景提出了基于跨语言预训练模型的迁移学习方法, 比 Kim 等的效果进一步有所提升^[37].

整体来看, 目标语言端的迁移学习比源语言端的迁移学习更具挑战性. 因为不同的目标语言需要特定的目标语言表示形式, 而迁移学习则更喜欢目标语言不变的表示形式. 迁移学习的成功取决于在这些因素之间取得适当的平衡. 随着各种预训练模型的流行, 迁移学习能够将尽可能多的知识迁移到翻译模型, 在一定程度上也增加了翻译模型的可解释性.

总结: 以预训练模型为主要方式的迁移学习方法已经广泛应用于自然语言处理的很多应用领域和场景, 也极大地改善了翻译质量. 但随着各种超大规模的预训练模型的发布, 在一定程度上对于迁移学习涉及的语言之间的相似性和算力资源提出了更高的挑战, 在训练过程中也会存在不易优化等困难.

2.3 半监督和无监督方法

低资源语言机器翻译通常缺乏大规模双语数据, 而单语数据比较充足也更容易获取. 半监督方法(Semi-supervised)主要利用单语数据和小规模双语数据实现翻译过程. 其中回译技术(Back-translation, 或称反向翻译)是一种主要手段, 将在下一小节中详细论述.

Cheng 等^[38]利用自动编码器(Autoencoder)分别重构单语的源语言 X 和目标语言 Y . 即, 目标语言的自动编码器首先利用 $Y \rightarrow X$ 的翻译模型将目标语言编码为潜在的源语言, 然后利用 $X \rightarrow Y$ 翻译模型解码重构新的目标语言 Y' , 同理, 源语言亦

如此.

Skorokhodov 等^[39]则将预训练的语言模型与翻译模型结合起来, 首先分别训练源语言和目标语言的语言模型, 然后初始化翻译模型, 在俄语-英语这一极低资源数据集上使 BLEU 分数比 baseline 提高了 1.4 个百分点.

Gulcehre 等^[40]利用“浅层融合”(类似于常规 SMT 解码器中使用语言模型)和“深层融合”(将语言模型和解码器的隐藏状态拼接在一起)两种机制将单语目标语言的神经网络语言模型融合到翻译模型中, 在土耳其-英语等低资源数据集上使 BLEU 分数提升了接近 2 个百分点.

南京大学团队^[41]最近提出了一种“镜像生成式”(Mirror-generative)的机器翻译模型, 结合了回译、对偶学习等技术, 可以更好地利用非平行语料改善翻译效果.

在无监督翻译(Unsupervised)方法中, 则不依赖双语平行数据, 只依赖单语数据训练翻译模型(如图 4 所示). Facebook 的团队非常重视低资源翻译研究, 做了很多开创性的工作^[42-45]. 现有的无监督翻译方法通常包括两个阶段: 首先预训练语言模型^[46-47], 然后在微调阶段利用回译等训练翻译模型.

此后, 无监督方法吸引了更多研究者的关注. 在过去两年的 NLP 顶级会议上就有多篇关于无监督翻译方法的论文. 例如: Yang 等^[48]对无监督翻译模型增加了权重共享约束, 使得模型能够使用两个独立的编码器, 然后使用生成对抗网络(GAN)改善了翻译效果.

Gu 等^[49]针对零资源翻译场景, 首先定量分析了零资源翻译的退化问题. 然后通过预训练解码器的语言模型和对零资源语言对进行反向翻译, 在数十种欧洲语言对上验证了方法的有效性. 最近他们又利用 25 种语言预训练去噪的模型, 然后在不同规模的语言对数据集上调优, 进行了句子级别和文档级别的翻译实验, 在英语 - 越南语等低资源数据集上将 BLEU 提升了 10 多个百分点, 取得了非常好的效果^[50].

除了文献 [43] 等少数研究工作, 大部分无监督翻译方法通常首先需要利用无监督跨语言向量模型(如 Facebook 的 LASER 和 MUSE 工具等)^[51-52]将两种语言的向量表示映射到共享向量空间, 因此对于相似性较高的低资源语言之间的翻译比较有效, 而在远距离语言对之间的翻译效果比较差.

总结: 半监督和无监督翻译方法的一个优势在于可以更多地依赖于单语数据资源, 而不必局限于双语数据, 这在低资源翻译场景中具有很好的价值, 在未来仍将是非常有希望的研究方向之一. 但是这

类方法的实际效果在不同语言之间的差别较大,在相似性更高的语言对上具有更好的效果.尽管有研究在日语-俄语等远距离语言对上的无监督翻译提升了翻译效果^[53-54],但在大多数语言对上的效果仍然落后于传统的有监督方法.另外,有研究也指出,无监督方法的有效性依赖于大量辅助数据,以及其他条件是否满足^[55].例如,当语言在形态上不同或训练领域不匹配时,无监督方法的有效性会受到损害.

2.4 数据增强方法

数据增强方法(Data augmentation)最初广泛应用于计算机视觉领域,是图像处理的标准处理技术,其目的是利用有限的训练样本数据增加健壮性并改进学习目标.后来开始应用于机器翻译等自然语言处理领域.在低资源语言机器翻译中,由于缺少足够的双语数据,数据增强方法主要利用已有的单语数据达到增加训练数据的目的,从而更好地训练翻译模型,改善翻译效果.单语数据主要来自目标语言一端,但也有研究利用源语言端的单语数据改善翻译效果.例如:中科院自动化研究所的宗成庆研究员团队早在2016年就提出通过自学习算法和多任务模型框架利用源语言一侧的数据有效提升翻译质量,引起了广泛关注^[56].

Gibadullin等^[57]对低资源翻译中如何利用单语数据做了比较全面的综述分析,他们把利用单语数据的方法分为“独立于翻译模型”和“依赖于翻译模型”两大类.本文将常见的数据增强方法同样分为以下两种类型.

1) 回译方法

爱丁堡大学的团队首次将回译技术用于NMT,利用目标端的单语数据有效地提升了翻译性能^[58].此后,这一方法被广泛应用在NMT中,也被证明对于改善翻译质量特别是低资源翻译场景有很大帮助,回译已经成为NMT的标准技术之一.

尽管回译对NMT的性能提升有很大帮助,但其中仍有很多因素值得深入研究.很多工作针对回译从多个方面进行了探索.

Park等^[59]分析了仅使用反向翻译生成的多种合成数据对翻译模型的影响.Poncelas等^[60]也对多种训练数据类型(仅真实数据、仅合成数据、混合数据)以及回译数据的比例对于翻译效果的影响进行了实证分析.他们认为,与基于最大化推断生成的合成数据相比,基于采样和加入噪声的束搜索生成的合成数据能够提供更好的训练优势.后来他们又分析了SMT和NMT产生的合成数据对于NMT效果的影响,认为二者结合可以进一步提高翻译

效果^[61].

回译方法的一个主要局限是需要平衡真实数据和合成数据的规模.Edunov等^[62]对生成回译的多种方法进行了比较全面的实证研究和对比分析.他们表明,在合成数据中增加噪声数据不仅可以提高翻译质量,还可以使训练在合成句子与真实句子的比率较高的情况下更加可靠.Fadaee和Monz^[63]也进行了类似的工作.

回译方法虽然操作简单,方便有效,但产生的伪平行语料可能会面临数据质量较低的问题,在一定程度上会影响翻译性能.针对这个问题,有研究者做了相关工作.

Hoang等^[64]在回译的基础上进一步提出了迭代式回译,即不断重复回译的过程,直到获得更好的翻译效果.Imankulova等^[65-66]通过多次翻译筛选译文语句构建高质量的伪双语数据,以改善翻译质量.

Wu等^[67]提出了一种名为“Extract-Edit”的双语数据抽取方法用于替代广泛使用的回译方法,以产生高质量的双语数据.

Currey等^[68]则利用反向翻译将第三方的单语枢轴语言分别生成源语言和目标语言,组成伪双语数据,然后实现源语言到目标语言的翻译.

2) 词语替换方法

不同于回译方法,这种方法的主要做法是有针对性地替换训练数据中的词语,而无需训练回译需要的翻译模型.

Fadaee等^[69]在不改变训练语句句法和语义的前提下,将训练语料中的词语替换成一些低频词语.这样就增加了训练数据的规模.但该方法需要一些复杂的预处理步骤,而且实验表明只对低资源数据集有效.

Wang等^[70]对源语言和目标语言两侧均进行数据增强.利用从两侧词汇中统一采样的其他单词分别替换源语言句子和目标语言句子中的单词.在多个语言数据集上均有效提升了BLEU分数.

Xia等^[71]结合枢轴语言方法提出了一种实现数据增强的通用框架(图5),不仅可以利用目标语言一端的数据,还可以将与稀缺资源语言有一定联系的丰富资源语言作为枢轴语言,获得丰富语言-稀缺语言的数据.

微软团队^[72]提出了一种“软”语境数据增强方法.该方法可以将随机选定的词语替换为与该词语语境密切相关的“软”词语.软性词语可以通过语言模型获得,可以认为是基于语料数据的词表概率分布.

Zhou等^[73]则将句法信息引入数据增强方法,

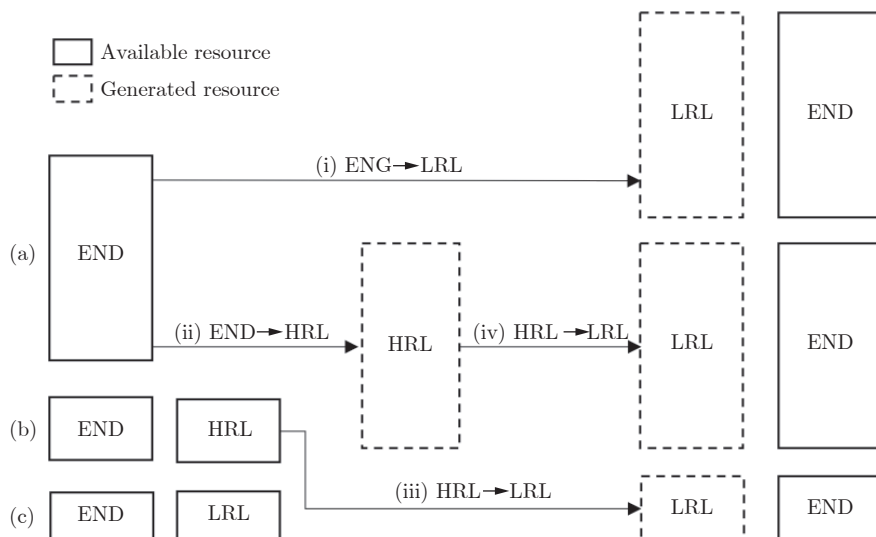


图5 数据增强框架. 其中, (i) 和 (ii) 是传统数据增强方法, (iii) 和 (iv) 是新提出的方法

Fig.5 Data augmentation method, where (i) and (ii) are traditional methods, while (iii) and (iv) are new ones

首先将目标语言调整为具有源语言句法结构和语序的目标语言, 然后利用双语词典将调序后的目标语言中的词语替换为源语言词语, 最后将其加入伪平行语料数据中, 从而实现数据增强的目的.

除了以上两种类型的方法, Currey 等^[74]通过复制目标端的数据也实现了提升翻译效果的目的.

数据增强方法在不同翻译任务上对于提升翻译效果存在差异性, 也就是说, 不是所有的数据增强方法在所有翻译任务上都能够提升效果. 为了分析数据增强在不同方法和任务中通常能够获得什么益处, 在深度学习理论的启发下, Li 等^[75]从输入灵敏度 (Input sensitivity) 和预测余量 (Prediction margin) 两个维度进行了全面的实证检验.

总结: 在缺乏大规模双语数据的场景中, 数据增强方法能够快速地扩充训练数据, 回译方法也以其自身的优势成为 NMT 的标准处理技术和数据增强的重要途径. 但这类方法通常需要考虑真实数据与新产生的合成数据之间的规模比例问题, 有时候当伪数据超过一定规模后, 反而会影响翻译性能, 另外, 通过数据增强产生的数据也可能存在质量较差的情况, 在一定程度上也会增加很多噪声.

2.5 多语言/多任务翻译方法

多语言 (Multi-lingual) 翻译的目标是通过单一翻译模型实现多种语言之间的互相翻译^[76]. 文献 [77] 对多语言神经网络机器翻译做了详细的回顾. 为了更好地改善低资源场景的翻译质量, 近几年来这种翻译方法引起了很多关注, 例如微软团队和 CMU 团队等都开展了有针对性的集中研究, 包括知识蒸

馏 (Knowledge distillation) 和语言集束方法^[78-79], 使用共享词向量表征所有语言, 以及软解耦的编码方法等^[80-81].

在多语言 NMT 模型中, 根据源语言和目标语言的数量, 通常存在三种翻译策略: 多对一翻译、一对多翻译以及多对多翻译. 多对一模型学习将在源语端的任何语言翻译成目标端的一种特定的语言, 这种情况下通常会选取语料丰富的语言比如英语; 相似地, 一对多模型学习将在源语端的一种语言翻译成目标端任意一种语言. 多对一这种模式通常会看作多领域学习的问题, 类似于源语端的输入分布是不一致的, 但是都会翻译成一种目标语. 而一对多模式可以被看作是多任务的问题, 类似于每一对源语到目标语看作一个单独的任务. 多对多这种模式就是这两个模式的集合. Google 的神经网络翻译系统在去年的最新进展中实现了以上三种策略在超过 100 种语言之间的低资源和零资源翻译^[82].

Firat 等^[83]对零资源翻译场景提出了一种多语言翻译的调优算法, 认为多对一的翻译策略更优于一对一的翻译方法. Zhou 等^[84]对多个不同语族的多种欧洲语言进行了全面的定性和定量分析, 得出了基于与低资源目标语言最相近的语族的几种语言训练多语言翻译模型, 能够有效改善翻译质量的结论. Maimaiti 等^[85]同样得到了类似的结论, 与仅使用一对高资源语言对进行迁移学习相比, 使用多个语言关系高度相关的高资源语言对并进行多个回合的微调可以提高翻译性能. 类似工作还有文献 [86] 等. 从这些研究可以看出, 无论是多语言翻译, 还是迁移学习方法, 更强的语言相关性对于改善翻译效

果都会有更多帮助。

除了欧洲语言,也有人研究了亚洲语言。Dabre 等^[87]将多语言翻译方法与迁移学习结合,实现了英语到亚洲多种语言的翻译。他们在多语言、多阶段微调过程中没有关注语言差异,而是表明数据大小很重要。但这种多语言、多阶段的迁移学习需要进一步的研究。Murthy 等^[88]基于 Zoph 等^[28]的工作,使用英语作为辅助语言,以英语-印地语作为预训练的父模型,在预训练时还考虑了句法语序信息,最后在几种亚洲语言-印地语的子翻译模型上进行微调。

Imankulova 等^[89]将领域适用性、多语言翻译和反向翻译等技术结合起来,首先利用领域外的数据训练多语言翻译模型,然后在领域内数据集上微调,在日语-俄语这一语言特点差异很大的极端低资源数据集上使翻译效果有了明显提升。

Neubig 和 Hu^[90]将多语言作为种子模型,提出了一种能够快速有效地适用于新语言的多语言翻译模型。类似地, Lu 等^[91]在多语言翻译框架中引入国际辅助语言“因特语”(Interlingua)模块,在 zero-shot 场景中将特定语言的编码器输出转化为独立于特定语言的解码器输入表示,实现了语言的直接翻译,而无需借助枢轴语言。

Sestorain 等^[92]则利用强化学习方法共同训练融入语言模型的多语言翻译模型和重构目标,以保证译文的语法准确性和翻译质量。但这种方法在训练中非常耗时,而且在数次迭代后效果很难继续有所改善。

多语言翻译对于低资源翻译场景尽管具有很多优势,但同时面临着在独立模型下无法实现知识共享,且需要大规模存储和计算资源的问题。CMU 的团队为翻译模型设计了一种适用于多种语言的语境参数生成器,使翻译模型变得更加通用,并且适用于零资源和低资源翻译^[81]。中科院自动化研究所的宗成庆老师团队提出了一种结构紧凑且语言敏感的多语言机器翻译方法,能够很好地利用语言之间的共性,为低资源和零资源的翻译场景提供了新的解决思路^[93]。

多任务 (Multi-task) 翻译方法是 NMT 的一种常见方法之一,其主要思想是利用多个相关的其他任务(如句法分析、语义分析等),来提升翻译任务的质量。

文献 [94] 利用词性标注和依存分析任务将句法信息引入翻译任务中,使得 BLEU 分数有了 1~2 个百分点的提升;文献 [95-96] 改进模型架构设计,希望学习到多个任务之间共享的有效参数,同时改进训练安排,即调整不同任务之间的优先程度,保

证将改善翻译质量作为主任务,其他相关任务作为辅助任务。

总结:多语言翻译方法对于低资源翻译任务是很有必要的,因为来自多种语言的翻译模型能够帮助资源匮乏语言获得额外的知识,同时不同语言也使得模型具有更好的泛化能力,与双语翻译相比具有更好的迁移学习能力^[97]。目前主流的多语言翻译模型主要有三种思路:1) 使用不同的编码器和解码器表征不同语言;2) 使用统一的编码器和解码器用于所有语言之间的翻译;3) 在编码器和解码器中共享一部分参数表示语言之间的共性特征,另一部分表示语言的特有属性。但正如前面提到的,通常都会不可避免地面临模型结构复杂,计算复杂度增加以及知识共享存在障碍等问题。

以上回顾了低资源翻译的主要方法。另外,还有一些研究不一定能够严格地归入上述的某一类方法,如微软团队提出的对偶学习 (Dual learning)^[98]就与前面方法的学习范式有很多不同之处,自从提出以来产生了比较大的影响,以及最近提出的语言图蒸馏 (Language graph distillation) 方法^[99]等。再如文献 [100] 等使用的方法同样证明了对于提升翻译质量的有效性。此处不再赘述。

3 翻译方法之间的关系

从前面的梳理可以看到,很多研究文献中结合了不止一种翻译方法和技术,表 2 列出了一些这样的文献。

表 2 使用多种翻译方法的一些文献
Table 2 Literatures with more than one MT method

文献	使用的方法
[84-87]	多语言, 迁移学习
[89]	多语言, 反向翻译, 领域迁移
[18], [23], [49]	多语言, 枢轴语言方法
[102]	多语言, 无监督方法
[41-45], [58]	反向翻译, 半监督方法
[68], [71]	数据增强, 枢轴语言方法
[56]	数据增强, 多任务方法
[39]	迁移学习, 半监督方法
[36]	迁移学习, 枢轴语言方法

研究中使用的不同方法各有特点,也存在比较密切的联系,下面根据图 6 依次进行简要分析。

迁移学习是目前非常流行而又高效的方法,特别是最近两年,在稀缺资源翻译研究中占有较高的比重,这种“预训练+微调”的模式可以应用到其他各类的翻译方法中。因此与其他各类方法都有

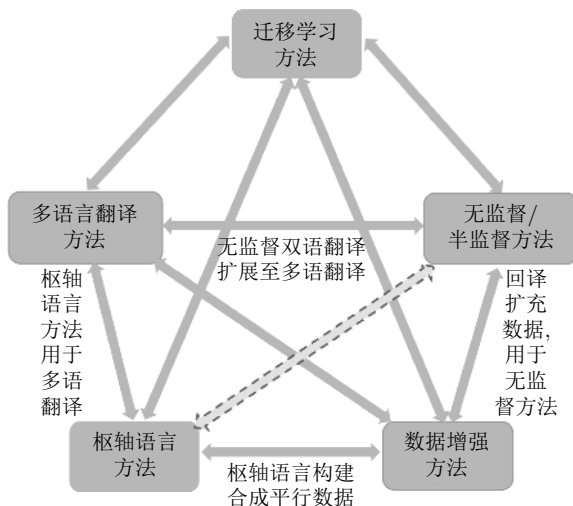


图 6 几类翻译方法之间的关系

Fig. 6 Relations between the translation methods

关联。

枢轴语言翻译方法由来已久, 在用于 NMT 以前, 经常用于统计翻译, 甚至在更早的规则翻译方法中也有所体现. 枢轴语言方法从一个语言对之间的翻译还可以扩展到多语种之间的翻译, 比如多语种之间的零资源翻译场景. 数据增强方法同样可以扩展到多语种翻译场景^[101].

另外, 枢轴语言还可以通过反向翻译构造枢轴语言与源语言/目标语言之间的合成双语数据, 从而实现了数据增强的目的。

值得注意的是, 在图 6 中, 枢轴语言翻译方法与无监督/半监督方法之间采用虚线箭头连接, 这是因为多数场景 (如 zero-shot) 的枢轴语言翻译可以看做是有监督翻译或半监督翻译, 而不属于无监督方法。

无监督方法具有很大的挑战性. 由于不使用双语数据, 而摆脱了 NMT 严重依赖大规模双语数据的局限, 对数据资源的依赖较小, 尤其适合于低资源和零资源翻译场景. 尽管目前的翻译效果仍然相对不是非常理想, 但在近期和未来一段时间应该会有很好的发展潜力。

将无监督翻译方法用于多语言翻译场景的研究目前暂时不多, 因为多语言翻译通常属于有监督翻译方法. 不过最近 Google 团队尝试了无监督方法的多语种翻译, 在罗马尼亚语-英语数据集上使最好 BLEU 分数提高了近两个百分点^[102]. 相信未来会有更多这方面的尝试。

由于很多文献在不同领域、不同规模和不同语种的测试集上进行实验, 我们认为难以比较以上不同类别方法之间的效果差异. 但我们分析了过去三

年 (2017~2019 年) WMT 比赛的新闻领域翻译评测任务使用的方法, 希望从实际比赛和工程实践中观察各类翻译方法的使用情况。

根据历年的评测总结报告, 在 2017 年的评测参赛队伍中, 约有一半左右 (12~15 个) 使用了回译技术, 而迁移学习和无监督方法等在当时还未出现或者尚不流行, 几乎没有被用于参赛. 在 2018 年的比赛中, 随着 Transformer 的出现和流行, 35 个左右的参赛队伍中的绝大多数 (约 30 个) 都使用了 Transformer 架构, 而其中都使用了回译作为基本的技术手段, 另外还有三四个队伍分别使用了无监督方法和多语言结合的方法. 2019 年的比赛任务吸引了包括多家国内单位在内的更多参赛队伍 (近 50 个) 参加, 有效提交系统数量超过 150 个, 同时新增了无监督翻译任务, 比赛中使用的方法也更加多元化. 表 3 总结了本文介绍的几类方法在参赛系统中出现的大概频次。

表 3 几类方法在 WMT2019 中的使用情况
Table 3 The methods in WMT2019

方法	频次
回译方法	45
多次回译方法	19
迁移学习和微调	24
使用额外语言 (包括枢轴语言和多语种)	12
无监督方法	9

除了以上方法, 微软亚洲研究院、百度、小牛翻译等多家单位还使用了领域适用性 (Domain adaptation) 和知识蒸馏等多种有效的综合方法和技术。

图 7 是 WMT2019 提供的一些训练方法的统计和占比, 从中也可以看出包括低资源翻译在内的神经网络机器翻译在国际比赛中的技术概况。

4 当前研究现状的特点

图 8 简要总结了这几类翻译方法各自的优势与局限性。

通过梳理已有的研究方法, 我们还总结了目前研究现状呈现的主要特点:

1) 从目前工作的类型来看, 已有研究主要分为两种类型: 一种主要的类型是采用某种或者某些方法改善低资源翻译的质量; 另一种则是对某种方法 (如回译) 进行不同程度的定性和定量研究和分析, 希望对这些方法有更深入的了解等. 这类工作不一定涉及方法改进, 往往只是对于某个方法的实证性探究。

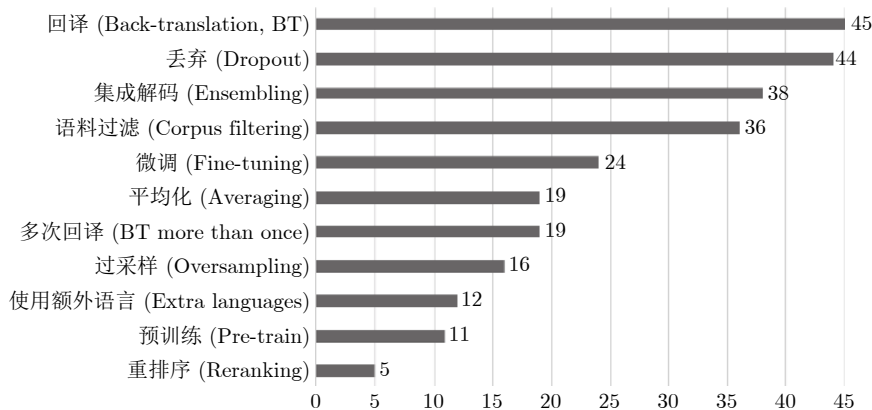


图 7 WMT2019 中涉及的主要方法和技术

Fig.7 Main methods in WMT2019



图 8 各类翻译方法的优势与局限

Fig.8 Advantages and limits of translation methods

2) 从机器翻译的类型来看, 绝大多数研究集中于低资源文本翻译, 这也是机器翻译最重要的应用场景之一; 还有一些工作围绕语音翻译开展研究, 包括语音转录 (Speech transcription)^[103], 语音到文本的翻译等^[104-106]; 甚至还有人研究低资源方言的机器翻译等^[107-108].

3) 从机器翻译涉及的领域来看, 更多的研究通常更关注新闻、(Technology, Entertainment, Design) 演讲等常见的通用领域的低资源语言翻译, 一个主要原因在于这些领域的的数据资源相对较多. 文献 [109] 针对医疗领域中的医患交流口语翻译率先做了尝试. 低资源翻译从常用领域逐渐扩展到更多的专门领域和场景很有必要, 同时也更有挑战性.

4) 从研究涉及的语种分布来看, 低资源语种的分布很不均衡. 绝大部分的工作主要集中于欧洲语言场景, 英语毫无疑问是最为广泛的语言, 无论是作为枢轴语言, 还是与其他低资源语种之间的翻译; 另有很少一部分研究以亚洲等其他地区的语种为研究对象, 如日语-越南语等^[110].

值得一提的是, 就我们目前掌握的研究文献,

汉语与其他低资源语言之间的 NMT 研究数量仍然非常少, 仅有少量研究采用统计翻译方法. 以汉语-越南语神经网络翻译为例, 昆明理工大学的团队近年来做了比较集中的研究^[111-115].

5 研究趋势展望

稀缺资源语言机器翻译目前仍然具有很多挑战和困难有待解决, 例如以下几个科学问题:

预训练模型与翻译模型的结合问题; 不同语言, 特别是远距离语言之间的有效表示学习问题; 以及领域适用性等问题.

我们认为低资源语言机器翻译在未来可能会呈现如下研究趋势和发展方向:

1) 加强预训练模型和迁移学习在低资源翻译中的研究. 预训练模型和迁移学习在 NLP 和机器翻译领域已被证明其有效性, 并成为一种新的研究范式. 一些研究工作已经开始关注将 BERT (Bidirectional encoder representations from transformers)^[116] 等预训练模型融入低资源语言的翻译模型中^[117-118], 未来将有更多值得探索和研究的方

例如: 如何更好地将 BERT/GPT (Generative pre-trained transformer)^[119] 等流行的预训练模型融入到翻译模型中, 并将迁移学习的功效尽可能最大化等. 目前基于大规模单语数据的无监督预训练模型进行迁移学习和无监督翻译等研究正在增强.

2) 加强语言之间的相关性等方面的语言学分析研究. 整体来看, 低资源翻译的研究多以改进算法模型提高翻译质量为主, 而缺乏必要的语言学分析. 目前虽然有一些研究从语言学角度出发, 证明了语族接近的语言有助于改善低资源语言的翻译效果等, 但值得进一步深入探索. 如果能够发现不同语言之间更多的深层次的语言学特征, 实现语言特征和知识的迁移和传递, 相信会进一步促进稀缺资源语言翻译的发展, 同时也可能使翻译过程更具有解释性.

3) 加强更有效的语言表示学习研究. 低资源语言对之间在词语、句法和语义等层面往往存在较大差异, 甚至具有不同的书写系统以及属于不同的语族. 在训练翻译模型时, 如何处理并减小语言之间在编码表示中的差异, 更好地平衡不同语言在向量空间中的表示始终是影响低资源翻译的核心问题. 例如: 构建语言无关的编码器和具有语言意识的解码器等都将是值得深入研究的话题.

4) 加强汉语和低资源语言之间的机器翻译研究. 我国目前正在大力推行“一带一路”倡议, 构建人类命运共同体. 语言互通是实现“一带一路”建设的重要保障, 而机器翻译又是实现语言互通的加速器和催化剂. 在未来应该进一步加强汉语与低资源语言, 特别是“一带一路”沿线国家和地区语言之间的机器翻译研究, 包括构建多样化的语言数据资源, 利用不同方法全面提升翻译质量等. 只有这样, 才能更好地服务于国家的重大战略需求.

5) 加强口语和方言等低资源语言的语音翻译研究. 正如上文所述, 目前大部分的低资源翻译主要关注文本翻译. 语音翻译, 包括自动语音识别, 语音-文本转换以及语音-语音翻译等都具有很多应用场景, 也存在很多机遇和挑战. 未来关于低资源语言在语音上的翻译也将是一个很有趣和有价值的研究方向.

References

- 1 Kalchbrenner N, Blunsom P. Recurrent continuous translation models. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Seattle, Washington, USA: Association for Computational Linguistics, 2013. 1700-1709
- 2 Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks. In: Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press, 2014. 3104-3112
- 3 Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. In: Proceedings of the 3rd International Conference on Learning Representations. San Diego, USA, 2015.
- 4 Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, et al. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc., 2017. 6000-6010
- 5 Liu Yang. Recent advances in neural machine translation. *Journal of Computer Research and Development*, 2017, **54**(6): 1144-1149
(刘洋. 神经机器翻译前沿进展. 计算机研究与发展, 2017, **54**(6): 1144-1149)
- 6 Li Ya-Chao, Xiong De-Yi, Zhang Min. A survey of neural machine translation. *Chinese Journal of Computers*, 2018, **41**(12): 2734-2755
(李亚超, 熊德意, 张民. 神经机器翻译综述. 计算机学报, 2018, **41**(12): 2734-2755)
- 7 Lin Qian, Liu Qing, Su Jin-Song, Lin Huan, Yang Jing, Luo Bin. Focuses and frontiers tendency in neural machine translation research. *Journal of Chinese Information Processing*, 2019, **33**(11): 1-14
(林倩, 刘庆, 苏劲松, 林欢, 杨静, 罗斌. 神经网络机器翻译研究热点与前沿趋势分析. 中文信息学报, 2019, **33**(11): 1-14)
- 8 Zhao Yang, Zhou Long, Wang Qian, Ma Cong, Liu Yu-Chen, Wang Yi-Ning, et al. The study on ethnic-to-Chinese scarce-resource neural machine translation. *Journal of Jiangxi Normal University (Natural Sciences Edition)*, 2019, **43**(6): 630-637
(赵阳, 周龙, 王迁, 马聪, 刘宇宸, 王亦宁, 等. 民汉稀缺资源神经机器翻译技术研究. 江西师范大学学报(自然科学版), 2019, **43**(6): 630-637)
- 9 Bojar O, Chatterjee R, Federmann C, Graham Y, Haddow B, Huck M, et al. Findings of the 2016 conference on machine translation. In: Proceedings of the 1st Conference on Machine Translation: Volume 2, Shared Task Papers. Berlin, Germany: Association for Computational Linguistics, 2016. 131-198
- 10 Bojar O, Chatterjee R, Federmann C, Graham Y, Haddow B, Huang S J, et al. Findings of the 2017 conference on machine translation (WMT17). In: Proceedings of the 2nd Conference on Machine Translation. Copenhagen, Denmark: Association for Computational Linguistics, 2017. 169-214
- 11 Bojar O, Federmann C, Fishel M, Graham Y, Haddow B, Koehn P, et al. Findings of the 2018 conference on machine translation (WMT18). In: Proceedings of the 3rd Conference on Machine Translation: Shared Task Papers. Brussels, Belgium: Association for Computational Linguistics, 2018. 272-303
- 12 Barrault L, Bojar O, Costa-Jussà M, Federmann C, Fishel M, Graham Y, et al. Findings of the 2019 conference on machine translation (WMT19). In: Proceedings of the 4th Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1). Florence, Italy: Association for Computational Linguistics, 2019. 1-61
- 13 Agić Ž, Vulić I. JW300: A wide-coverage parallel corpus for low-resource languages. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019. 3204-3210
- 14 Schwenk H, Chaudhary V, Sun S, Gong H Y, Guzmán F. WikiMatrix: Mining 135m parallel sentences in 1620 language pairs from Wikipedia. arXiv preprint arXiv: 1907.05791, 2019
- 15 Jha G N. The TDIL program and the Indian language corpora initiative (ILCI). In: Proceedings of the 7th International Conference on Language Resources and Evaluation. Valletta, Malta, 2010. 982-985
- 16 Thu Y K, Pa W P, Utiyama M, Finch A, Sumita E. Introducing the Asian language Treebank (ALT). In: Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC' 16). Portorož, Slovenia: European Language Resources Association (ELRA), 2016. 1574-1578
- 17 Ahmadnia B, Serrano J, Haffari G. Persian-Spanish low-resource statistical machine translation through English as pivot language. In: Proceedings of the 2017 Recent Advances in Nat-

- ural Language Processing. Varna, Bulgaria, 2017. 24–30
- 18 Johnson M, Schuster M, Le Q V, Krikun M, Wu Y H, Chen Z F, et al. Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 2017, 5: 339–351
 - 19 Cheng Y, Yang Q, Liu Y, Sun M S, Xu W. Joint training for pivot-based neural machine translation. In: Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI-17). Melbourne, Australia, 2017. 3974–3980
 - 20 Zheng H, Cheng Y, Liu Y. Maximum expected likelihood estimation for zero-resource neural machine translation. In: Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne, Australia: AAAI Press, 2017. 4251–4257
 - 21 Chen Y, Liu Y, Cheng Y, Li V O K. A teacher-student framework for zero-resource neural machine translation. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Vancouver, Canada: Association for Computational Linguistics, 2017. 1925–1935
 - 22 Ren S, Chen W H, Liu S J, Li M, Zhou M, Ma S. Triangular architecture for rare language translation. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne, Australia: Association for Computational Linguistics, 2018. 56–65
 - 23 Lakew S M, Lotito Q F, Negri M, Turchi M, Federico M. Improving zero-shot translation of low-resource languages. In: Proceedings of the 14th International Workshop on Spoken Language Translation. Tokyo, Japan, 2017. 113–119
 - 24 Nakayama H, Nishida N. Zero-resource machine translation by multimodal encoder-decoder network with multimedia pivot. *Machine Translation*, 2017, 31(1): 49–64
 - 25 Chowdhury K D, Hasanuzzaman M, Liu Q. Multimodal neural machine translation for low-resource language pairs using synthetic data. In: Proceedings of the Workshop on Deep Learning Approaches for Low-Resource NLP. Melbourne, Australia: Association for Computational Linguistics, 2018. 33–42
 - 26 Pan S J, Yang Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 2010, 22(10): 1345–1359
 - 27 Ruder S. Neural Transfer Learning for Natural Language Processing [Ph. D. dissertation], National University of Ireland, Ireland, 2019
 - 28 Zoph B, Yuret D, May J, Knight K. Transfer learning for low-resource neural machine translation. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, Texas: Association for Computational Linguistics, 2016. 1568–1575
 - 29 Nguyen T Q, Chiang D. Transfer learning across low-resource, related languages for neural machine translation. In: Proceedings of the 8th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). Taipei, China: Asian Federation of Natural Language Processing, 2017. 296–301
 - 30 Dabre R, Nakagawa T, Kazawa H. An empirical study of language relatedness for transfer learning in neural machine translation. In: Proceedings of the 31st Pacific Asia Conference on Language, Information and Computation. Cebu City, Philippines: The National University, 2017. 282–286
 - 31 Kocmi T, Bojar O. Trivial transfer learning for low-resource neural machine translation. In: Proceedings of the 3rd Conference on Machine Translation: Research Papers. Brussels, Belgium: Association for Computational Linguistics, 2018. 244–252
 - 32 Gu J T, Hassan H, Devlin J, Li V O K. Universal neural machine translation for extremely low resource languages. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). New Orleans, Louisiana: Association for Computational Linguistics, 2018. 344–354
 - 33 Gu J T, Wang Y, Chen Y, Li V O K, Cho K. Meta-learning for low-resource neural machine translation. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium: Association for Computational Linguistics, 2018. 3622–3631
 - 34 Li R M, Wang X, Yu H. MetaMT, a meta learning method leveraging multiple domain data for low resource machine translation. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York, USA: AAAI, 2020. 8245–8252
 - 35 Kim Y, Gao Y B, Ney H. Effective cross-lingual transfer of neural machine translation models without shared vocabularies. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019. 1246–1257
 - 36 Kim Y, Petrov P, Petrushkov P, Khadivi S, Ney H. Pivot-based transfer learning for neural machine translation between non-English languages. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China: Association for Computational Linguistics, 2019. 866–876
 - 37 Ji B J, Zhang Z R, Duan X Y, Zhang M, Chen B X, Luo W H. Cross-lingual pre-training based transfer for zero-shot neural machine translation. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI-20). New York, USA: AAAI, 2020. 115–122
 - 38 Cheng Y, Xu W, He Z J, He W, Wu H, Sun M S, et al. Semi-supervised learning for neural machine translation. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Berlin, Germany: Association for Computational Linguistics, 2016. 1965–1974
 - 39 Skorokhodov I, Rykachevskiy A, Emelyanenko D, Slotin S, Ponkratov A. Semi-supervised neural machine translation with language models. In: Proceedings of the AMTA 2018 Workshop on Technologies for MT of Low Resource Languages (LoResMT 2018). Boston, MA, USA: Association for Machine Translation in the Americas, 2018. 37–44
 - 40 Gulcehre C, Firat O, Xu K, Cho K, Bengio Y. On integrating a language model into neural machine translation. *Computer Speech and Language*, 2017, 45: 137–148
 - 41 Zheng Z X, Zhou H, Huang S J, Li L, Dai X Y, Chen J J. Mirror-generative neural machine translation. In: Proceedings of the 8th International Conference on Learning Representations. Addis Ababa, Ethiopia, 2020.
 - 42 Lample G, Conneau A, Denoyer L, Ranzato M A. Unsupervised machine translation using monolingual corpora only. In: Proceedings of the 6th International Conference on Learning Representations. Vancouver, Canada, 2018.
 - 43 Lample G, Ott M, Conneau A, Denoyer L, Ranzato M A. Phrase-based & neural unsupervised machine translation. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium: Association for Computational Linguistics, 2018. 5039–5049
 - 44 Artetxe M, Labaka G, Agirre E, Cho K. Unsupervised neural machine translation. In: Proceedings of the 6th International Conference on Learning Representations. Vancouver, Canada, 2018.
 - 45 Artetxe M, Labaka G, Agirre E. An effective approach to unsupervised machine translation. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019. 194–203
 - 46 Artetxe M, Labaka G, Agirre E. A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne, Australia: Association for Computational Linguistics, 2018. 789–798
 - 47 Artetxe M, Labaka G, Agirre E. Generalizing and improving bilingual word embedding mappings with a multi-step framework of linear transformations. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans, USA: AAAI, 2018. 5012–5019

- 48 Yang Z, Chen W, Wang F, Xu B. Unsupervised neural machine translation with weight sharing. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne, Australia: Association for Computational Linguistics, 2018. 46–55
- 49 Gu J T, Wang Y, Cho K, Li V O K. Improved zero-shot neural machine translation via ignoring spurious correlations. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019. 1258–1268
- 50 Liu Y H, Gu J T, Goyal N, Li X, Edunov S, Ghazvininejad M, et al. Multilingual denoising pre-training for neural machine translation. arXiv preprint arXiv: 2001.08210, 2020
- 51 Artetxe M, Labaka G, Agirre E. 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). Vancouver, Canada: Association for Computational Linguistics, 2017. 451–462
- 52 Conneau A, Lample G. Cross-lingual language model pretraining. In: Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Vancouver, Canada, 2019.
- 53 Pourdamghani N, Aldarrab N, Ghazvininejad M, Knight K, May J. Translating translationese: A two-step approach to unsupervised machine translation. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019. 3057–3062
- 54 Leng Y C, Tan X, Qin T, Li X Y, Liu T Y. Unsupervised pivot translation for distant languages. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019. 175–183
- 55 Sennrich R, Zhang B. Revisiting low-resource neural machine translation: A case study. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019. 211–221
- 56 Zhang J J, Zong C Q. Exploiting source-side monolingual data in neural machine translation. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, Texas: Association for Computational Linguistics, 2016. 1535–1545
- 57 Gibadullin I, Valeev A, Khushainova A, Khan A. A survey of methods to leverage monolingual data in low-resource neural machine translation. In: Proceedings of the 2019 International Conference on Advanced Technologies and Humanitarian Sciences. Rabat, Morocco, 2019.
- 58 Sennrich R, Haddow B, Birch A. Improving neural machine translation models with monolingual data. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Berlin, Germany: Association for Computational Linguistics, 2016. 86–96
- 59 Park J, Na B, Yoon S. Building a neural machine translation system using only synthetic parallel data. arXiv preprint arXiv: 1704.00253, 2017
- 60 Poncelas A, Shterionov D, Way A, de Buy Wenniger G M, Passban P. Investigating backtranslation in neural machine translation. In: Proceedings of the 21st Annual Conference of the European Association for Machine Translation. Alacant, Spain, 2018. 249–258
- 61 Poncelas A, Popović M, Shterionov D, de Buy Wenniger G M, Way A. Combining SMT and NMT back-translated data for efficient NMT. In: Proceedings of the 2019 Recent Advances in Natural Language Processing. Varna, Bulgaria, 2019. 922–931
- 62 Edunov S, Ott M, Auli M, Grangier D. Understanding back-translation at scale. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium: Association for Computational Linguistics, 2018. 489–500
- 63 Fadaee M, Monz C. Back-translation sampling by targeting difficult words in neural machine translation. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium: Association for Computational Linguistics, 2018. 436–446
- 64 Hoang V C D, Koehn P, Haffari G, Cohn T. Iterative back-translation for neural machine translation. In: Proceedings of the 2nd Workshop on Neural Machine Translation and Generation. Melbourne, Australia: Association for Computational Linguistics, 2018. 18–24
- 65 Imankulova A, Dabre R, Fujita A, Imamura K. Exploiting out-of-domain parallel data through multilingual transfer learning for low-resource neural machine translation. In: Proceedings of Machine Translation Summit XVII Volume 1: Research Track. Dublin, Ireland: European Association for Machine Translation, 2019. 128–139
- 66 Imankulova A, Sato T, Komachi M. Improving low-resource neural machine translation with filtered pseudo-parallel corpus. In: Proceedings of the 4th Workshop on Asian Translation (WAT2017). Taipei, China: Asian Federation of Natural Language Processing, 2017. 70–78
- 67 Wu J W, Wang X, Wang W Y. Extract and edit: An alternative to back-translation for unsupervised neural machine translation. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1(Long and Short Papers). Minneapolis, Minnesota: Association for Computational Linguistics, 2019. 1173–1183
- 68 Currey A, Heafield K. Zero-resource neural machine translation with monolingual pivot data. In: Proceedings of the 3rd Workshop on Neural Generation and Translation. Hong Kong, China: Association for Computational Linguistics, 2019. 99–107
- 69 Fadaee M, Bisazza A, Monz C. Data augmentation for low-resource neural machine translation. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Vancouver, Canada: Association for Computational Linguistics, 2017. 567–573
- 70 Wang X Y, Pham H, Dai Z H, Neubig G. SwitchOut: An efficient data augmentation algorithm for neural machine translation. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium: Association for Computational Linguistics, 2018. 856–861
- 71 Xia M Z, Kong X, Anastasopoulos A, Neubig G. Generalized data augmentation for low-resource translation. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019. 5786–5796
- 72 Gao F, Zhu J H, Wu L J, Xia Y C, Qin T, Cheng X Q, et al. Soft contextual data augmentation for neural machine translation. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019. 5539–5544
- 73 Zhou C T, Ma X Z, Hu J J, Neubig G. Handling syntactic divergence in low-resource machine translation. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China: Association for Computational Linguistics, 2019. 1388–1394
- 74 Currey A, Barone A V M, Heafield K. Copied monolingual data improves low-resource neural machine translation. In: Proceedings of the 2nd Conference on Machine Translation. Copenhagen, Denmark: Association for Computational Linguistics, 2017. 148–156
- 75 Li G L, Liu L M, Huang G P, Zhu C H, Zhao T J. Understanding data augmentation in neural machine translation: Two perspectives towards generalization. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China: Association for Computational Linguistics, 2019. 5689–5695
- 76 Firat O, Cho K, Bengio Y. Multi-way, multilingual neural ma-

- chine translation with a shared attention mechanism. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego, California: Association for Computational Linguistics, 2016. 866–875
- 77 Dabre R, Chu C H, Kunchukuttan A. A survey of multilingual neural machine translation. *ACM Computing Surveys*, to be published
- 78 Tan X, Ren Y, He D, Qin T, Zhao Z, Liu T Y. Multilingual neural machine translation with knowledge distillation. In: Proceedings of the 7th International Conference on Learning Representations. New Orleans, USA, 2019.
- 79 Tan X, Chen J L, He D, Xia Y C, Qin T, Liu T Y. Multilingual Neural Machine Translation with Language Clustering. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China: Association for Computational Linguistics, 2019. 963–973
- 80 Wang X Y, Pham H, Arthur P, Neubig G. Multilingual neural machine translation with soft decoupled encoding. In: Proceedings of the 7th International Conference on Learning Representations. New Orleans, USA, 2019.
- 81 Platanios E A, Sachan M, Neubig G, Mitchell T. Contextual parameter generation for universal neural machine translation. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium: Association for Computational Linguistics, 2018. 425–435
- 82 Arivazhagan N, Bapna A, Firat O, Lepikhin D, Johnson M, Krikun M, et al. Massively multilingual neural machine translation in the wild: Findings and challenges. arXiv preprint arXiv: 1907.05019, 2019
- 83 Firat O, Sankaran B, Al-Onaizan Y, Vural F T Y, Cho K. Zero-resource translation with multi-lingual neural machine translation. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, Texas: Association for Computational Linguistics, 2016. 268–277
- 84 Zhou Z, Sperber M, Waibel A. Massively parallel cross-lingual learning in low-resource target language translation. In: Proceedings of 3rd Conference on Machine Translation: Research Papers. Brussels, Belgium: Association for Computational Linguistics, 2018. 232–243
- 85 Maimaiti M, Liu Y, Luan H B, Sun M S. Multi-round transfer learning for low-resource NMT using multiple high-resource languages. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 2019, 18(4): Article No. 38
- 86 Wang X Y, Neubig G. Target conditioned sampling: Optimizing data selection for multilingual neural machine translation. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019. 5823–5828
- 87 Dabre R, Fujita A, Chu C H. Exploiting multilingualism through multistage fine-tuning for low-resource neural machine translation. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China: Association for Computational Linguistics, 2019. 1410–1416
- 88 Murthy R, Kunchukuttan A, Bhattacharyya P. Addressing word-order divergence in multilingual neural machine translation for extremely low resource languages. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis, Minnesota: Association for Computational Linguistics, 2019. 3868–3873
- 89 Imankulova A, Sato T, Komachi M. Filtered pseudo-parallel corpus improves low-resource neural machine translation. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 2019, 19(2): Article No. 24
- 90 Neubig G, Hu J J. Rapid adaptation of neural machine translation to new languages. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium: Association for Computational Linguistics, 2018. 875–880
- 91 Lu Y C, Keung P, Ladhak F, Bhardwaj V, Zhang S N, Sun J. A neural interlingua for multilingual machine translation. In: Proceedings of the 3rd Conference on Machine Translation: Research Papers. Belgium, Brussels: Association for Computational Linguistics, 2018. 84–92
- 92 Sestorain L, Ciaramita M, Buck C, Hofmann T. Zero-shot dual machine translation. arXiv preprint arXiv: 1805.10338, 2018
- 93 Wang Y N, Zhou L, Zhang J J, Zhai F F, Xu J F, Zong C Q. A compact and language-sensitive multilingual translation method. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019. 1213–1223
- 94 Kiperwasser E, Ballesteros M. Scheduled multi-task learning: From syntax to translation. *Transactions of the Association for Computational Linguistics*, 2018, 6: 225–240
- 95 Zareemoodi P, Buntine W, Haffari G. Adaptive knowledge sharing in multi-task learning: Improving low-resource neural machine translation. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Melbourne, Australia: Association for Computational Linguistics, 2018. 656–661
- 96 Zareemoodi P, Haffari G. Adaptively scheduled multitask learning: The case of low-resource neural machine translation. In: Proceedings of the 3rd Workshop on Neural Generation and Translation. Hong Kong, China: Association for Computational Linguistics, 2019. 177–186
- 97 Zoph B, Knight K. Multi-source neural translation. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego, California, USA: Association for Computational Linguistics, 2016. 30–34
- 98 He D, Xia Y C, Qin T, Wang L W, Yu N H, Liu T Y, et al. Dual learning for machine translation. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: Curran Associates Inc., 2016. 820–828
- 99 He T Y, Chen J L, Tan X, Qin T. Language graph distillation for low-resource machine translation. arXiv preprint arXiv: 1908.06258, 2019
- 100 Östling R, Tiedemann J. Neural machine translation for low-resource languages. arXiv preprint arXiv: 1708.05729, 2017
- 101 Nishimura Y, Sudoh K, Neubig G, Nakamura S. Multi-source neural machine translation with data augmentation. In: Proceedings of the 15th International Workshop on Spoken Language Translation. Bruges, Belgium, 2018. 48–53
- 102 Garcia X, Foret P, Sellam T, Parikh A P. A multilingual view of unsupervised machine translation. arXiv preprint arXiv: 2002.02955, 2020
- 103 Anastasopoulos A, Chiang D. Leveraging translations for speech transcription in low-resource settings. In: Proceedings of Interspeech 2018. Hyderabad, India, 2018. 1279–1283
- 104 Stoian M C, Bansal S, Goldwater S. Analyzing ASR pretraining for low-resource speech-to-text translation. In: Proceedings of the 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Barcelona, Spain: IEEE, 2020. 7909–7913
- 105 Bansal S, Kamper H, Livescu K, Lopez A, Goldwater S. Low-resource speech-to-text translation. In: Proceedings of Interspeech 2018. Hyderabad, India, 2018. 1298–1302
- 106 Anastasopoulos A, Chiang D, Duong L. An unsupervised probability model for speech-to-translation alignment of low-resource languages. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, Texas: Association for Computational Linguistics, 2016. 1255–1263
- 107 Erdmann A, Habash N, Taji D, Bouamor H. Low resourced ma-

chine translation via morpho-syntactic modeling: The case of dialectal Arabic. In: Proceedings of Machine Translation Summit XVI. Nagoya, Japan, 2017. 185–200

- 108 Honnet P E, Popescu-Belis A, Musat C, Baeriswyl M. Machine translation of low-resource spoken dialects: Strategies for normalizing Swiss German. In: Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018). Miyazaki, Japan: European Language Resources Association (ELRA), 2018. 3781–3788
- 109 Musleh A, Durrani N, Temnikova I, Nakov P, Vogel S, Alsaad O. Enabling medical translation for low-resource languages. In: Proceedings of the 2016 International Conference on Intelligent Text Processing and Computational Linguistics. Konya, Turkey: Springer, 2016. 3–16
- 110 Ngo T V, Ha T L, Nguyen P T, Nguyen L M. Overcoming the rare word problem for low-resource language pairs in neural machine translation. In: Proceedings of the 6th Workshop on Asian Translation. Hong Kong, China: Association for Computational Linguistics, 2019. 207–214
- 111 Wang Zhuo, Yu Zheng-Tao, Wen Yong-Hua, Gao Sheng-Xiang, Wu Fei. Chinese-Vietnamese neural machine translation integrated with lexicon probability. *Journal of Kunming University of Science and Technology (Natural Science Edition)*, 2019, 44(1): 54–60
(王卓, 余正涛, 文永华, 高盛祥, 吴飞. 融合词汇翻译概率的汉越神经机器翻译方法. 昆明理工大学学报(自然科学版), 2019, 44(1): 54–60)
- 112 Che Wan-Jin, Yu Zheng-Tao, Guo Jun-Jun, Wen Yong-Hua, Yu Zhi-Qiang. Unknown words processing method for Chinese-Vietnamese neural machine translation based on hybrid network integrating classification dictionaries. *Journal of Chinese Information Processing*, 2019, 33(12): 67–75
(车万金, 余正涛, 郭军军, 文永华, 于志强. 融入分类词典的汉越混合神经网络机器翻译集外词处理方法. 中文信息学报, 2019, 33(12): 67–75)
- 113 Xu Yu, Lai Hua, Yu Zheng-Tao, Gao Sheng-Xiang, Wen Yong-Hua. Chinese-Vietnamese neural machine translation based on deep separable convolution. *Journal of Xiamen University (Natural Science)*, 2020, 59(2): 220–224
(徐毓, 赖华, 余正涛, 高盛祥, 文永华. 基于深度可分离卷积的汉越神经机器翻译. 厦门大学学报(自然科学版), 2020, 59(2): 220–224)
- 114 Jia Cheng-Xun, Lai Hua, Yu Zheng-Tao, Wen Yong-Hua, Yu Zhi-Qiang. Pseudo-parallel corpus generation for Chinese-Vietnamese neural machine translation based on pivot language. *Computer Engineering & Science*, 2021, 43(3): 542–550
(贾承勋, 赖华, 余正涛, 文永华, 于志强. 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, 2021, 43(3): 542–550)
- 115 Yu Zhi-Qiang, Yu Zheng-Tao, Huang Yu-Xin, Guo Jun-Jun, Gao Sheng-Xiang. Improving semi-supervised neural machine translation with variational information bottleneck. *Acta Automatica Sinica*, DOI: 10.16383/j.aas.c190477
(于志强, 余正涛, 黄于欣, 郭军军, 高盛祥. 基于变分信息瓶颈的半监督神经机器翻译. 自动化学报, DOI: 10.16383/j.aas.c190477)
- 116 Devlin J, Chang M W, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis, Minnesota: Association for Computational Linguistics, 2019. 4171–4186
- 117 Clinchant S, Jung K W, Nikoulina V. On the use of BERT for neural machine translation. In: Proceedings of the 3rd Workshop on Neural Generation and Translation. Hong Kong, China: Association for Computational Linguistics, 2019. 108–117
- 118 Zhu J H, Xia Y C, Wu L J, He D, Qin T, Zhou W G, et al. Incorporating BERT into neural machine translation. In: Proceedings of the 8th International Conference on Learning Representations. Addis Ababa, Ethiopia, 2020.
- 119 Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training [Online],

available: <https://www.cs.ubc.ca/~amuham01/LING530/papers/radford2018improving.pdf>, April, 4, 2020



李洪政 北京理工大学外国语学院助理教授(特别副研究员)。2018–2020年在北京理工大学计算机学院从事博士后研究工作, 2018年获北京师范大学中文信息处理研究所博士学位。主要研究方向为自然语言处理, 机器翻译, 语料库语言学。

E-mail: lihongzheng@bit.edu.cn

(LI Hong-Zheng Assistant professor (special associate research fellow) at the School of Foreign Languages, Beijing Institute of Technology (BIT). He was a postdoctoral researcher at the School of Computer Science and Technology, BIT from 2018 to 2020. He received his Ph. D. degree from the Institute of Chinese Information and Processing, Beijing Normal University in 2018. His research interest covers natural language processing, machine translation and corpus linguistics.)



冯冲 北京理工大学计算机学院教授。2005年获中国科学技术大学计算机科学系博士学位。主要研究方向为自然语言处理, 信息抽取, 机器翻译。

E-mail: fengchong@bit.edu.cn

(FENG Chong Professor at the School of Computer Science and Technology, Beijing Institute of Technology. He received his Ph. D. degree from the Department of Computer Science, University of Science and Technology of China in 2005. His research interest covers natural language processing, information extraction, and machine translation.)



黄河燕 北京理工大学计算机学院教授。1989年获中国科学院计算技术研究所计算机科学与技术博士学位。主要研究方向为自然语言处理, 机器翻译, 社交网络, 信息检索, 智能处理系统。本文通信作者。

E-mail: hhy63@bit.edu.cn

(HUANG He-Yan Professor at the School of Computer Science and Technology, Beijing Institute of Technology. She received her Ph. D. degree from the Institute of Computing Technology, Chinese Academy of Sciences in 1989. Her research interest covers natural language processing, machine translation, social network, information retrieval, and intelligent processing system. Corresponding author of this paper.)