

文章编号: 2095-2163(2021)02-0001-06

中图分类号: TP391; TH166

文献标志码: A

# 无监督神经机器翻译综述

孙海鹏, 赵铁军

(哈尔滨工业大学 计算机科学与技术学院, 哈尔滨 150001)

**摘要:** 神经机器翻译模型的学习往往依赖于大规模的平行语料, 然而并不是所有语言对都有充足的平行语料。为了解决神经机器翻译中平行语料缺失的问题, 研究人员提出了无监督神经机器翻译, 通过无监督预训练策略、去噪自编码器、反向翻译和共享潜在表示机制仅依赖于单语语料对翻译任务进行建模。本文从模型主要机制、训练过程出发来分析无监督神经机器翻译, 并介绍无监督神经机器翻译现阶段的发展以及存在的挑战。

**关键词:** 无监督神经机器翻译; 去噪自编码器; 反向翻译

## A survey on unsupervised neural machine translation

SUN Haipeng, ZHAO Tiejun

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

**[Abstract]** The training of neural machine translation often relies on large-scale parallel corpora, but not all language pairs have sufficient parallel corpora. In order to alleviate the problem of the lack of parallel corpora in neural machine translation, unsupervised neural machine translation is proposed to model translation relying solely on monolingual corpora with the help of a combination of diverse mechanisms such as unsupervised pre-training strategy, denoising auto-encoder, back-translation, and shared latent representation. This paper analyzes unsupervised neural machine translation from the perspective of model mechanism and training process, and introduces the advances and challenges of unsupervised neural machine translation at the present stage.

**[Key words]** unsupervised neural machine translation; denoising auto-encoder; back-translation

## 0 引言

自注意力机制和 transformer<sup>[1]</sup>模型的提出显著提升了神经机器翻译模型的性能, 但是神经机器翻译模型训练需要大量的平行语料, 而实际上大多数语言对的平行语料并不充足, 甚至一些稀缺语言对只有少量的平行语料。通常构建平行语料库的成本很高, 而平行语料的缺失往往会导致神经机器翻译模型性能的急速下降<sup>[2]</sup>, 为了缓解平行语料缺失对神经机器翻译模型性能产生的消极影响, 研究者们开始尝试仅使用单语语料来训练无监督神经机器翻译模型<sup>[3-4]</sup>。相较于平行语料, 单语语料更容易挖掘与获取。许多只有有限平行数据的语言仍然拥有大量的单语数据。无监督双语词嵌入<sup>[4-5]</sup>的出现也使得神经机器翻译模型在无监督模式下进行成为了可能, 利用去噪自编码器<sup>[6]</sup>和反向翻译<sup>[7]</sup>训练在不使用平行语料的前提下完成了双语间的翻译建模。

本文的框架如下: 首先介绍无监督神经机器翻译的模型主要机制及训练过程, 然后详述无监督神

经机器翻译的发展过程, 再给出现阶段无监督神经机器翻译存在的挑战, 最后是论文研究结论。

## 1 无监督神经机器翻译的主要机制

传统的无监督神经机器翻译模型主要借助于无监督预训练方法、去噪自编码器、反向翻译和共享潜在表示机制在单语数据集上实现了双语间的翻译建模。具体地, 给定2种语言 $L_1$ 和 $L_2$ 的单语数据 $X$ 和 $Y$ ,  $X$ 和 $Y$ 的单语数据空间被表示为 $\varphi_{L_1}$ 和 $\varphi_{L_2}$ 。

无监督神经机器翻译通过最大化的目标函数来进行训练, 此时需用到如下公式:

$$L_{all} = L_{DN} + L_{BT} \quad (1)$$

其中,  $L_{DN}$ 表示去噪自编码器的目标函数;  $L_{BT}$ 表示反向翻译的目标函数。

无监督神经机器翻译的训练过程如图1所示。由图1可知, 无监督神经机器翻译模型的整个训练过程中需要考虑各自语言的去噪自编码器过程和2种语言间的反向翻译过程, 无监督神经机器翻译训练在 $L_1$ 语言和 $L_2$ 语言的句子之间交替进行。具体

基金项目: 国家重点研发计划项目(2017YFB1002102)。

作者简介: 孙海鹏(1990-), 男, 博士研究生, 主要研究方向: 自然语言处理、机器翻译; 赵铁军(1962-), 男, 博士, 教授, 博士生导师, 主要研究方向: 自然语言处理、机器翻译、机器学习与人工智能。

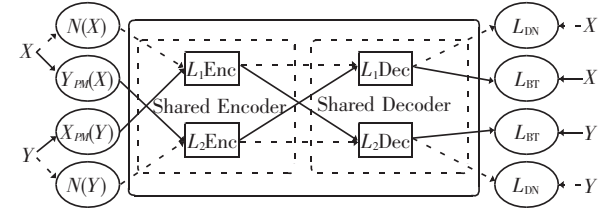
通讯作者: 赵铁军 Email: tjzhao@hit.edu.cn

收稿日期: 2020-10-21

哈尔滨工业大学主办 ◆ 学术研究与应用

地,每次无监督神经机器翻译训练迭代过程包括: $L_1 \rightarrow L_1$  去噪自编码器训练、 $L_2 \rightarrow L_2$  去噪自编码器训练、 $L_1 \rightarrow L_2$  反向翻译训练和  $L_2 \rightarrow L_1$  反向翻译训练。

$L_1$  monolingual data



$L_2$  monolingual data

DN: $L_1$ Enc $\rightarrow L_1$ Dec BT: $L_1$ Enc $\rightarrow L_2$ Dec generation  $L_2$ Enc $\rightarrow L_1$ Dec training  
 $L_2$ Enc $\rightarrow L_2$ Dec  $L_2$ Enc $\rightarrow L_1$ Dec generation  $L_1$ Enc $\rightarrow L_2$ Dec training

图1 无监督神经机器翻译的训练过程图

Fig. 1 The training process of unsupervised neural machine translation

### 1.1 无监督预训练方法

与有监督神经机器翻译模型<sup>[1,8]</sup>相比,无监督神经机器翻译模型在训练过程中并不存在监督的双语信号。因此,在无监督神经机器翻译模型训练开始前,需要通过无监督预训练模型对无监督神经机器翻译模型进行初始化提供朴素的双语信号。

最初,无监督双语词嵌入被用来对无监督神经机器翻译模型编码器和解码器的词嵌入层进行初始化。无监督双语词嵌入<sup>[4-5]</sup>可以从2个非平行的单语语料库中学习了词对翻译。通过双语词嵌入初始化后,编码器就得到了语言独立的词级表示,且只需要学习如何将其组合起来以构建更大的短语表示。无监督双语词嵌入提供朴素的翻译知识,使反向翻译可以产生伪监督双语信号<sup>[3-4]</sup>。

随着预训练语言模型的广泛研究,预训练语言模型<sup>[9]</sup>也被用来初始化无监督神经机器翻译模型的编码器和解码器。通过语言模型训练来建立一个跨语言编码器,该编码器将2个单语句子编码到共享的潜在空间中。与无监督双语词嵌入相比,预训练语言模型可以为无监督神经机器翻译模型提供更多的跨语言信息。

### 1.2 去噪自编码器

由于无监督神经机器翻译模型的对偶结构,可以直接通过重建其自身的输入进行训练。具体来说,将指定语言的输入句子进行编码,再用该语言的解码器重新构造原始的句子,来对无监督神经机器翻译模型进行优化。如果不添加一些约束条件,普通的自编码器很难学习到有用的知识,将变成一个

复制任务,只学习到对输入单词的一对一复制<sup>[3-4]</sup>。为了改进这一问题不足,引入了去噪自编码器策略<sup>[6]</sup>,并在输入语句中增加了噪声,以提高模型学习能力<sup>[10-11]</sup>。研究中,即选择了2种不同类型噪声加入到输入语句中<sup>[4]</sup>。在输入语句中的任意单词以概率被删除,概率越大,输入语句中被删除的单词越多。第二类噪声是对输入语句的语序进行轻微调整。通过这种去噪自编码器策略,无监督神经机器翻译系统需要了解所涉及语言的内部结构,以便能够恢复正确的词序。同时,通过阻止无监督神经机器翻译系统过多地依赖输入语句的词序,语言之间的实际语序差异可以更好地被解释。

去噪自编码器对有上述噪声的输入语句  $N(X)$  和  $N(Y)$  进行编码并用同种语言  $L_1$  和  $L_2$  的解码器对其进行重构,通过如下公式的最小化所表示的目标函数来进行优化,即:

$$L_{DN} = E_{X \sim \varphi_{L_1}} [-\log P_{L_1 \rightarrow L_1}(X | N(X))] + E_{Y \sim \varphi_{L_2}} [-\log P_{L_2 \rightarrow L_2}(Y | N(Y))] \quad (2)$$

其中,  $P_{L_1 \rightarrow L_1}$  和  $P_{L_2 \rightarrow L_2}$  表示同一种语言间的重构概率。

### 1.3 反向翻译

去噪自编码器作为一个语言模型,在同一种语言中进行训练,并没有考虑在2种语言之间翻译的最终目标。因此将引入反向翻译<sup>[7]</sup>来训练这个仅使用单语语料的无监督翻译系统。反向翻译可以将这个无监督问题转换成为有监督的学习任务,尽管是带有噪声的源端语句。在给定一种语言  $L_1$  的输入语句的情况下,推理模式使用  $L_1$  编码器进行编码,用  $L_2$  解码器进行解码,同时又使用了贪心解码来将其转换为另一种语言的译文。通过这种方法,即获取了伪平行句对,并将其用来训练无监督神经机器翻译系统,再从这个带有噪声的合成翻译中预测出原有语句。标准的反向翻译是使用一个独立的模型一次性反向翻译整个语料库,与标准的反向翻译不同,由于无监督神经机器翻译系统的对偶结构,在无监督神经机器翻译模型训练过程中,正在训练无监督神经机器翻译模型将会用于实时反向翻译每一个 batch 的语句。这样,随着训练的进行和模型的改进,通过反向翻译会产生更好的伪平行句对,必将有助于在接下来的迭代中进一步改进模型。

这里,给出单语语料  $X$  和  $Y$ , 句子  $Y_{PM}(X)$  和  $X_{PM}(Y)$  将分别由之前迭代的模型生成,并已获得伪平行句对  $(Y_{PM}(X), X)$  和  $(X_{PM}(Y), Y)$  来训练新的

翻译模型。在此基础上,反向翻译模型即对生成的输入语句  $Y_{PM}(X)$  和  $X_{PM}(Y)$  进行编码,再用另一种语言的解码器产生译文,其通过如下公式的最小化所表示的目标函数来进行优化,即:

$$L_{BT} = E_{X \sim \varphi_{L_1}} [-\log P_{L_2 \rightarrow L_1}(X | Y_{PM}(X))] + E_{Y \sim \varphi_{L_2}} [-\log P_{L_1 \rightarrow L_2}(Y | X_{PM}(Y))] \quad (3)$$

其中,  $P_{L_2 \rightarrow L_1}$  和  $P_{L_1 \rightarrow L_2}$  表示 2 种语言间的翻译概率。

#### 1.4 共享潜在表示

2 种语言  $L_1$  和  $L_2$  的编码器和解码器是共享的或者部分共享的,2 种语言  $L_1$  和  $L_2$  使用相同的词汇表。不管输入源端语言是什么,共享编码器的目的是产生输入文本的语言独立的潜在表示,解码器可以通过这个潜在表示转换成相应语言的目标译文。这确保了通过去噪自编码器实现的语言模型的效益,可以从有噪声的源端句子很好地转换到相应的目标译文,最终有助于无监督神经机器翻译模型来获取更流畅的译文。共享编码器是实现无监督翻译模型工作的关键,而共享解码器只是引入有用的正则化。

## 2 无监督神经机器翻译的优化方法

### 2.1 模型参数共享

Artetxe 等人<sup>[3]</sup>和 Lample 等人<sup>[4]</sup>使用去噪自编码器和反向翻译在没有平行语料的情况下进行无监督神经机器翻译训练,与 Artetxe 等人<sup>[3]</sup>仅采用共享编码器方法不同的是,Lample 等人<sup>[4]</sup>不仅采用共享编码器,也将共享解码器,只是针对不同语言会使用不同的词嵌入层。Yang 等人<sup>[12]</sup>认为共享编码器对于将不同语言的句子映射到共享的潜在空间至关重要,但却并不能很好地保持每种语言的独特性和内部特征。由于每种语言都有自己的特性,源语言和目标语言应独立编码和学习,共享编码器会是限制无监督神经机器翻译性能的一个因素。为了解决这个问题,Yang 等人<sup>[12]</sup>利用 2 个独立的编码器来代替原有的共享编码器,每个编码器对应一种语言,也同样会用到 2 个独立的解码器。为了在训练过程中更好地将不同语言的潜在表示映射到共享潜在空间,可利用权重共享来约束现有的无监督神经机器翻译模型。总地来说,负责抽取输入句子的高级表示的 2 个编码器的末端多层的权重以及 2 个解码器的前端多层的权重将被共享。稍后的大部分无监督翻译工作也都是基于这三种参数共享框架开展的。

### 2.2 模型机制优化

通常,无监督神经机器翻译模型主要包括无监督预训练模型、去噪自编码器、反向翻译和共享潜在表示四个机制,很多工作通过优化这 4 个机制来提升无监督神经机器翻译的性能。

与先前研究工作<sup>[3-4,12]</sup>使用双语词嵌入作为翻译模型的初始化不同,Lample 等人<sup>[13]</sup>提出了一种更适合相似语言对的初始化方法,并没有学习源语言和目标语言之间的显式映射,而是将 2 个单语语料库合并成一个语料库,对其进行 BPE<sup>[14]</sup>操作后,在这个语料库上学习 token 级的词嵌入<sup>[15]</sup>,这个词嵌入将用来初始化翻译模型中编码器和解码器的词嵌入层。然而,这种初始化方法并不适用于一些不相似语言对。

随着预训练语言模型被广泛研究,Conneau 等人<sup>[9]</sup>,Song 等人<sup>[16]</sup>和 Ren 等人<sup>[17]</sup>使用跨语言掩码语言模型预训练方法代替原有的词嵌入预训练方法对翻译模型中的所有参数进行初始化,大幅度提升无监督神经机器翻译模型性能。与词嵌入预训练方法相比,语言模型预训练方法可以为无监督神经机器翻译模型提供更多的跨语言信息。Sun 等人<sup>[18-19]</sup>提出跨语言一致性方法来进一步提升翻译性能。

Kim 等人<sup>[20]</sup>并没有使用反向翻译进行模型训练,提出一种基于双语词嵌入单词翻译的方法,来提升模型的训练速度。与此同时,Wu 等人<sup>[21]</sup>也认为由于生成的伪平行句对通常质量不高,翻译误差容易积累,为此,Wu 等人<sup>[21]</sup>从可比的单语语料库中抽取真实的平行语句来训练翻译模型。

### 2.3 语言对选择

无监督神经机器翻译研究工作通常是在法英、德英等相似语言对上对系统的有效性进行验证。后来一些研究者也在不相似语言对和多语言方面开展无监督神经机器翻译研究。Leng 等人<sup>[22]</sup>通过引入中轴语翻译来解决无监督神经机器翻译在不相似语言对中性能低的问题。Xu 等人<sup>[23]</sup>提出多边形网络(Polygon-Net)框架,利用额外的辅助语言联合学习增强无监督神经机器翻译模型的性能。Sen 等人<sup>[24]</sup>通过一个共享编码器和多个特定语言的解码器来联合训练多语言无监督神经机器翻译模型。Sun 等人<sup>[25]</sup>在大规模欧洲单语语料上训练多语言无监督神经机器翻译模型,并提出特有的知识蒸馏方法改善翻译性能。

### 2.4 引入统计机器翻译

随着无监督神经机器翻译模型在一些语言对上

可以获得显著性能的同时,许多研究者开始探索统计机器翻译的模块化特性是否更适合这种仅依赖单语语料库的无监督设置,提出了全新的无监督统计机器翻译模型<sup>[13,26-27]</sup>。Lample 等人<sup>[13]</sup>, Marie 等人<sup>[28]</sup>, Ren 等人<sup>[29]</sup>和 Artetxe 等人<sup>[27]</sup>将无监督神经机器翻译与无监督统计机器翻译相结合进一步提升翻译性能。Lample 等人<sup>[13]</sup>将无监督统计机器翻译系统生成的数据添加到无监督神经机器翻译模型生成的反向翻译数据中进一步优化无监督神经机器翻译模型。Marie 等人<sup>[28]</sup>提出在无监督翻译场景下使用监督的神经机器翻译框架,只是使用无监督统计机器翻译模型生成的伪平行语料代替真实的平行语料,并且使用正向翻译代替反向翻译以提高翻译质量,消除对2个翻译方向同时进行训练的需要。Ren 等人<sup>[29]</sup>利用统计机器翻译作为后验正则化在迭代反向翻译过程中去噪和指导无监督神经机器翻译模型训练,来消除反向翻译带来随机误差。Ren 等人<sup>[30]</sup>认为使用双语词嵌入方法初始化短语表很难恢复一些训练语句中的复杂句子结构,并且这个短语表中会存在大量噪声。为了解决这一问题, Ren 等人<sup>[30]</sup>利用无监督的跨语言句子嵌入方法从2种语言的单语语料库中检索语义上相似的句子对,再通过 GIZA++<sup>[31]</sup>来获得单词对齐,删除检索到的目标句子中未对齐的单词,并使用一个序列到序列的重写模型对修改后的目标句进行重写,以减少源语言语句和目标语言语句之间的语义差距。

### 2.5 引入额外信息

源端语句和目标端语句在潜在空间中进行关联是无监督神经机器翻译模型的一个难点。Su 等人<sup>[32]</sup>和 Huang 等人<sup>[33]</sup>提出多模态无监督神经机器翻译模型,利用图像信息在翻译训练过程中消除歧义,促进潜在空间的对齐,进一步提升翻译性能。Duan 等人<sup>[34]</sup>使用额外的双语词典来建立锚点,缩小源语言和目标语言之间的差距以提升翻译性能。Garcia 等人<sup>[35]</sup>和 Li 等人<sup>[36]</sup>引入第三种语言,这种语言与源端或者目标端之间存在平行语料,利用额外引入的平行语料来提升翻译性能。

### 2.6 其他

Pourdamghani 等人<sup>[37]</sup>通过将无监督神经机器翻译训练分解为2个阶段。第一阶段,通过一个源端到目标端的词典生成粗略的译文(Translationese),翻译过程中采用 Kim 等人<sup>[20]</sup>的方法加入语言模型,基于上下文信息进行词汇选择。第二阶段,使用一些高资源语言对的平行语料训练

一个由 Translationese 到流畅目标端译文的翻译模型。在迁移到新的源语言时,该方法不需要对翻译模型进行重新训练,只需要一个源端到目标端的字典,这使得该方法可以快速地应用到其他新的语言中。Sun 等人<sup>[38]</sup>通过对抗训练方法提高无监督神经机器翻译的鲁棒性。近年来,国际机器翻译评测(WMT)也引入了无监督机器翻译任务,吸引许多研究者们关注与参与<sup>[39-40]</sup>。

## 3 无监督神经机器翻译的挑战

双语词嵌入的出现使机器翻译在单语语料上以无监督方式完成双语间的翻译成为了可能。尽管现如今大量工作集中在无监督神经机器翻译的相关研究上,但是无监督神经机器翻译的研究依然存在一些不足和局限性<sup>[41-42]</sup>。

目前,无监督神经机器翻译的研究工作多数情况下是在英法、英德等相似语言对上验证方法的有效性,并没有考虑语言的多样性。通常,在一些不相似语言对上双语词嵌入准确率并不高<sup>[43]</sup>,无监督神经机器翻译在这些语言对上则会获得更差的性能。如何在诸如中英、日英等不相似语言对上提升无监督神经机器翻译质量是一个值得研究的课题。此外,在一些真正低资源语言对上进行无监督神经机器翻译训练也亟待进一步探索。

相较于有监督机器翻译,无监督神经机器翻译领域自适应问题并未获得充分研究。除了有监督存在的训练集和测试集领域不一致以外,无监督翻译的训练集间也可能存在领域不一致的问题<sup>[44]</sup>,从而使得无监督神经机器翻译领域自适应问题变得更为复杂。此外,单语数据大小以及2种语言间单语数据间存在的数量不平衡性<sup>[45]</sup>也会影响无监督神经机器翻译性能。

传统的无监督神经机器翻译模型在训练过程中会同时对一个语言对的2个不同翻译方向进行训练,无监督神经机器翻译的训练过程主要依赖无监督预训练策略、去噪自编码器、反向翻译和共享潜在表示机制,然而这些机制本身存在自身的局限性。预训练模型的质量会显著影响翻译模型的质量。通过去噪自编码器训练可以使无监督神经机器翻译模型获得更流畅通顺的译文,然而通过不断修改训练数据进行去噪自编码器训练也极大地减缓了模型的训练速度。反向翻译是在单语语料上实现双语间翻译的关键,但是反向翻译也存在自身的不足。反向翻译生成的伪平行句对通常质量不高,翻译误差容

易累积叠加,这会阻碍无监督神经机器翻译系统的进一步提升。此外,无监督神经机器翻译训练过程中,需要通过反向翻译不断生成译文,反向翻译的解码速度也限制了无监督神经机器翻译模型的训练速度。如何设计高效的反向翻译方法将是未来无监督神经机器翻译研究的热点课题。共享潜在表示只是在无监督神经机器翻译训练前期是必需的,而随着无监督神经机器翻译训练,共享潜在表示机制会迅速变得低效。学习共享潜在表示会限制了2个不同方向翻译性能的进一步提升,特别是对于不相似语言对。

## 4 结束语

无监督神经机器翻译解决了因平行语料缺失造成神经机器翻译出现瓶颈的问题,显著降低了训练成本。本文介绍了无监督神经机器翻译的模型主要机制、训练方法及其发展过程,并分析了无监督神经机器翻译存在的局限性。未来期待更多的工作,研究完善无监督神经机器翻译,进一步缩小与有监督神经机器翻译的性能差距。

## 参考文献

[1] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]// Advances in Neural Information Processing Systems 30. Long Beach, CA, USA: Curran Associates, Inc., 2017: 5998–6008.

[2] KOEHN P, KNOWLES R. Six challenges for neural machine translation [C]// Proceedings of the First Workshop on Neural Machine Translation. Vancouver: Association for Computational Linguistics, 2017: 28–39.

[3] ARTETXE M, LABAKA G, AGIRRE E, et al. Unsupervised neural machine translation [C]// Proceedings of the Sixth International Conference on Learning Representations. Vancouver, Canada: OpenReview.net, 2018:1–11.

[4] LAMPLE G, CONNEAU A, DENOYER L, et al. Unsupervised machine translation using monolingual corpora only [C]// Proceedings of the Sixth International Conference on Learning Representations. Vancouver, Canada: OpenReview.net, 2018:1–14.

[5] CONNEAU A, LAMPLE G, RANZATO M, et al. Word translation without parallel data [C]// Proceedings of the Sixth International Conference on Learning Representations. Vancouver, Canada: OpenReview.net, 2018:1–14.

[6] VINCENT P, LAROCHELLE H, LAJOIE I, et al. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion [J]. Journal of Machine Learning Research, 2010, 11: 3371–3408.

[7] SENNRICH R, HADDOW B, BIRCH A. Improving neural machine translation models with monolingual data [C]// Proceedings of the 54<sup>th</sup> Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Berlin,

Germany: Association for Computational Linguistics, 2016: 86–96.

[8] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate [C]// Proceedings of the 3<sup>rd</sup> International Conference on Learning Representations. San Diego, CA, USA: DBLP, 2015:1–15.

[9] CONNEAU A, LAMPLE G. Cross-lingual language model pretraining [C]// Advances in Neural Information Processing Systems 32. Vancouver, Canada: Curran Associates, Inc., 2019: 7059–7069.

[10] HE D, XIA Y, QIN T, et al. Dual learning for machine translation [C]// Advances in Neural Information Processing Systems 29. Barcelona, Spain: Curran Associates, Inc., 2016: 820–828.

[11] HILL F, CHO K, KORHONEN A. Learning distributed representations of sentences from unlabelled data [C]// 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: 1367–1377.

[12] YANG Zhen, CHEN Wei, WANG Feng, et al. Unsupervised neural machine translation with weight sharing [C]// Proceedings of the 56<sup>th</sup> Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne, Australia: Association for Computational Linguistics, 2018: 46–55.

[13] LAMPLE G, OTT M, CONNEAU A, et al. Phrase-based neural unsupervised machine translation [C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium: Association for Computational Linguistics, 2018: 5039–5049.

[14] SENNRICH R, HADDOW B, BIRCH A. Neural machine translation of rare words with subword units [C]// Proceedings of the 54<sup>th</sup> Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Berlin, Germany: Association for Computational Linguistics, 2016: 1715–1725.

[15] MIKOLOV T, SUTSKEVER I, CHEN Kai, et al. Distributed representations of words and phrases and their compositionality [C]// Advances in Neural Information Processing Systems 26. Lake Tahoe, Nevada, United States: Curran Associates, Inc., 2013: 3111–3119.

[16] SONG Kaitao, TAN Xu, QIN Tao, et al. MASS: Masked sequence to sequence pre-training for language generation [C]// Proceedings of the 36<sup>th</sup> International Conference on Machine Learning. Long Beach, California, USA: PMLR, 2019: 5926–5936.

[17] REN Shuo, WU Yu, LIU Shujie, et al. Explicit cross-lingual pre-training for unsupervised machine translation [C]// Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9<sup>th</sup> International Joint Conference on Natural Language Processing. Hong Kong, China: Association for Computational Linguistics, 2019: 770–779.

[18] SUN Haipeng, WANG Rui, CHEN Kehai, et al. Unsupervised bilingual word embedding agreement for unsupervised neural machine translation [C]// The 57<sup>th</sup> Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019: 1235–1245.

[19] SUN H, WANG R, CHEN K, et al. Unsupervised neural machine translation with cross-lingual language representation



- agreement[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2020, 28:1170–1182.
- [20] KIM Y, GENG J, NEY H. Improving unsupervised word-by-word translation with language model and denoising autoencoder [C] // *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Brussels, Belgium: Association for Computational Linguistics, 2018: 862–868.
- [21] WU Jiawei, WANG Xin, WANG Yang. Extract and edit: An alternative to back-translation for unsupervised neural machine translation [C] // *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.
- [22] LEMG Yichong, TAN Xu, QIN Tao, et al. Unsupervised pivot translation for distant languages [C] // *Proceedings of the 57<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, 2019: 175–183.
- [23] XU Chang, QIN Tao, WANG Gang, et al. Polygon-Net: A general framework for jointly boosting multiple nsupervised neural machine translation models [C] // *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*. Macao, China: ijcai.org, 2019: 5320–5326.
- [24] SEN S, GUPTA K K, EKBAL A, et al. Multilingual unsupervised NMT using shared encoder and language-specific decoders [C] // *Proceedings of the 57<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, 2019: 3083–3089.
- [25] SUN Haipeng, WANG Rui, CHEN Kehai, et al. Knowledge distillation for multilingual unsupervised neural machine translation [C] // *The 58<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*, Online: Association for Computational Linguistics, 2020: 3525–3535.
- [26] ARTETXE M, LABAKA G, AGIRRE E. Unsupervised statistical machine translation [C] // *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Brussels, Belgium: Association for Computational Linguistics, 2018:3632–3642.
- [27] ARTETXE M, LABAKA G, AGIRRE E. An effective approach to unsupervised machine translation [C] // *Proceedings of the 57<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, 2019: 194–203.
- [28] MARIE B, FUJITA A. Unsupervised neural machine translation initialized by unsupervised statistical machine translation [J]. *CoRR*, 2018, abs/1810.12703.
- [29] REN Shuo, ZHANG Zhirui, LIU Shujie, et al. Unsupervised neural machine translation with SMT as posterior regularization [C] // *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence*. Honolulu, Hawaii, USA: AAAI Press, 2019: 241–248.
- [30] REN Shuo, WU Yu, LIU Shujie, et al. A retrieve-and-rewrite initialization method for unsupervised machine translation [C] // *Proceedings of the 58<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics, 2020: 3498–3504.
- [31] OCH F J, NEY H. A systematic comparison of various statistical alignment models [J]. *Computational Linguistics*, 2003, 29(1): 19–51.
- [32] SU Yuanhang, FAN Kai, BACH N, et al. Unsupervised multi-modal neural machine translation [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Long Beach, California, USA: Computer Vision Foundation / IEEE, 2019: 10482–10491.
- [33] HUANG Poyao, HU Junjie, CHANG Xiaojun, et al. Unsupervised multimodal neural machine translation with pseudo visual pivoting [C] // *Proceedings of the 58<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics, 2020: 8226–8237.
- [34] DUAN Xiangyu, JI Baijun, JIA Hao, et al. Bilingual dictionary based neural machine translation without using parallel sentences [C] // *Proceedings of the 58<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics, 2020: 1570–1579.
- [35] GARCIA X, FORET P, SELAM T, et al. A multilingual view of Unsupervised Machine Translation [J]. *CoRR*, 2020, abs/2002.02955.
- [36] LI Zuchao, ZHAO Hai, WANG Rui, et al. Reference language based unsupervised neural machine translation [J]. *CoRR*, 2020, abs/2004.02127.
- [37] POURDAMGHANI N, ALDARRAB N, GHAZVININEJAD M, et al. Translating translationese: A two-step approach to unsupervised machine translation [C] // *Proceedings of the 57<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, 2019: 3057–3062.
- [38] SUN Haipeng, WANG Rui, CHEN Kehai, et al. Robust unsupervised neural machine translation with adversarial training [J]. *CoRR*, 2020, abs/2002.12549.
- [39] MARIE B, SUN H, WANG R, et al. NICT’s unsupervised neural and statistical machine translation systems for the WMT19 news translation task [C] // *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1)*. Florence, Italy: Association for Computational Linguistics, 2019: 294–301.
- [40] LI Zuchao, ZHAO Hai, WANG Rui, et al. SJTU–NICT’s supervised and unsupervised neural machine translation systems for the WMT20 news translation task [J]. *CoRR*, 2020, abs/2010.05122.
- [41] KIM Y, GRACA M, NEY H. When and why is unsupervised neural machine translation useless? [C] // *Proceedings of the 22<sup>nd</sup> Annual Conference of the European Association for Machine Translation*. Lisboa, Portugal: European Association for Machine Translation, 2020: 35–44.
- [42] MARCHISIO K, DUH K, KOEHN P. When does unsupervised machine translation work? [J]. *CoRR*, 2020, abs/2004.05516.
- [43] SØGAARD A, RUDER S, VULIĆ I. On the limitations of unsupervised bilingual dictionary induction [C] // *Proceedings of the 56<sup>th</sup> Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne, Australia: Association for Computational Linguistics, 2018: 778–788.
- [44] SUN Haipeng, WANG Rui, CHEN Kehai, et al. Revisiting simple domain adaptation methods in unsupervised neural machine translation [J]. *CoRR*, 2019, abs/1908.09605.
- [45] SUN Haipeng, WANG Rui, CHEN Kehai, et al. Self-training for unsupervised neural machine translation in unbalanced training data scenarios [J]. *CoRR*, 2020, abs/2004.04507.