

# 短语后验证的无监督领域适应电商机器翻译

宁秋怡, 史小静, 段湘煜

(苏州大学, 江苏 苏州 215006)

**摘要:**目前机器翻译训练普遍依赖于大规模平行语料,但在电子商务领域中,公开的大规模平行语料几乎不存在且很难构建。为了解决平行资源稀缺的问题,将大量外领域平行数据迁移应用于内领域非平行数据上的方法称为无监督领域适应。但神经网络机器翻译的领域适应内领域中存在词对匹配不佳的问题,为了解决该问题提出了短语后验证的无监督领域适应电商机器翻译。获取了中文和英文电子商务的短语数据,在无监督领域适应的迭代过程中,采取逐步数据混合训练策略,提升翻译性能,在此基础上引入短语后验证改善词对匹配问题。根据大量电商机器翻译基准实验对比,结果表明,短语后验证的无监督领域适应方法在中英双向电子商务领域上较最强基线系统提升约1.5 Bleu点。

**关键词:**无监督;领域适应;词对匹配;短语后验证;电子商务领域

中图分类号:TP183

文献标识码:A

文章编号:1673-629X(2021)12-0001-06

doi:10.3969/j.issn.1673-629X.2021.12.001

## Unsupervised Domain Adaptation E-commerce Machine Translation Based on Phrase Post-validation

NING Qiu-yi, SHI Xiao-jing, DUAN Xiang-yu

(Soochow University, Suzhou 215006, China)

**Abstract:** At present, machine translation training commonly relies on large-scale parallel corpus, but in the e-commerce domain, public available parallel corpus are non-existent and difficult to construct. In order to solve the problem of scarce parallel resources, the method of applying a large number of parallel data migration from out domain to non-parallel data in domain is called unsupervised domain adaptation. However, there is a problem of poor word pair matching in the in-domain of domain adaptation for neural network machine translation. To solve this problem, an unsupervised domain adaptation e-commerce machine translation based on phrase post-validation is proposed. The monolingual data of Chinese and English e-commerce are obtained. In the iterative process of unsupervised domain adaptation, the step-by-step data mixing training strategy is adopted to improve the translation performance. On this basis, phrase post-validation is introduced to improve the word pair matching problem. According to a large number of baseline experiments of e-commerce machine translation, the unsupervised domain adaptation method based on post-phrase verification is about 1.5 Bleu points higher than the strongest baseline system.

**Key words:** unsupervised; domain adaptation; word pair matching; phrase post-validation; e-commerce domain

## 0 引言

机器翻译(machine translation, MT)<sup>[1-2]</sup>是基于自然语言数据研究的项目,近年来,机器翻译的发展已经达到一个较高的水平,特别是基于编码器和解码器结构的神经机器翻译系统(neural machine translation, NMT)<sup>[3-5]</sup>。但在现阶段,不论是统计机器翻译(statistical machine translation, SMT)<sup>[6]</sup>模型还是神经机器翻译模型,它们的性能取决于语料库的规模和质量。高质量的机器翻译系统训练都依赖于大规模的平行数据,这些数据一般在通用领域中大量存在,如新闻

领域。然而在一些不常用的低资源领域,往往只有小规模平行语料,甚至在一些领域中不存在平行数据,如电子商务领域。如今电子商务活动日益频繁,随之而来的电子商务平台产品信息翻译成为了一个突出的问题。通过人工进行产品信息翻译将消耗大量的时间和资金,因此希望通过机器翻译来辅助电商产品信息翻译,构建电子商务翻译系统。

通过利用其他通用领域训练的翻译系统进行稀缺领域资源的翻译,但是两个领域数据差异较大,翻译效果不佳。可以利用资源丰富的领域语料来帮助语料稀

收稿日期:2021-01-21

修回日期:2021-05-24

基金项目:国家自然科学基金项目(项目号 61673289)

作者简介:宁秋怡(1995-),女,硕士研究生,CCF会员(B1928G),研究方向为机器翻译;段湘煜,博士,教授,研究方向为自然语言处理。

少的领域提升翻译质量,称为机器翻译的领域适应(machine translation adaptation)<sup>[7]</sup>。在领域适应机器翻译中,资源丰富的领域被称为外领域(out-domain),资源稀缺的领域被称为内领域(in-domain)。

在机器翻译领域适应中,有监督的领域适应要求内领域和外领域都具有平行语料,然而在电子商务领域,不存在公开的大规模平行数据,并且构建电子商务平行语料是十分困难的,所以采用无监督领域适应的方法,避免了对内领域平行语料的需求。目前无监督领域适应中,提升模型翻译效果主要有两种方式。第一种方法是生成内领域的伪平行数据。最为显著的方法是 2015 年 Sennrich 等人<sup>[8]</sup>通过将内领域目标语言反向翻译来构建伪平行数据。反向翻译(back-translation)是给定目标语言句子  $y$ , 用训练好的目标语言到源语言的翻译模型得到伪句子对  $(x', y)$ , 训练过程始终保持目标端的真实性。此外,2017 年 Anna Currey 等人<sup>[9]</sup>复制内领域目标语言文本的方法都是不断更新扩充领域内的数据。2019 年 Hu 等人<sup>[10]</sup>通过词库归纳构建域内伪平行语料库,大大提高了翻译质量。第二种方法是在现有的网络基础上设计出新的网络模型,Xia 等人<sup>[11]</sup>在利用伪平行的数据的基础上,提出双向学习的方法提升模型的性能。2018 年 Zhen Yang 等人<sup>[12]</sup>在翻译模型中引入对抗分类器的方法实现无监督的领域适应。2019 年 Dou 等人<sup>[13]</sup>提出将领域特征嵌入到网络中编码端的无监督领域适应方法,并通过多任务学习训练了整个网络。

为了解决内领域平行资源稀缺的问题,该文使用无监督领域适应,并在迭代训练的过程中,通过数据逐步混合训练策略提升翻译性能,充分利用单语数据。同时针对于无监督领域适应过程中词对匹配不佳的问题,提出短语后验证的方法进一步强化。通过大量实验对比,在电子商务领域上该方法超于现有的方法,较最强基线系统高出约 1.5 Bleu 点。

该文的贡献包含以下几点:

(1) 获取电子商务单语数据,以无监督领域适应方法搭建电子商务机器翻译平台;

(2) 无监督领域适应电商机器翻译系统采用混合策略为主体框架,充分利用内领域数据,探索内外领域混合比例,提升翻译性能;

(3) 提出短语后验证的方法解决无监督领域适应机器翻译中词对匹配不佳问题,同时最大化利用内领域数据。

首先介绍了近年在领域适应机器翻译方面的相关研究以及工作,其次详细描述无监督领域适应电商机器翻译的主体框架以及短语后验证方法,再次介绍实验设置和实验结果,并对相关实验进行详细的分析,最

后给出结论与未来工作计划。

## 1 短语后验证的无监督领域适应

### 1.1 混合策略的主体框架

短语后验证的无监督领域适应电商机器翻译是在混合策略基础上实现的,它是训练双向翻译过程,整个框架见图 1。从源端到目标端系统,目标端到源端系统,记做  $\text{src} \rightarrow \text{tgt}$  和  $\text{tgt} \rightarrow \text{src}$ 。

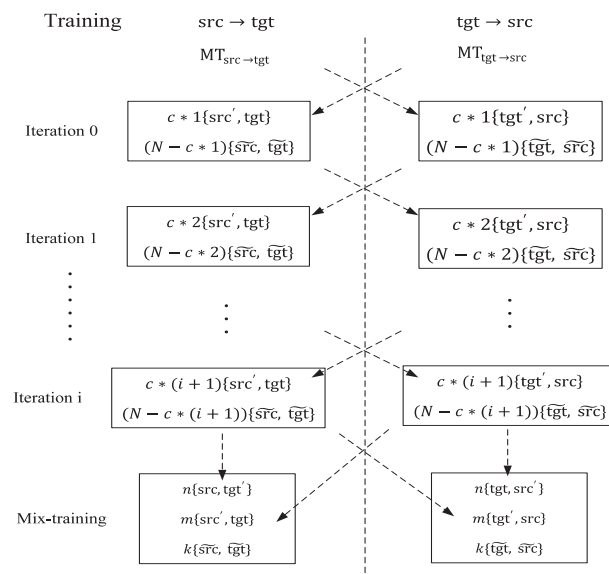


图 1 基于混合策略的无监督领域适应框架结构

首先由外领域翻译模型获得内领域的初始伪平行数据,如图 1 中的  $\{\text{src}', \text{tgt}\}$  和  $\{\text{tgt}', \text{src}\}$ 。该过程是神经机器翻译模型,它是一种非线性结构,其经典的架构是编码器—解码器(encoder—decoder)。编码器首先将源端语句  $X = \{x_1, x_2, \dots, x_n\}$  输入转换成  $E_{\text{encoder}} = \{e_1, e_2, \dots, e_n\}$  编码,然后将其编码成隐藏状态  $h$ , 并发送到解码器 decoder。解码器类似,将目标端语言输入  $Y = \{y_1, y_2, \dots, y_m\}$  转换成  $E_{\text{decoder}} = \{e_1, e_2, \dots, e_m\}$ , 然后给定输入历史和隐藏状态  $h$ , 生成  $Z$ , 具体公式如下:

$$P(Y|X; \theta) = \prod_i P(y_i | z \leq i, h; \theta) \quad (1)$$

其中,  $\theta$  是神经模型参数,  $z \leq i$  是历史输入。训练过程的损失如公式(2):

$$L(\theta) = - \sum_i \log_p(y_i | x, y_{<i}; \theta) \quad (2)$$

然后将内领域和外领域数据混合,外领域数据在图中表示为  $\{\widetilde{\text{src}}, \widetilde{\text{tgt}}\}$  和  $\{\widetilde{\text{tgt}}, \widetilde{\text{src}}\}$ , 其中数据总量为  $N$ , 内领域保持数据总量不变, 每轮训练迭代增加  $c * i$  数量伪平行数据  $\{\text{src}', \text{tgt}\}$  和  $\{\text{tgt}', \text{src}\}$ ,  $c$  为常数,  $i$  为训练迭代数, 更新模型, 重复这个过程直到收敛。得到最佳模型后将内领域数据与外领域数据混合训练, 如图 1 中所示, 训练  $\text{src} \rightarrow \text{tgt}$  方向模型, 通过混

合  $\{\text{src}', \text{tgt}'\}$ 、 $\{\text{src}, \text{tgt}'\}$  以及  $\{\widetilde{\text{src}}, \widetilde{\text{tgt}}\}$  三者数据训练更新模型,其中  $m$ 、 $n$ 、 $k$  三者为该数据的比例,  $\text{tgt} \rightarrow \text{src}$  方向模型训练亦同。

## 1.2 短语后验证

混合策略的无监督领域适应在训练过程中通过数据混合充分利用了内领域单语数据以及外领域的平行语料,大大地提升了翻译效果。但反向翻译产生的伪平行语料质量难以保证,所以在混合策略训练的前期始终保持目标语言的数据是真实的,避免因使用低质量的伪数据作为目标端,导致在神经网络训练过程中翻译模型的偏离。但由外领域训练产生的词对存在匹配不佳的问题,通过引入短语后验证方法解决该问题。

整个方法的实现过程见图 2。该方法将前向翻译(forward-translation)联合反向翻译(back-translation)训练出新的短语表,更具体地使用 fasttext<sup>[14]</sup> 抽取图中反向翻译数据  $\{\text{tgt}', \text{src}'\}$  对齐文件,根据该对齐文件抽出短语表  $W = \{(w'_{s1}, w_{t1}), (w'_{s2}, w_{t2}), \dots, (w'_{sq}, w_{tq})\}$ , 同样  $\{\text{tgt}, \text{src}'\} \cup \{\text{tgt}', \text{src}'\}$  为联合数据,其产生短语表  $W^* = \{(w'_{s1}, w_{t1}), (w'_{s2}, w_{t2}), (w'_{s3}, w_{t3}), \dots, (w'_{sq}, w_{tq}), \dots, (w'_{sk}, w_{tk})\}$ ,  $W^*$  中产生新的词对  $(w'_{s3}, w_{t3})$  等,并将原词对  $(w'_{sq}, w_{tq})$  更新为  $(w'_{sq}, w_{tq})$ , 然后使用这个新的短语表抽取出新的词对替换原伪语句对的对齐短语,构成新的伪语料进行训练。

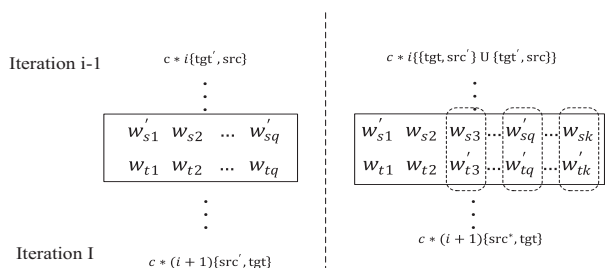


图 2 前向翻译验证示意图

## 1.3 综合训练

在混合训练策略的无监督领域适应框架结构基础上加入前向翻译验证的方法,由于训练开始产生的前向翻译数据质量较低,为了保证训练的新验证短语质量,每次迭代训练后,增加  $c * i$  数量前向翻译与反向

翻译学习校验原句中源端数据的词进行替换,从而产生新的伪语料继续训练,重复迭代训练与前向翻译验证过程直至收敛。

## 2 实验设置及过程

### 2.1 实验设置

该文所有实验采用开源代码 Fairseq<sup>[15]</sup>,将模型设置为 Transformer<sup>[16]</sup>,dropout 设置为 0.3,编码器和解码器层数均为 6 层,其他基本的超参数设置为 Fairseq 中的默认参数选项。解码时,采用集束搜索 (beam search),其中 beam\_size 设置为 5,其余参数采用默认设置。训练和测试均基于 GTX1080Ti。采用双语互译评估 (bilingual evaluation understudy, BLEU)<sup>[17]</sup>,它已经被广泛应用于自然语言处理和机器翻译任务。该文采用 BLEU 自动化脚本 multi-bleu.perl。

### 2.2 实验数据集及预处理

在新闻领域,平行数据非常丰富,因此通常被用作域外领域数据。选择 linguistic data consortium (LDC) 作为领域外的数据集,这是一个中英新闻语料,大约有 125 万个平行句。领域内的数据是本节中获取的电子商务领域数据。

目前,电子商务领域不存在公开的语料,为了构建电子商务翻译系统,在此之前首先构建一份电子商务语料。这是一份包含英文和中文的单语电商语料库,分别通过中英电商平台获取产品文本信息。数据包含四类:男士服装、女士服装、食物和玩具。笔者构建了训练集为单语数据,平行数据分别为测试集和验证集,数据的详细数量见表 1。表中数字为数据集的句子数量,其中训练集是以百万 (M) 为单位。

对训练数据进行了预处理,去掉网络中特殊字符,并且针对中文,由于字符之间没有明显的分割符号,一般采用中文分词工具进行分词,这里使用 jieba 分词工具;针对英文,可以依据单词之间的空格进行切分,同时进行词串识别 (tokenization) 处理。为了去除分词工具在模型评估中的影响,在测试过程中,在英文到中文方向上,以字级别文本文件作为参考译文。

表 1 电商领域数据统计

		男士服装	女士服装	食物	玩具	总计
训练集	中文	0.70M	0.90M	0.59M	0.31M	2.50M
	英文	0.62M	0.67M	0.70M	0.64M	2.63M
验证集	中文	587	731	780	429	2 527
	英文	591	388	462	391	1 832
测试集	中文	889	1 905	1 603	911	5 308
	英文	747	653	358	571	2 329

### 2.2.1 数据 bpe 设置

在神经机器翻译的训练中,使用了 (byte pair encoder, BPE)<sup>[18]</sup> 技术处理所有的数据,进行了词表的压缩。使用所有的数据训练得到 bpe 编码,并且分别对不同大小的 bpe 进行了实验,对源端与目标端分别做 bpe 和联合源端与目标端做 bpe 进行了实验。在外领域数据上训练得到的内领域四类测试数据结果如表 2 所示,英—中和中—英方向均是 64K bpe 效果最佳,最终决定中英分别使用 64K bpe 编码。

表 2 不同 bpe 设置实验对比

		男士服装	女士服装	食物	玩具	平均值
英—中	32K bpe	10.64	10.55	15.74	12.43	12.34
	64K bpe	11.16	10.77	15.76	13.30	12.75
	联合 64K bpe	11.11	10.44	15.32	13.40	12.57
中—英	32K bpe	6.68	6.04	13.11	11.92	9.44
	64K bpe	7.34	7.00	12.66	12.02	9.76
	联合 64K bpe	6.98	6.78	13.17	11.26	9.55

表 3 不同的数据比例实验结果

	混合比率	男士服装	女士服装	食物	玩具	平均值
英—中	3 : 1	12.11	10.18	15.96	13.79	13.01
	2 : 1	11.52	10.61	16.15	13.88	13.04
	1 : 1	12.87	11.28	17.49	14.76	14.10
中—英	3 : 1	11.17	9.43	13.73	16.72	12.76
	2 : 1	12.01	9.76	17.86	19.45	14.77
	1 : 1	11.94	10.87	18.92	21.27	15.75

## 2.3 实验过程

基于上述的实验设置,采用中英分别 64K bpe,内领域与外领域数据 1 : 1 的设置,并且数据总量保持不变。分别训练了混合策略的无监督领域适应电商机器翻译以及短语后验证方法,此外在电子商务领域数据上分别采用了下面基准实验进行性能的对比,实验具体结果见表 4。

### 2.3.1 基准系统

为了更好地探索短语后验证的无监督在电商上的效果,利用以下多个无监督基准系统:

外领域 nmt:在外领域平行语料上使用 fairseq 训练神经网络机器翻译系统。

nmt 反向翻译:由外领域 nmt 系统进行反向翻译,联合外领域数据构成伪平行语料,训练神经翻译模型。

nmt 反向翻译+前向翻译:在 nmt 反向翻译的基础上,前期融合前向翻译数据,联合外领域数据构成伪平行语料,训练神经翻译模型。

目标端复制:在内领域,目标语言句子被直接复制到源语言,以构建伪平行语料。联合外领域平行数据,

### 2.2.2 数据比例设置

为了探寻基于混合策略的无监督训练过程中,内领域和外领域数据在不同混合比例下的效果,在整个训练前进行了不同比例的多组实验。通过外领域平行句与内领域伪平行句不同比例组合,该伪语料是获取的反向翻译。实验结果如表 3 所示。表中的混合比率表示域内数据和域外数据的比率。根据结果,当外部数据和内部数据以 1 : 1 混合时,与 Sennrich 等人<sup>[19]</sup>得到的结论一致。

训练新的神经翻译模型。

DAFE 反向翻译:在内领域伪平行语料库和外领域平行数据的组合上训练神经翻译系统,在编码器的每一层添加领域感知函数嵌入和任务特定函数嵌入。

无监督词归纳领域适应:通过所有领域单语数据训练一个词嵌入,使用最近邻搜索获得归纳词,根据该词表构建内领域伪平行数据。然后,将内领域伪数据和真实外领域数据结合起来训练新模型。

监督词归纳领域适应:在无监督词归纳的基础上用外领域词典作为种子词典来进行有监督的词汇归纳。

### 2.3.2 短语后验证的无监督领域适应

混合训练:根据混合策略,先由外领域 nmt 模型获得初始数据,在迭代中逐步混合内领域和外领域数据,其中常数  $c$  为 50k,迭代训练  $i$  经过 6 轮,最后混合前向翻译训练更新得到最佳模型,  $m : n : k = 1 : 1 : 1$ ,整个训练过程数据总量保持不变。

混合训练+短语后验证:在混合训练的每轮更新模型后,加上短语后验证从而产生新的伪平行句对,代



入继续训练。重复这个过程直至收敛。

表 4 电子商务产品翻译测试 BLEU 值

		男士服装	女士服装	食物	玩具	平均
英—中	外领域 nmt	11.16	10.77	15.76	13.30	12.75
	nmt 反向翻译	12.87	11.28	17.49	14.76	14.10
	nmt 反向翻译+前向翻译	11.67	10.97	17.13	15.38	13.79
	目标端复制	13.63	15.87	17.78	18.19	16.37
	DAFE 反向翻译	12.74	11.23	17.14	14.92	14.01
	无监督词归纳领域适应	16.16	12.79	19.88	16.47	16.33
	监督词归纳领域适应	15.59	13.54	21.81	17.28	17.06
	混合训练	15.84	13.62	20.47	17.59	16.88
	混合训练+短语后验证	16.94	13.80	22.63	21.99	18.84
	外领域 nmt	7.34	7.00	12.66	12.02	9.76
中—英	nmt 反向翻译	11.94	10.87	18.92	21.27	15.75
	nmt 反向翻译+前向翻译	8.27	7.40	13.94	13.66	10.82
	目标端复制	12.48	16.02	18.27	19.42	16.55
	DAFE 反向翻译	12.57	13.45	19.28	17.46	15.69
	无监督词归纳领域适应	11.96	14.15	17.25	22.17	16.38
	监督词归纳领域适应	14.40	15.57	18.72	17.08	16.44
	混合训练	12.35	14.66	21.51	19.40	16.98
	混合训练+短语后验证	16.17	15.89	20.09	20.21	18.09

### 3 实验分析

#### 3.1 基准系统

通过基准实验系统结果表明, nmt 反向翻译以及 DAFE 反向翻译高于外领域 nmt 系统, 这表明内领域数据构建伪平行数据能够提升翻译性能。此外 nmt 反向翻译+前向翻译较外领域 nmt 系统有提升, 但由于前向翻译质量影响, 在效果上低于反向翻译方法。

词归纳通过归纳词构建初始数据进行训练和通过目标端复制构建伪平行语料进行训练, 结果表现出较强的基准, 这是因为电商数据包含大量的低频词, 并且在相关产品的描述时, 会存在短语的堆叠现象。例如: “尺码: m x xl xxl”, “品牌: adidas”, 这体现了词匹配优化在电子商务中是具有一定效果。

#### 3.2 短语后验证 vs. 基准系统

混合训练充分利用内领域与外领域数据, 采用渐进增加方法, 避开了前期的前向翻译质量较低问题, 从而使得训练能够达到基准系统。通过加入短语的后验证, 在此基础上, 不仅最大化利用了内领域数据, 同时解决了无监督领域适应的词匹配问题, 在中—英和英—中方向超过最强基线分别为 1.54 Bleu 点和 1.78 Bleu 点。

此外为了更好地验证混合训练与短语后验证方法对数据质量的提升, 在测试集进行单词匹配验证。通

过 fasttext<sup>[14]</sup> 对测试集及参考译文构建出内领域词对, 在构建内领域词对过程中, 若抽取的该领域内的词对在外领域中存在则去除该词对, 剩余的对齐词集合构成新词对的参考集合。

根据该参考集合, 在验证集上分别测出其精确率、召回率和 F 值。其过程与构建内领域过程相同, 分别对最强的基线系统、混合训练+短语后验证方法进行新词对集合提取。

表 5 中显示各方法验证集的单词配对三个值。其中混合训练、混合训练+短语后验证方法在中—英和英—中两个方向的召回率远超于基准系统的召回率。而在精确率上英—中的基准系统偏高, 中—英几种模型之间差距较小。根据精确率以及召回率, 综合指标 F 值三者参考各方法产生的词对质量, 表明文中方法能够改善词对质量, 提升翻译性能。

表 5 验证集的单词配对评分

		精确率	召回率	F 值
英—中	监督词归纳领域适应	58.61	50.22	54.09
	混合训练	56.18	63.27	59.51
	混合训练+短语后验证	56.70	65.44	60.76
	目标端复制	76.17	70.01	72.96
中—英	混合训练	75.87	79.41	77.60
	混合训练+短语后验证	76.75	80.53	78.59

## 4 结束语

在电子商务机器翻译中,资源十分稀缺。获取大量电子商务数据,通过一系列预处理构建了电子商务中英单语语料库。在此语料库上构建短语后验证的无监督领域适应电商机器翻译,通过无监督领域适应避免了对电商平行资源的需求。并且该方法在混合策略的无监督领域基础上,充分利用内领域与外领域数据,将短语后验证方法融入该训练,解决无监督领域适应的词对匹配不佳问题,使得电商产品信息翻译模型性能够得到进一步的提升。在未来将拓展更多的语言的电商产品数据信息,进行进一步的研究与创新。

### 参考文献:

- [1] 戴新宇,尹存燕,陈家骏,等. 机器翻译研究现状与展望[J]. 计算机科学,2004,31(11):176-179.
- [2] 高明虎,于志强. 神经机器翻译综述[J]. 云南民族大学学报:自然科学版,2019,28(1):72-76.
- [3] 李亚超,熊德意,张 民. 神经机器翻译综述[J]. 计算机学报,2018,41(12):2734-2755.
- [4] 刘 洋. 神经机器翻译前沿进展[J]. 计算机研究与发展,2017,54(6):1144-1149.
- [5] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. arXiv:1409.0473,2014.
- [6] BROWN P F, PIETRA S A D, PIETRA V D J, et al. The mathematics of statistical machine translation: parameter estimation[J]. Computational Linguistics, 1993, 19(2):263-311.
- [7] WANG Rui, ZHAO Hai, LU Baoliang, et al. Connecting phrase based statistical machine translation adaptation[C]//26th international conference on computational linguistics. Osaka, Japan: The COLING Organizing Committee, 2016.
- [8] SENNRICH R, HADDOW B, BIRCH A. Improving neural machine translation models with monolingual data[J]. arXiv:1511.06709,2016.
- [9] CURREY A, BARONE A V M, HEAFIELD K. Copied monolingual data improves low-resource neural machine translation[C]//Proceedings of the second conference on machine translation. Denmark: Copenhagen, 2017.
- [10] HU Junjie, XIA Mengzhou, NEUBIG G, et al. Domain adaptation of neural machine translation by lexicon induction[C]//Proceedings of the 57th conference of the association for computational linguistics. Italy: Association for Computational Linguistics, 2019:2989-3001.
- [11] XIA Yingce, HE Di, QIN Tao, et al. Dual learning for machine translation[J]. arXiv:1611.00179,2016.
- [12] ZHEN Yang, WEI Chen, FENG Wang, et al. Unsupervised domain adaptation for neural machine translation[C]//2018 24th international conference on pattern recognition (ICPR). China, Beijing: IEEE, 2018:338-343.
- [13] DOU Ziyi, HU Junjie, ANASTASOPOULOS A, et al. Unsupervised domain adaptation for neural machine translation with domain-aware feature embeddings[C]//Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing. Hong Kong, China: Association for Computational Linguistics, 2019:1417-1422.
- [14] GRAVE E, JOULIN A, BERTHET Q. Unsupervised alignment of embeddings with wasserstein procrustes[C]//22nd international conference on artificial intelligence and statistics. Japan: PMLR, 2019:1880-1890.
- [15] OTT M, EDUNOV S, BAEVSKI A, et al. Fairseq: a fast, extensible toolkit for sequence modeling[C]//Proceedings of the 2019 conference of the North American Chapter of the association for computational linguistics. Minneapolis: Association for Computational Linguistics, 2019:48-53.
- [16] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st international conference on neural information processing systems. Long Beach, California, USA: Curran Associates Inc, 2017:6000-6010.
- [17] PAPINENI K, ROUKOS S, WARD T, et al. Bleu: a method for automatic evaluation of machine translation[C]//Proceedings of the 40th annual meeting of the association for computational linguistics. Philadelphia: ACL, 2002:311-318.
- [18] SENNRICH R, HADDOW B, BIRCH A. Neural machine translation of rare words with subword units[C]//Proceedings of the 54th annual meeting of the association for computational linguistics. Berlin: Association for Computer Linguistics, 2016:1715-1725.
- [19] SENNRICH R, HADDOW B, BIRCH A. Improving neural machine translation models with monolingual data[C]//Proceedings of the 54th annual meeting of the association for computational linguistics. Berlin: Association for Computational Linguistics, 2016:86-96.