赵忠超,唐忠,谢京天,等. 基于内容词融合的神经机器翻译方法[J]. 智能计算机与应用,2024,14(12):157-162. DOI:10.20169/j.issn.2095-2163.241222

基于内容词融合的神经机器翻译方法

赵忠超[1], 唐 忠[1], 谢京天[1], 李付学[2], 闫 红[2]
(1 沈阳化工大学 计算机科学与技术学院, 沈阳 110142; 2 营口理工学院 电气工程学院, 辽宁 营口 115014)

摘 要: 基于 Transformer 的神经机器翻译模型是目前主流的机器翻译范式,达到了最先进的性能水平。然而,这种模型无法捕捉单词在句子语义中的重要性,如内容词(实词)比功能词(虚词)表达的更重要,进而会导致翻译错误或歧义。为了解决这个问题,提出了一种基于内容词融合的方法来改进模型。首先,根据内容词识别算法将句子中的单词分为内容词和功能词;然后,利用不同的融合策略将源语言句子中的内容词嵌入到模型中,指导模型翻译过程。在多个翻译任务上的实验证明,基于内容词融合的方法优于基线模型,提升了模型的翻译性能。

关键词: 神经机器翻译; 词嵌入; 内容词; Transformer; 融合策略
中图分类号: TP391.1 **文献标志码**: A **文章编号**: 2095-2163(2024)12-0157-06

A content word fusion method for neural machine translation

ZHAO Zhongchao[1], TANG Zhong[1], XIE Jingtian[1], LI Fuxue[2], YAN Hong[2]
(1 Institute for Computer Sciences and Technology, Shenyang University of Chemical Technology, Shenyang 110142, China; 2 College of Electrical Engineering, Yingkou Institute of Technology, Yingkou 115014, Liaoning, China)

Abstract: The Transformer-based neural machine translation model is currently the prevailing paradigm in the research field, achieving state-of-the-art performance. However, these models fails to capture the importance of words in the semantic context of a sentence, such as content words being more crucial than function words. This limitation can lead to translation errors or ambiguities. To alleviate this issue, the paper proposes an approach called content-word fusion to enhance the model's performance. Firstly, the paper utilizes a content word identification algorithm to categorize words in a sentence into content words and function words. Next, the paper employs various fusion strategies to incorporate the content words from the source language sentence into the model, thereby providing guidance for the translation process. Experimental results from multiple translation tasks demonstrate that the content-word fusion method outperforms the baseline models, significantly improving translation performance.

Key words: neural machine translation; embedding; content word; Transformer; fusion strategy

0 引 言

近年来,随着深度学习技术的发展,基于神经网络的机器翻译模型(Neural Machine Translation, NMT)已经逐渐取代了传统的统计机器翻译模型(Statistical Machine Translation, SMT),成为机器翻译领域的主流范式[1-2]。而在众多神经机器翻译方法中,基于自注意力机制的 Transformer 模型[2]在多个翻译任务表现出优异的性能,在学术界和工业界被广泛研究和使用。该模型由编码器和解码器两部分组成。其中,编码器将输入数据编码成固定长度的向量表示,而解码器则将向量表示翻译成目标语言文本。通过自注意力机制,Transformer 模型能够更加准确地关注与输入相关的部分,并减少对不相关部分的关注,从而更好地学习到源语言与目标语言的映射关系。然而,该模型采用标准的序列到序列的架构,模型本身没有对先验知识进行显性建模,这也限制了模型性能的进一步提升。为了缓解这一问题,一些研究人员尝试将先验知识融合到神经网络之中。例如,Arthur 等学者[3]借助单词在词典的概率,缓解了罕见词翻译的问题。Wang 等学者[4]使用统计机器翻译模型生成的短语来指导神经机器

基金项目: 辽宁省自然科学基金(2021-YKLH-12, 2022-YKLH-18)。
作者简介: 赵忠超(1999—),男,硕士研究生,主要研究方向:神经机器翻译。
通信作者: 闫 红(1984—),女,副教授,主要研究方向:自然语言处理。Email:yanhong@yku.edu.cn。
收稿日期: 2023-07-09

翻译模型。龚龙超等学者[5]显式地引入源语言句子的句法信息指导模型解码来挖掘深层语言知识,从而提高模型性能。与这些引用额外资源作为先验知识来提高 NMT 模型翻译质量的方法不同,本文提出一种简单而有效的方法,利用源语言句子中存在的内容词来增强编码器,从而进一步提高翻译质量。

1 相关工作

在机器翻译发展的过程中,先验知识一直发挥着不可或缺的重要作用[6]。所谓的先验知识指的是特定领域相关的知识,包括语言学规则、专业术语等。为了提高翻译质量,许多研究人员尝试将先验知识纳入序列到序列模型中,如字符或单词结构[7-9]、短语或块结构[4,10]和句法结构[11-13]。

在模型训练或解码阶段,Jean 等学者[7]提出了一种使用小目标词汇的方法来针对不同训练数据。同时,Mi 等学者[8]和 L'Hostis 等学者[9]通过预测目标词来优化模型以降低计算成本。Weng 等学者[14]提出了一种单词预测方法,在解码器开始时预测目标句子中的所有单词,并将其用作训练过程中的控制机制。这些方法侧重于提高训练效率或减低计算成本。然而,本文提出的方法与上述不同,通过采用源语言句子中存在的内容词作为先验知识,将其纳入编码器中,以改进 Transformer 模型中的编码器。

除此之外,Wang 等学者[15]和韩冬等学者[16]也尝试将先验知识融合到神经机器翻译模型中,研究中分别使用基于神经网络的分类器和统计机器翻译模型生成的单词翻译知识来调整 NMT 模型生成的单词概率。本文提出的内容词融合方法与这些工作相似,旨在提高编码器的学习能力。不同的是,本文采用源语言句子中存在的内容词作为先验知识,而不是源语言单词所对应的单词翻译。

2 Transformer 模型

Transformer 作为当前工业界和学术界主流的神经机器翻译模型,在许多翻译任务中取得了最佳的翻译效果。因此,本文提出的内容词融合方法以 Transformer 模型为研究基础。该模型是基于自注意力机制的神经机器翻译模型,采用编码器-解码器架构。与传统的循环神经网络不同,使用自注意力机制来构建模型。Transformer 模型由编码器和解码器两个部分组成,每个部分分别由 N 个相同的层堆叠而成。每层由自注意力模块和前馈全连接神经网络构成。其中,自注意力模块被称为多头注意力机制,由多个点积注意力机制组成,利用 3 个向量 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 进行计算:

$$Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = S\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{1}$$

其中,S 表示 *Softmax* 函数;$\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}$ 分别表示查询、键和值向量;d_k 表示查询和键向量的维度。

最终,将计算得到的结果进行拼接送入到前馈全连接神经网络中:

$$Head^i = Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}), \text{for } i = 1\cdots h \tag{2}$$

$$AttentionOutput(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V},\boldsymbol{W}) = concat(Head^1, Head^2, \cdots, Head^h)\boldsymbol{W} \tag{3}$$

其中,$\boldsymbol{W}$ 表示组合多头自注意力输出的权重矩阵。

前馈全连接神经网络是一个 2 层的全连接网络层,使用 RELU 作为激活函数。具体地,给定一个输入向量 $\boldsymbol{x}$,输出向量可以表示为:

$$FFN(\boldsymbol{x}) = \max(0, \boldsymbol{x}\boldsymbol{W}_1 + \boldsymbol{b}_1)\boldsymbol{W}_2 + \boldsymbol{b}_2 \tag{4}$$

其中,$\boldsymbol{W}_1, \boldsymbol{W}_2, \boldsymbol{b}_1, \boldsymbol{b}_2$ 表示模型可训练参数。

3 内容词融合方法

基于内容词融合的神经机器翻译方法通过提取源语言句子中存在的内容词作为先验知识融入到 Transformer 模型,指导模型生成更好的译文。

3.1 内容词识别

本文受 Chen 等学者[17]的启发,根据 TF-IDF 算法将句子中的单词分成内容词和功能词。具体来说,给定双语语料库 $D_m = \{S, T\}$,其中,m 表示 D 中句子对的总数,S 和 T 分别表示源语言和目标语言句子。在 TF-IDF 算法中,假设句子 $S_i (S_i \in S)$ 由 n 个单词组成,句子中第 j 个单词 S_i^j 在 S_i 中的得分具体如下:

$$score(S_i^j) = \frac{count(S_i^j)}{len(S_i)} \cdot \log\left(\frac{m}{1 + count\left(S_i^j \text{ in } \sum_i^m S_i\right)}\right) \tag{5}$$

其中,$count\left(S_i^j \text{ in } \sum_i^m S_i\right)$ 表示源语言句子 S 中包含单词 S_i^j 的句子数。

通过 TF-IDF 算法,可以计算句中每个单词的分数,这代表了其在句子中的重要性。值得注意的是,这里选择内容词的标准是与词频相关的统计数据,而不是语言学意义上的标准。也就是将得分较高的单词视为内容词,这种近似方法消除了对额外特定语言的资源的需求。

3.2　内容词融合方法

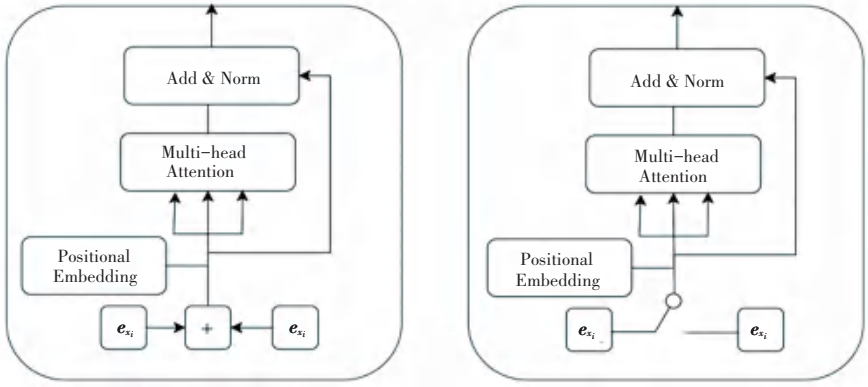
3.2.1　Blend 编码器

本文将内容词作为额外特征直接添加至编码器中的方式来增强翻译模型。具体而言，首先识别和标记源语言句子中的内容词，然后将内容词的词嵌入添加到其原始词嵌入中，作为编码器的输入。该方法可以描述如下：

$$\boldsymbol{e}_i = \boldsymbol{e}_{s_i} + \boldsymbol{e}_{s_i} \tag{6}$$

其中，$\boldsymbol{e}_{s_i}$ 表示源语言句子中存在内容词 s_i 对应的词嵌入，$\boldsymbol{e}_i$ 表示最终输入到编码器中的词嵌入。

融合源语言内容词方法如图 1 (a)可知，本文采用的 Blend 编码器方法是一种简单直接的方式，通过对源语言中存在的内容词进行融合来提高翻译效果。Blend 编码器方法具有 2 个优点。首先，能够更好地捕捉源语言句子中重要单词的信息，从而改善编码器性能；其次，该方法无需引入额外的资源，且计算效率高。



(a) Blend　　(b) Gated

图 1　融合源语言内容词方法

Fig. 1　Method of integrating source language content words

3.2.2　Gated 编码器

Blend 编码器将内容词的词嵌入添加到原始单词的词嵌入中，然而，简单直接相加的操作可能会导致对内容词过度关注。此外，使用 TF-IDF 算法不能保证所选内容词的准确性。因此，本文提出了一种新的融合方法：Gated 编码器。该方法通过添加门控机制，自动调节内容词在编码器中所占的权重，从而更好地保留和利用源语言句子中的重要信息。由图 1(b)可知，输入到编码器中的词向量 $\boldsymbol{e}_i$ 表示为：

$$\boldsymbol{e}_i = \boldsymbol{e}_{s_i} + \delta \circ \boldsymbol{e}_{s_i} \tag{7}$$

其中，“∘”表示元素相乘，δ 表示门控机制，计算如下：

$$\delta = \sigma(\boldsymbol{W}_a \boldsymbol{e}_{s_i} + \boldsymbol{W}_b \boldsymbol{e}_{s_i} + \boldsymbol{c}) \tag{8}$$

其中，$\boldsymbol{W}_a \in R^{d\times d}$，$\boldsymbol{W}_b \in R^{d\times d}$，$\boldsymbol{c} \in R^{1\times d}$ 为模型待训练的参数；d 表示词向量大小；σ 表示 Sigmoid 激活函数。

Blend 编码器采用简单直接的方式，将内容词的信息融入到翻译模型之中，虽然使模型在翻译的过程中考虑到了内容词的信息，但是对于误标记的内容词的信息将无法区分且可能过度关注。Gated 编码器是对 Blend 编码器方法的进一步改进，通过添加门控机制赋予模型自主选择的能力，实现根据具体情况自动选择最佳的模型翻译信息。

4　实验结果与分析

4.1　数据处理与模型设置

4.1.1　数据处理

为验证内容词融合方法性能，本文在 Ai Challenge 中英数据集、IWSLT14 德英平行语料库以及标准稀缺资源 IWSLT15 英越翻译任务和 WMT18 英土翻译任务进行实验。其中，中英数据集从 Ai Challenge 随机筛选 400 万句对组成训练集，验证集使用 Ai Challenge 提供的 8 000 句对，对于测试集则采用 IWSLT15 中英数据集之中的 tst2010、tst2011、tst2012 和 tst2013 合并作为测试集。在 IWSLT15 英越数据集中将 tst2012 作为验证集，tst2013 作为测试集。WMT18 英土翻译任务的数据设置与 Bugliarello 等学者[11]相同。对 IWSLT14 德英任务，本文跟随龚龙超等学者[5]的设置。表 1 记录了实验所使用的语料规模。对于所使用的语料库，需要进行数据预处理步骤，如符号标准化、规范化大小写、数据清洗，

以及使用由 Sennrich 等学者[18]提出的 BPE 算法进行子词切分。其中,对于数据清洗,中文数据使用 Jieba 工具进行分词,对于其他语言而言,则采用 Mosesdecoder 进行分词处理。

表 1 实验使用语料规模统计

Table 1 Statistics of corpus size used in the experiment

语料	训练集	验证集	测试集
AI-CHALLENGE 中英	4M	8 000	5 473
IWSLT15 英越	133 k	1 553	1 268
IWSLT14 德英	160 k	7 283	6 750
WMT18 英土	207 k	3 000	3 007

4.1.2 模型设置

本文使用开源工具 Fairseq[19]实现了将内容词融合到 Transformer 神经机器翻译模型的功能。具体而言,对于中英和英越数据集,采用 Fairseq 提供的标准 Transformer 模型,而在德英和英土数据集上,则采用 Transformer 轻量模型。其中,标准 Transformer 模型的编码器和解码器均包含 6 层,自注意力头的数量为 8 个,词向量维度大小为 512,前馈神经网络层的输出维度为 2 048。而轻量 Transformer 模型的编码器和解码器均包含 6 层,自注意力头的数量为 4 个,词向量维度大小为 512,前馈神经网络层的输出维度为 1 024。此外,实验使用 Adam[20]优化 1 器为模型更新参数,其中,$\beta_1 = 0.9$,$\beta_2 = 0.98$,并使用 RTX3090TI 设备训练实验模型。模型翻译性能使用 *BLEU*[21] 评测,对于其他设置均使用 Vaswani 系统[2]中的默认设置。

4.2 实验结果与分析

4.2.1 多翻译任务表现

本文方法与基线系统在不同数据集上的 *BLEU* 值见表 2。分析表 2 可知,本方法在测试集上表现出较好的翻译性能,具体而言,在中英数据集上,本文提出的方法均优于 Transformer 基线系统,其中 Gated 编码器较基线系统提升了 0.37 点 *BLEU* 值。在英土数据集上,Gated 编码器的提升效果要优于 Blend 编码器,其相对于基线系统提升了 0.32 点 *BLEU* 值。在德英数据集上,Blend 编码器的提升效果要优于 Gated 编码器,其相对于基线系统提升了 0.24 点 BLEU 值。在英越数据集上,Gated 编码器的提升效果要优于 Blend 编码器,其相对于基线系统提升了 0.57 点 BLEU 值。这些实验结果表明,本文提出的增强方法可以显著提升机器翻译系统的性能。值得注意的是,在不同数据集上,2 种不同融合方式的编码器的表现存在差异,特别地,对于中-英、英-土和英-越数据集而言,Gated 编码器的表现更为优秀,而在德英数据集上,Blend 编码器的表现更为出色。这些实验结果表明,本文提出的内容词融合方法可以有效地提高模型的翻译性能。

分析可知,模型翻译性能提高的主要原因包括以下方面:首先,根据 TF-IDF 算法可以识别出在句子中出现频率较高,但在整个语料库中出现频率较低的重要单词,这些单词通常是句子中最具有信息量的部分。其次,将这些单词融入到编码器中,可以使编码器更好地学习这些信息,并更准确地表示输入句子的含义。

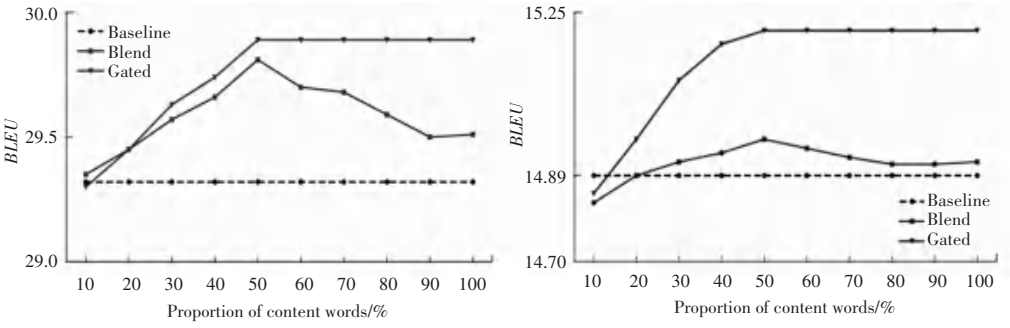
表 2 本文方法与基线系统在不同数据集上的 *BLEU* 值

Table 2 Performance on several translation tasks with the *BLEU* metric

方法	汉语-英语	英语-越南语	德语-英语	英语-土耳其语
基线系统	19.01	29.32	34.43	14.89
Blend 编码器	19.15	29.81	**34.67**	14.97
Gated 编码器	**19.38**	**29.89**	34.55	**15.21**

4.2.2 内容词比例

内容词的选择是通过 TF-IDF 算法得到的。显然易见的是,模型的性能可能会受到内容词数量的影响。因此,通过改变源语言句子中存在的内容词比例,验证内容词数量对模型性能的影响。实验结果如图 2 所示。图 2 中 Blend 编码器策略的表现随着句子中的内容词的覆盖率的不断上升,*BLEU* 值也不断上升,但是在 50%这里达到最大值,之后随着覆盖率的上升而出现下降。而对于 Gated 编码器策略,内容词的覆盖率从 10%增加到 50%时,*BLEU* 分数均有提高。然而,内容词的比例超过 50%,NMT 模型的 *BLEU* 得分几乎没有变化,保持在同一水平。分析后可知,门控机制控制了内容词集成到编码器中的信息量,50%的内容词足以使 NMT 模型表现良好,而 100%的内容词无法进一步提高模型的性能。因此,在其他实验中,Blend 和 Gated 编码器策略的内容词百分比设置为 50%。



（a）英-越数据集　　（b）英-土数据集

图 2　不同比例的内容词在英-越数据集和英-土数据集的表现

Fig. 2　Performance of the NMT model by varying proportion of content words on En-Vi and En-Tr translation tasks

4.2.3　对比其他方法

本文方法与韩冬等学者[16]研究类似，都将单词级的知识纳入 Transformer 模型。然而，内容词融合方法与韩冬等学者[16]提出的方法有所不同，本文采用源语言句子中存在的内容词作为先验知识，而不是源语言单词所对应的目标翻译单词。内容词融合方法与韩冬等学者[16]提出的方法在 IWSLT15 英越数据集的表现见表 3。由表 3 可知，内容词融合方法优于韩冬等学者[16]所提出的方法，Gated 优于 Factored 编码器，证明了内容词融合方法具有一定的优越性。

表 3　对比其他方法

Table 3　Comparison with other methods

方法	英语-越南语
Baseline	29.32
Factored 编码器	29.69
Blend 编码器	29.81
Gated 编码器	**29.89**

此外，为进一步验证内容词融合方法的优越性，本文在 WMT18 英土数据集上与已有的先验知识相关的工作进行比较。其中，包括：Bugliarello 等学者[11]提出的句法增强方法 PASCAl（Parent-Scaled Self-Attention）以及参数优化的 Multi-Task；将句法依赖标签加入到 Transformer 编码器 S&H（Sennrich 等学者[12]）；将自注意力与句法解析相融合的 LISA（Linguistically-Informed Self-Attention[13]）。实验结果见表 4。由表 4 可知，本文提出的内容词融合方法均优于同样使用先验知识来增强模型的方法，对于 WMT18 En-Tr 数据集，与表 4 中最好的方法（PASCAL）相比，内容词融合方法提升 1.21 点 *BLEU*。由此，可以看出，内容词融合方法与之前的方法对比，仍然具备更佳的模型性能，进一步证明本文方法的有效性。

表 4　对比其他方法

Table 4　Comparison with other methods

方法	英语-土耳其语
Multi-Task	14.00
S&H	13.00
LISA	13.60
PASCAL	14.00
Gated	**15.21**

5　结束语

本文提出了一种简单有效的方法来增强基于 Transformer 的神经机器翻译模型的编码器。利用 TF-IDF 算法识别源语言句子中存在的内容词，通过 Blend 编码器和 Gated 编码器两种方式将内容词纳入编码器中，使其学习到额外的信息。在多个翻译任务上获得的实验结果证明了方法的有效性，与当前最佳的 Transformer 模型相比也取得了较好的性能表现。

在下一步工作中，将探索以下方向：通过 TF-IDF 算法区分的内容词，是否还有其他更优的分类方法；对于内容词融合到编码器中的方法，是否有更加高效的方式；以及，利用短语或块结构是否可以增强编码器。

参考文献

[1] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[C]//Proceedings of the 6th International Conference on Learning Representations. Piscataway, NJ: IEEE, 2015:1-15.

[2] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Advances in Neural Information Processing Systems. Long Beach, USA: NIPS Foundation, 2017, 30: 5998-6008.

[3] ARTHUR P, NEUBIG G, NAKAMURA S. Incorporating discrete translation lexicons into neural machine Translation [C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. ACL, 2016: 1557-1567.

[4] WANG Xing, TU Zhaopeng, XIONG Deyi, et al. Translating phrases in neural machine translation [C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. ACL, 2017:1421-1431.

[5] 龚龙超,郭军军,余正涛. 基于源语言句法增强解码的神经机器翻译方法[J]. 计算机应用, 2022,42(11): 3386-3394.

[6] 郭望皓,范江威,张克亮. 融合语言学知识的神经机器翻译研究进展[J]. 计算机科学与探索, 2021,15(7): 1183-1194.

[7] JEAN S, CHO K, MEMISEVIC R, et al. On using very large target vocabulary for neural machine translation[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. ACL, 2015: 1-10.

[8] MI Haitao, WANG Zhiguo, ITTYCHERIAH A. Vocabulary manipulation for neural machine translation[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. ACL, 2016: 124-129.

[9] L'HOSTIS G, GRANGIER D, AULI M. Vocabulary selection strategies for neural machine translation[J]. arXiv preprint arXiv, 1610.00072,2016.

[10] ERIGUCHI A, HASHIMOTO K, TSURUOKA Y. Tree-to-sequence attentional neural machine translation[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. ACL, 2016:823-833.

[11] BUGLIARELLO E, OKAZAKI N. Enhancing machine translation with dependency-aware self-attention[J]. arXiv preprint arXiv, 1909.03149, 2019.

[12] SENNRICH R, HADDOW B. Linguistic input features improve neural machine translation[J]. arXiv preprint arXiv,1606.02892, 2016.

[13] STRUBELL E, VERGA P, ANDOR D, et al. Linguistically-informed self-attention for semantic role labeling [J]. arXiv preprint arXiv,1804.08199, 2018.

[14] WENG Rongxiang, HUANG Shujian, ZHENG Zaixiang, et al. Neural machine translation with word predictions [C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. ACL, 2017: 136-145.

[15] WANG Xing , TU Zhaopeng, ZHANG Min. Incorporating statistical machine translation word knowledge into neural machine translation[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2018, 26:2255-2266.

[16] 韩冬,李军辉,周国栋. 融合单词翻译的神经机器翻译[J]. 中文信息学报, 2019,33(7): 40-45.

[17] CHEN K, WANG R, UTIYAMA M, et al. Content word aware neural machine translation[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. ACL, 2020: 358-364.

[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. ACL, 2015: 1715-1725.

[19] 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. ACL, 2019: 48-53.

[20] KINGMA D P, BA J. Adam: A method for stochastic optimization [J]. arXiv preprint arXiv,1412.6980, 2014.

[21] PAPINENI K, ROUKOS S, WARD T, et al. Bleu: A method for automatic evaluation of machine translation[C]// Proceedings of the 40th Annual Meeting on Association for Computational Linguistics(ACL '02). ACL, 2002:311-318