计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.190100048

融合语义角色的神经机器翻译

乔博文　李军辉
苏州大学计算机科学与技术学院　江苏 苏州 215006
(20164227047@stu.suda.edu.cn)

摘　要　近年来,深度学习取得了重大突破,融合深度学习技术的神经机器翻译逐渐取代统计机器翻译,成为学术界主流的机器翻译方法。然而,传统的神经机器翻译将源端句子看作一个词序列,没有考虑句子的隐含语义信息,使得翻译结果与源端语义不一致。为了解决这个问题,一些语言学知识如句法、语义等被相继应用于神经机器翻译,并取得了不错的实验效果。语义角色也可用于表达句子语义信息,在神经机器翻译中具有一定的应用价值。文中提出了两种融合句子语义角色信息的神经机器翻译编码模型,一方面,在句子词序列中添加语义角色标签,标记每段词序列在句子中担当的语义角色,语义角色标签与源端词汇共同构成句子词序列;另一方面,通过构建源端句子的语义角色树,获取每个词在该语义角色树中的位置信息,将其作为特征向量与词向量进行拼接,构成含语义角色信息的词向量。在大规模中-英翻译任务上的实验结果表明,相较基准系统,文中提出的两种方法分别在所有测试集上平均提高了0.9和0.72个BLEU点,在其他评测指标如TER(Translation Edit Rate)和RIBES(Rank-based Intuitive Bilingual Evaluation Score)上也有不同程度的性能提升。进一步的实验分析显示,相较基准系统,文中提出的融合语义角色的神经机器翻译编码模型具有更佳的长句翻译效果和翻译充分性。

关键词:神经机器翻译;语义角色标注;语义特征;编码模型
中图法分类号　TP391

Neural Machine Translation Combining Source Semantic Roles

QIAO Bo-wen and LI Jun-hui
School of Computer Science and Technology,Soochow University,Suzhou,Jiangsu 215006,China

Abstract　With the rapid development of deep learning in recent years,neural machine translation combining deep learning has gradually replaced statistical machine translation and becomes the mainstream machine translation method in the academic circle. However,the traditional neural machine translation regards the source-side sentence as a word sequence and does not take into account the implicit semantic information of sentences,resulting in the inconsistency between the translation results and source-side semantics. To solve this problem,some linguistic knowledges,such as syntax and semantics,are applied to neural machine translation and achieve good experimental results. Semantic roles can also be used to express the semantic information of sentences and have a certain application value in neural machine translation. This paper proposed two neural machine translation encoding models that incorporate semantic role information of sentences. On the one hand,semantic role played by labels are added to the word sequences to mark the semantic role played by each word in the sentence. The semantic role labels and source-side words together constitute the word sequence. On the other hand,by constructing the semantic role tree of source sentences,the position information of each word in the semantic role tree is obtained,which is spliced with the word vector as a feature vector to form a word vector containing semantic role information. Experimental results on large-scale Chinese-English translation show that,compared with the baseline system,the two methods proposed in this paper not only improve 0.9 BLEU points and 0.72 BLEU points on average in all test sets respectively,but also improve performance in other evaluation indexes,such as TER (Translation Edit Rate) and RIBES (Rank-based Intuitive Bilingual Evaluation Score). Further experimental analysis shows that the proposed neural machine translation encoding models combining semantic roles have better translation effect on long sentences and translation adequacy than the baseline system.

Keywords　Neural machine translation,Semantic role labeling,Semantic feature,Encoding model

到稿日期:2019-01-07　返修日期:2019-04-24　本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家自然科学基金(61876120)
This work was supported by the National Natural Science Foundation of China(61876120).
通信作者:李军辉(jhli@suda.edu.cn)

1 引言

近年来，Sutskever 等[1]提出了一种端到端的神经机器翻译(Neural Machine Translation，NMT)模型，在自然语言处理领域取得了很好的效果。在此基础上，Bahdanau 等[2]提出了一种基于注意力机制的 NMT 模型，该模型在某些语言对上的翻译性能超越了传统的统计机器翻译，在机器翻译领域产生了很大的影响。经过不断发展和完善，NMT 已经成为一种常用的机器翻译方法[3]。

在 NMT 的相关研究中，使用语言学知识辅助翻译是一项重要的研究内容，通过引入语言学知识能够显著提高译文质量。语义角色标注[4]是一种浅层语义分析方法，可以表示文本中含有的浅层语义信息。对文本进行语义角色标注后可以生成不同的语义角色标签，其中包含有丰富的语义信息。

传统的 NMT 模型存在着错译、漏译和过译等问题。此外，从语义角色的角度观察，其还存在语义角色在源端和目标端不一致的问题，包括不连续翻译。这里的不连续翻译是指在源端句子中担任某语义角色的片段，在目标端被翻译为两个或多个不连续的片段，这种情况的出现往往代表译文词序出现问题。表 1 列出了一个中-英翻译示例，“给酒店医生”这个片段在源端句子中担任谓词“打”的一个语义角色 A2，从参考译文中可以看到，这个词汇片段本应被译为一个连续的整体“the hotel doctor”，但基准系统却把它翻译为“the hotel”和“the doctor”这两个不连续的片段。

表 1 NMT 中的不连续翻译现象

Table 1 Discontinuous translation phenomenon in NMT

输入	我给酒店医生打电话并得到了初步治疗。
基准系统译文	i called the hotel and received initial treatment of the doctor.
参考译文	I called the hotel doctor and received initial treatment.

为了解决 NMT 中语义角色不连续翻译的问题，本文在 NMT 模型中引入了源端语义角色信息，使得 NMT 编码器中的词表示向量不仅包含词及其上下文信息，还包含词的语义角色信息。具体地，本文提出并比较了两种融合源端语义角色信息的 NMT 编码器。第一种编码器简称横向编码器，将语义角色标签嵌入到句子词汇序列中合适的位置，用于表示句子哪些片段担当了何种语义角色；第二种编码器简称纵向编码器，为源端句子的每个单词制定含有语义角色信息的特征，并将词向量与该语义角色特征向量拼接以获得新的词向量。实验表明，在中-英机器翻译任务上，本文提出的两种融合语义角色的 NMT 编码器都能显著提高翻译性能。

2 相关工作

2.1 语义角色在统计机器翻译中的应用

在统计机器翻译领域，大量的研究工作均表明语义角色信息对机器翻译是有用的[5-12]。Wu 等[5]首次探索了语义角色标注在统计机器翻译中的应用，提出了融合语义角色标注结果的统计机器翻译模型。该模型充分利用语义角色标注以及基于短语的统计机器翻译的优势，建立了一个双层结构。其第一层使用传统的基于短语的统计机器翻译模型，第二层通过使用浅层语义分析器生成语义框架和角色标签，实现了重排序策略。实验表明，该双层模型的 BLEU 值相较基准系统提高了约 0.5。

为了利用语义角色信息来提高统计机器翻译的准确率和流畅性，Liu 等[6]针对树到串统计机器翻译模型，提出了两类语义角色特征用于建立源端语义角色的重排序/删除模型，这些语义特征以及树到串模板都基于一个条件对数线性模型进行参数训练。实验结果表明，通过添加额外的语义角色特征以及改进参数训练模型，BLEU 值得到了有效提升，译文流畅度也比基准系统更好。

为了向串到树统计机器翻译系统中添加语义角色信息，Bazrafshan 等[7]基于规则提取过程提出了两种方法。第一种方法通过使用已被语义角色标注的语料库和增加非终结符集合来学习翻译规则。第二种方法首先提取同步上下文无关文法(Synchro-nous Context-Free Grammar，SCFG)，然后学习一些包含谓词完整语义结构的语义规则，最后添加特征来对这些规则加以区分。实验表明，第一种方法的性能低于基准系统，而第二种方法的 BLEU 值相较基准系统提升了 0.92。

2.2 语言学知识在神经机器翻译中的应用

在神经机器翻译中，尚未发现利用语义角色信息的研究工作。但是，一些其他的语言学知识[13-22]已经在神经机器翻译中得到了广泛应用，有效提高了译文质量。

为了在神经机器翻译中利用语言学知识，Sennrich 等[13]提出了一种使用语言学输入特征来扩展神经机器翻译的方法。该方法向神经机器翻译源端编码器中添加词元、子词标签、形态学特征、词性标签和依赖性标签这五大语言学特征来扩展传统的神经机器翻译系统。实验表明，添加语言学输入特征的神经机器翻译系统在困惑度、BLEU 值和 chr-F3 等评测指标上均有较好的表现。

句法作为众多语言学知识中的一种，在神经机器翻译中也得到了广泛应用。Li 等[14]通过向源端编码器中添加句法标签，提出了 3 种不同的编码器模型，分别是平行 RNN 编码器、层次 RNN 编码器和混合 RNN 编码器。通过应用这些添加了句法信息的编码器模型，改进后的翻译模型在包括 BLEU 值在内的多项评测指标上均有较大提升。

句法树是一种表示句法信息的常用方式，使用句法树能够充分地表示句子的句法结构。Eriguchi 等[15]据此提出了一种树到序列的句法神经机器翻译模型，该模型通过构建一个基于句法树的编码器，使得源端句子中的句法结构信息能够被充分利用。实验表明，改进后的翻译模型在 RIBES 和 BLEU 值上均有不错的表现。

3 基于注意力机制的神经机器翻译

本节主要介绍基于注意力机制的神经机器翻译。如图 1 所示，传统的神经机器翻译包含编码器和解码器两个组成部分[23]。

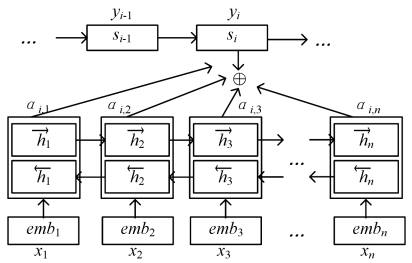


图 1 基于注意力机制的神经机器翻译

Fig. 1 Attention mechanism-based neural machine translation

对于源端句子 $\boldsymbol{X}=\{x_1,x_2,\cdots,x_n\}$（其中 n 表示句子总词数），编码器使用循环神经网络[24]正序处理源端句子词向量得到隐藏状态序列 $\overrightarrow{\boldsymbol{h}}=\{\overrightarrow{h_1},\overrightarrow{h_2},\cdots,\overrightarrow{h_n}\}$，接着逆序处理源端句子词向量得到 $\overleftarrow{\boldsymbol{h}}=\{\overleftarrow{h_1},\overleftarrow{h_2},\cdots,\overleftarrow{h_n}\}$，然后将两个隐藏状态序列拼接起来，获得所有词汇的隐藏层表示，以此作为编码器的输出。解码器使用单向循环神经网络来预测目标句子 $\boldsymbol{y}=\{y_1,y_2,\cdots,y_n\}$，其中每个单词 y_i 需要由当前时刻的隐藏状态 s_i、上一时刻的预测单词 y_{i-1} 和当前时刻的源端上下文向量 $\boldsymbol{c}_i$ 来决定，如式(1)所示：

$$p(y_i\mid y_1,y_2,\cdots,y_{i-1},x)=g(y_{i-1},s_i,\boldsymbol{c}_i) \tag{1}$$

其中，s_i 表示当前时刻的隐藏状态，需要由上一时刻的隐藏状态 s_{i-1}、上一时刻的预测单词 y_{i-1} 和当前的源端上下文向量 $\boldsymbol{c}_i$ 来决定，如式(2)、式(3)所示：

$$s_i=f(s_{i-1},y_{i-1},\boldsymbol{c}_i) \tag{2}$$

$$c_i=\sum_j a_{ij}h_j \tag{3}$$

其中，a_{ij} 为注意力权值，如式(4)所示：

$$a_{ij}=\frac{\exp(e_{ij})}{\sum_{k=1}^{n}\exp(e_{ik})} \tag{4}$$

式(4)中的 e_{ij} 度量了第 i 个输出词与第 j 个输入词之间的匹配程度，如式(5)所示：

$$e_{ij}=a(s_{i-1},h_j) \tag{5}$$

其中，h_j 表示源端句子第 j 个词的隐藏状态。

4 融合语义角色的横向编码器

给定一个源端句子，语义角色标注为该句子中的每一个谓词识别出其语义角色。如表 2 所列，列句有两个谓词“雇用”和“超过”，每个谓词都有多个语义角色。参考 Li 等[14]提出的融合句法信息的混合编码器模型，本文提出在词序列中引入语义角色标签。具体地，假设源端某个片段 w_i^j 表示源端第 i 个词到第 j 个词担当某个谓词的类别为 R 的语义角色，则将该片段的前后分别插入语义角色标签 B_R 和 E_R，分别表示该语义角色的起始位置和终止位置。例如，“新加坡境内”担当谓词“雇用”的 ARG0 语义角色，则其对应的序列为“B_ARG0 新加坡境内 E_ARG0”。此外，考虑到一个句子中可能出现多个谓词，不同谓词的语义角色之间会出现内容左边界或右边界一致的情况。为了方便起见，本文制定以下 3 条规则来减少语义标签的复杂性。

表 2 横向编码器输入句子

Table 2 Horizontal encoder input sentence

例句	新加坡境内雇用的外籍劳工人数超过十万人。
语义角色标注结果	(1)ARG0(新加坡境内) V(雇用) 的外籍 ARG1(劳工人数) 超过十万人。 (2)ARG0(新加坡境内雇用的外籍劳工人数) V(超过) ARG1(十万人)。
横向编码器输入	B_ARG0 B_ARG0 新加坡境内 E_ARG0 B_V 雇用 E_V 的外籍 B_ARG1 劳工人数 E_ARG1 E_ARG0 B_V 超过 E_V B_ARG1 十万人 E_ARG1。

(1)左右边界一致：片段 w_i^j 同时担当两个或更多谓词的语义角色。针对这种情况，仅保留与片段句法关系最近谓词的语义角色。为此，本文选择与该片段距离最近的谓词。

(2)左边界一致：片段 w_i^j 担当某谓词类别为 $A1$ 的语义角色，片段 $w_i^{j'}$ 担当另一谓词类别为 A2 的语义角色，并且 $j<j'$。针对这种情况，在词 w_i 之前依次添加语义角色标签“B_A2 B_A1”，在 w_j 和 $w_{j'}$ 之后分别添加语义角色标签“E_A1”和“E_A2”。

(3)右边界一致：片段 w_i^j 担当某谓词类别为 $A1$ 的语义角色，片段 $w_{i'}^j$ 担当另一谓词类别为 A2 的语义角色，并且 $i<i'$。针对这种情况，在词 w_i 和 $w_{i'}$ 词之前分别添加语义标签“B_A1”和“B_A2”，在 w_j 之后依次添加语义角色标签“E_A2”和“E_A1”。

表 2 最后一行列出了对应例句的融合语义角色标签的输入序列，该序列交替融入了词序列和语义角色标签序列。因此，基于该混合序列的编码器能够学习到新的词表示向量，其中既包含了词及其上下文信息，也包含了词的语义角色信息。

需要注意的是，如图 2 所示，由于语义角色标签（如 B_ARG0 等）本身并不包含词义信息，在编码器的输出序列中，仅词的表示向量用作解码器的输入，即语义角色标签的表示向量不直接用于预测目标端句子。

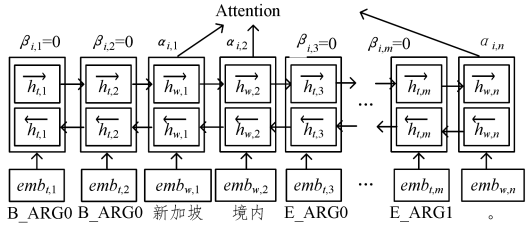


图 2 融合语义角色的横向编码器

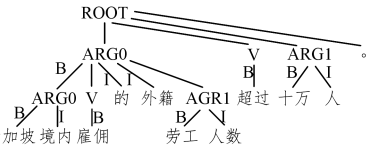
Fig. 2 Horizontal encoder combining semantic roles

5 融合语义角色的纵向编码器

给定一个源端句子，语义角色标注为该句子中的每一个谓词，识别出其语义角色，标注结果可以绘成一棵语义角色树。如表 3 中的语义角色树所示，其中叶子节点表示句子单词，非叶子节点表示语义角色（根节点除外），边 B 或 I 分别表示第一个儿子节点或其他儿子节点。为了获取每个单词表达的语义信息，本文提出语义角色路径特征。具体地，对于词 w，其语义角色路径特征为语义角色树上从根节点 ROOT 到词 w 所经过的节点和边。如表 3 所列，“雇用”这个词汇在语义角色树中的路径为“ROOT_ARG0_I_V_B”。表 3 的最后一行列出了部分单词所对应的语义角色路径特征。

表 3 语义角色路径特征

Table 3 Semantic role path features

例句	新加坡境内雇用的外籍劳工人数超过十万人。
语义角色树	
语义角色路径特征示例	新加坡/ROOT_ARG0_B_ARG0_B 境内/ROOT_ARG0_I_ARG0_I 雇用/ROOT_ARG0_I_V_B,的/ROOT_ARG0_I

如图 3 所示,本文将词向量与路径特征向量进行拼接,并将其作为源端编码器的输入。拼接后的词表示向量既包含词本身的信息,也包含其在语义角色树中的位置信息。

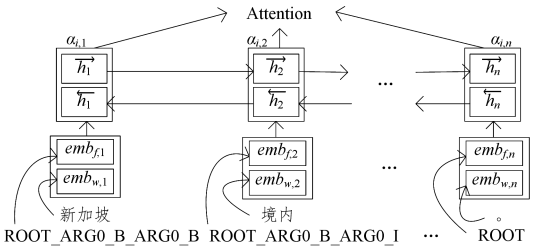


图 3 融合语义角色的纵向编码器

Fig. 3 Vertical encoder combining semantic roles

6 实验

6.1 实验设置

本文使用基于 Theano 深度学习框架实现的 RNNSearch 模型作为基准系统。模型参数使用批随机梯度下降算法进行优化,批大小设定为 80。学习率使用 Adadelta 算法[25]自动调节,算法的衰减常数 $\rho=0.95$,分母常数 $\varepsilon=1\times10^{-6}$。源端和目标端词向量维度都设为 620,隐藏层节点数设为 1000。仅在模型最终输出层使用 dropout 策略,dropout 率设为 0.5。测试时使用集束搜索(Beam Search)算法进行解码,集束宽度为 10。

本文实验采用 NIST MT 评测中-英翻译语料,训练集来自多个 LDC(Linguistic Data Conso-rtium)中-英翻译语料,共包含 124 万句对。源端和目标端词汇表大小均设为 3 万,未登录词统一使用“UNK”替代。使用数据集 NIST06 作为开发集,使用数据集 NIST02, NIST03, NIST04, NIST05, NIST08 作为测试集。评测时,使用大小写不敏感的 4 元 BLEU 值作为评测指标(Moses 开源脚本 mteval-v11b.pl),另外使用其他机器翻译性能评测指标 TER,RIBES 等来辅助评测。为了加快训练速度,源目标端最大句子长度均设置为 50。

为了从源端训练语料和测试语料中生成其语义角色标签,首先使用 Berkeley Parser 工具[26]对源端语料进行句法分析,然后使用中文语义角色标注工具[27]对句法分析结果进行处理,最终得到源端语料的语义角色标签。

在横向编码器模型中,语义角色标签和源端中文词共同构成源端训练语句,由于语义角色标签的类型较少,本文仍将词汇表大小设置为 3 万。在纵向编码器模型中,语义角色路径特征具有单独的词汇表,包含 14770 个词,即训练集中语义角色路径特征的类别数。此外,语义角色路径特征的向量维度设定为 620。

6.2 实验结果

本节使用多个常用的机器翻译评测指标对提出的模型进行评测,从不同的角度来观察模型的性能。

6.2.1 BLEU 值评测结果

本文提出的两种翻译模型和基准系统在不同测试集上的 BLEU 值[28]如表 4 所列。在所有测试集上,本文提出的两种模型较基准系统获均得了更好的翻译性能。相较基准系统,横向和纵向编码器分别在测试集上平均提高了 0.90% 和 0.72%个 BLEU 值,表明本文提出的两种模型能够有效地融合源端句子的语义角色信息。

表 4 BLEU 的评测结果

Table 4 Evaluation results of BLEU

(单位:%)

实验系统	NIST02	NIST03	NIST04	NIST05	NIST08	AVG.
基准系统	38.52	36.26	39.05	35.82	27.37	35.40
横向编码器	38.89	37.02^{+}	40.11^{++}	36.77^{+}	28.71^{++}	36.30
纵向编码器	39.59^{++}	37.05^{+}	39.95^{+}	36.54	27.49	36.12

注:+/++表示在显著性水平为 0.05/0.01 时,本文提出模型相较基准系统有显著提高

6.2.2 翻译编辑率评测结果

翻译编辑率(Translation Edit Rate,TER)是一种常用的机器翻译性能评测指标[29],其通过统计机器译文修改为参考译文的后编辑次数,来分析机器译文的质量。机器译文所需的后编辑次数越少,译文质量就越高。相关研究表明,与 BLEU 值评测方法相比,TER 评测结果更加接近人工评测结果。本文提出的两种模型的 TER 值评测结果如表 5 所列。可以观察到,相较基准系统,横向编码器的 TER 值降低了 0.90%,而纵向编码器的 TER 值则降低了 1.13%,并且在所有测试集上,本文提出的两种模型都具有更好的 TER 值。

表 5 TER 的评测结果

Table 5 Evaluation results of TER

(单位:%)

实验系统	NIST02	NIST03	NIST04	NIST05	NIST08	AVG.
基准系统	58.26	60.60	58.59	61.50	65.28	60.84
横向编码器	57.90	59.64	57.60	60.48	64.12	59.94
纵向编码器	57.52	59.44	57.35	60.36	63.92	59.71

6.2.3 RIBES 评测结果

RIBES(Rank-based Intuitive Bilingual Evaluation Score,RIBES)是另一种评测机器翻译性能的方法[30]。与 BLEU 评测方法不同,RIBES 评测方法更加关注译文的词序是否正确。本文提出的两种模型的 RIBES 评测结果如表 6 所列。

表 6 RIBES 的评测结果

Table 6 Evaluation results of RIBES

(单位:%)

实验系统	NIST02	NIST03	NIST04	NIST05	NIST08	AVG.
基准系统	81.31	80.37	80.90	79.96	75.59	76.62
横向编码器	81.43	80.21	81.16	80.13	76.42	79.87
纵向编码器	81.75	80.59	81.55	80.01	76.88	80.15

可以观察到,相较基准系统,横向编码器的 RIBES 值提高了 3.25%,而纵向编码器的 RIBES 值提高了 3.53%。除

测试集 NIST03 外，本文提出的两种模型都具有更好的 RIBES 值。实验结果表明，相较基准系统，本文提出的两种模型生成的译文具有更好的词序。

6.3 实验分析

除使用以上传统的翻译评测指标衡量翻译结果外，本节从长句性能和漏译两个角度分析基准系统与本文提出的两种模型的性能。

6.3.1 长句性能分析

本文将所有测试集中的句子合并，并按照长度将其分为 6 组，然后评测各组测试句子的 BLEU 值。长句性能评测实验结果如图 4 所示，可以观察到，本文提出的横向编码器模型在所有长度区间内的 BLEU 值均高于基准系统，而本文提出的纵向编码器模型除在长度区间[40,50)内的 BLEU 值稍差之外，在其余长度区间内也高于基准系统。实验表明，本文提出的两种模型不仅在短句翻译上具有不错的性能，在长句翻译上也要优于基准系统。

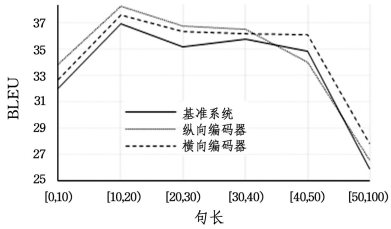


图 4 不同长度区间内的模型性能

Fig. 4 Model performance in different length intervals

6.3.2 语义角色漏译分析

在神经机器翻译中，漏译是一个十分常见的问题。从语义角色的角度来看，漏译指源端语义角色中的部分或者全部词汇没有在目标端译文中被翻译出来。为了分析这个问题，本文从所有测试集句子中随机选取 100 句进行人工分析，统计其中 ARG0(施事者)、ARG1(受事者)、V(谓词)、ARGM-TMP(时间)、ARGM-LOC(地点)以及 ARGM-ADV(状语)这 6 类主要语义角色的漏译情况。据统计，这 100 个句子共包含 1377 个语义角色，其中这 6 类所占比例达92.15%。本文使用式(6)计算单个语义角色的漏译分数，分数越小表示漏译词数越少；使用式(7)计算各类语义角色的漏译比例，比例越低表示模型的漏译现象越少，具有更好的翻译效果。本文提出的两个模型的漏译分析结果如表 7 所列。

$$\text{单个语义角色的漏译分数}=\frac{\text{语义角色中未翻译词数}}{\text{语义角色中总词数}} \tag{6}$$

$$\text{每类语义角色漏译比例}=\frac{\sum\text{每个语义角色的漏译分数}}{\text{每类语义角色的总个数}} \tag{7}$$

表 7 语义角色漏译比例

Table 7 Ratio of semantic roles missing translation

(单位：%)

实验系统	ARG0	ARG1	V	ARGM-TMP	ARGM-LOC	ARGM-ADV	AVG.
基准系统	27.87	18.92	27.37	26.40	21.77	32.59	25.82
横向编码器	17.59	10.17	15.09	18.42	19.09	17.85	16.36
纵向编码器	21.82	14.00	25.43	26.17	16.61	27.32	21.89

可以观察到，相较基准系统，横向编码器的漏译比例降低了 9.46%，而纵向编码器的漏译比例降低了 3.93%，在所有语义角色分类上，本文提出的两种模型都具有更低的漏译比例。

结束语 传统的神经机器翻译系统缺乏语言学知识的指导，使得神经机器翻译中漏译、不连续翻译的问题较为严重。针对上述问题，本文提出两种融合语义角色的神经机器翻译方法，在源端横向编码器和纵向编码器中利用额外的语义角色信息来改进传统的神经机器翻译模型。在 NIST MT 中-英翻译数据集上的实验结果表明，相较基准系统，本文提出的两种方法能有效提高译文质量，具有更好的长句翻译效果，词序也更加规范，在一定程度上解决了传统神经机器翻译出现的语义角色漏译、不连续翻译等问题。

后续我们将继续研究语义角色在机器翻译中的应用，一方面在源端寻求更加富有表现力的语义角色编码器模型，另一方面在目标端尝试将语义角色融入解码器中。

参 考 文 献

[1] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks[C]// Advances in neural information processing systems. Massachusetts: MIT Press, 2014: 3104-3112.

[2] BAHDANAU D, CHO K, BENGIO Y. Neural Machine Translation by Jointly Learning to Align and Translate[C]// Proceedings of the 3rd International Conference on Learning Representations. San Diego, CA, USA: ICLR, 2015: 1-15.

[3] LI Y C, XIONG D Y, ZHANG M. A Survey of Neural Machine Translation[J]. Chinese Journal of Computers, 2018, 41(12): 2734-2755.

[4] GILDEA D, JURAFSKY D. Automatic Labeling of Semantic Roles[C]// Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics. Hong Kong, China: Association for Computational Linguistics, 2000: 512-520.

[5] WU D K, FUNG P. Semantic Roles for SMT: A Hybrid Two-Pass Model[C]// Proceedings of the 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Boulder, Colorado: Association for Computational Linguistics, 2009: 13-16.

[6] LIU D, GILDEA D. Semantic Role Features for Machine Translation[C]// Proceedings of the 23rd International Conference on Computational Linguistics. Association for Computational Linguistics, Beijing, 2010: 716-724.

[7] BAZRAFSHAN M, GILDEA D. Semantic Roles for String to Tree Machine Translation[C]// Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics. Sofia, Bulgaria: Association for Computational Linguistics, 2013: 419-423.

[8] GAO Q, VOGEL S. Corpus Expansion for Statistical Machine Translation with Semantic Role Label Substitution Rules[C]// Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics. Portland, Oregon: Association for Computational Linguistics, 2011: 294-298.

[9] XIONG D Y, ZHANG M, LI H Z. Modeling the Translation of Predicate-Argument Structure for SMT[C]// Proceedings of the 50th Annual Meeting of the Association for Computational Lin-

guistics. Jeju, Republic of Korea: Association for Computational Linguistics, 2012: 902-911.
[10] GAO Q, VOGEL S. Utilizing Target-Side Semantic Role Labels to Assist Hierarchical Phrase-based Machine Translation[C]// Pro-ceedings of SSST-5, Fifth Workshop on Syntax, Semantics and Structure in Statistical Translation. Portland, Oregon, USA: Association for Computational Linguistics, 2011: 107-115.
[11] LI J H, RESNIK P, DAUMÉ H. Modeling Syntactic and Semantic Structures in Hierarchical Phrase-based Translation[C]// Proceedings of the 2013 Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Atlanta, Georgia: Association for Computational Linguistics, 2013: 540-549.
[12] LI J H, MARTON Y, RESNIK P, et al. A Unified Model for Soft Linguistic Reordering Constrains in Statistical Machine Translation[C]// Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Baltimore, Maryland, USA: Association for Computational Linguistics, 2014: 1123-1133.
[13] SENNRICH R, HADDOW B. Linguistic Input Features Improve Neural Machine Translation[C]// Proceedings of the First Conference on Machine Translation. Berlin, Germany: Association for Computational Linguistics, 2016: 83-91.
[14] LI J H, XIONG D Y, TU Z P, et al. Modeling Source Syntax for Neural Machine Translation[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada: Association for Computational Linguistics, 2017: 688-697.
[15] 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. Berlin, Germany: Association for Computational Linguistics, 2016: 823-833.
[16] CHEN H D, HUANG S J, CHIANG D, et al. Improved Neural Machine Translation with a Syntax-Aware Encoder and Decoder [C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada: Association for Computational Linguistics, 2017: 1936-1945.
[17] CHEN K H, WANG R, UTIYAMA M, et al. Neural Machine Translation with Source Dependency Representation[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen, Denmark: Association for Computational Linguistics, 2017: 2846-2852.
[18] WU S Z, ZHOU M, ZHANG D D. Improved Neural Machine Translation with Source Syntax[C]// Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. Melbourne, Australia: IJCAI, 2017: 4179-4185.
[19] AHARONI R, GOLDBERG Y. Towards String-to-Tree Neural Machine Translation [C] // Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada: Association for Computational Linguistics, 2017: 132-140.
[20] MORISHITA M, SUZUKI J, NAGATA M. Improving Neural-Machine Translation by Incorporating Hierarchical Subword Features[C]// Proceedings of the 27th International Conference on Computational Linguistics. Santa Fe, New-Mexico, USA: COLING, 2018: 618-629.
[21] XIONG D Y, LI J H, WANG X, et al. Neural Machine Translation with Constraints[J]. Scientia Sinica Informationis, 2018, 48(5): 574-588.
[22] WANG Q, DUAN X Y. Neural Machine Translation Based on Attention Convolution[J]. Computer Science, 2018, 45(11): 226-230.
[23] CHO K, MERRIENBOER B V, BAHDANAU D. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches[C]// Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation. Doha, Qatar: Association for Computational Linguistics, 2014: 103-111.
[24] CHUNG J, GULCEHRE C, CHO K, et al. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling [C]// Proceedings of the Twenty-eighth Conference on Neural Information Processing Systems. Montreal, Quebec, Canada: NIPS, 2014: 1-9.
[25] ZEILER M D. An Adaptive Learning Rate Method[J]. arXiv: 1212.5701.
[26] PETROV S, KLEIN D. Improved Inference for Unlexicalized Parsing[C]// Proceedings of the 2007 Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Rochester, NY: Association for Computational Linguistics, 2007: 404-411.
[27] LI J H, ZHOU G D, HWEE T N. Joint Syntactic and Semantic Parsing of Chinese[C]// Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Uppsala, Sweden: Association for Computational Linguistics, 2010: 1108-1117.
[28] 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: Association for Computational Linguistics, 2002: 311-318.
[29] SNOVER M, DORR B, SCHWARTZ R, et al. A Study of Translation Edit Rate with Targeted Human Annotation[C]// Proceedings of Association for Machine Translation in the Americas. 2006: 231-231.
[30] ISOZAKI H, HIRAO T, DUH K, et al. Automatic Evaluation of Translation Quality for Distant Language Pairs[C]// Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. MIT, Massachusetts: Association for Computational Linguistics, 2010: 944-952.

QIAO Bo-wen, born in 1994, postgraduate. His main research interests include machine translation and so on.

LI Jun-hui, born in 1983, Ph. D, associate professor. His main research interests include machine translation and natural language processing.