融合语言学知识的神经机器翻译研究进展

郭望皓[1*]，范江威[2]，张克亮[1]

（1. 战略支援部队信息工程大学 洛阳校区，河南 洛阳，471003；2. 郑州大学 信息工程学院，河南 郑州，450001）

摘要：尽管神经机器翻译已经成为目前机器翻译研究应用中的主流方法与范式，但同时也存在译文流利但不够忠实、罕见词处理困难、低资源语言表现不佳、跨领域适应性差、先验知识利用率低等问题。受统计机器翻译研究启发，在神经机器翻译模型中融入语言学信息，利用已有的语言学知识，缓解神经机器翻译面临的固有困境，提升翻译质量，成为神经机器翻译研究领域的一个热门话题。根据语法单位分类体系，可以将这方面的研究分为三类：分别是融合字词结构信息的神经机器翻译、融合短语结构的神经机器翻译和融合句法结构信息的神经机器翻译，目前的研究也集中在这三个方面。首先，梳理了神经机器翻译面临的主要挑战及原因，然后重点介绍了当前融合语言学知识的神经机器翻译研究现状与主要成果，最后总结归纳现有研究中仍在存在的问题，展望未来的研究方向。

关键词：神经机器翻译；语言学知识; 字词结构信息; 短语结构信息；句法结构信息

中图分类号：TP391.2　　**文献标志码：** A

自 1954 年世界上第一个机器翻译系统问世以来，到今天已经有 60 余年了。期间，机器翻译经历了百花齐放、百舸争流的盛况，也经历了万籁俱静的萧条与沉寂。主流机器翻译技术发展范式由基于规则的方法，演进到统计方法，再到时至今日的神经网络方法。随着机器翻译译文质量的提升，其应用也由实验室走向人们的日常生活之中，满足大家阅读、会谈、出行、购物等跨语言交际的需求。2013 年以来，神经机器翻译由于不需要设计复杂的特征工程，模型简洁高效得到了研究者与开发人员的青睐，加之并行计算、图形处理器、大数据的广泛应用，在学界和产业界迅速掀起了神经机器翻译的研发热潮，推动神经机器翻译向实用化、商业化方向不断迈进。尽管神经机器翻译取得了巨大成功，但是依然存在着诸如翻译不忠实、存在“过译”和“漏译”现象、罕见词（rare word）和集外词（OOV，out of vocabulary）处理困难、低资源语言表现不佳等问题[1-3]。神经机器翻译架构本身导致了上述问题的产生。表 1 显示了目前神经机器翻译存在的问题及原因。

表 1 目前神经机器翻译存在的问题及原因

Tab. 1 The problems and causes of neural machine translation

问题	原因
罕见词、集外词处理困难	神经网络计算量大，词表规模受限
存在“过译”、“漏译”现象	缺少“硬对齐”，注意力机制缺乏约束，词对齐效果不够好
译文流利但不够忠实	神经网络采用连续的词语表示方法
低资源语言表现不佳	模型构建所需数据量大
先验知识利用率低	神经网络架构单一，难以利用外部知识
跨领域适应性差	神经网络模型迁移困难

为了缓解上述问题，学者们提出了诸多方法改进神经机器翻译模型[4-8]。其中一项重要

* Email：guowanghao@yeah.net

的思路就是将语言学知识融合到神经网络之中，从而提升系统性能，提高翻译质量。本文就是对这一方向的研究进行梳理、归纳和总结，为进一步的相关研究提供文献支撑。

1 融合字词结构信息的神经机器翻译研究

在融合字词结构信息方面，最主要的思路是通过对词以下的结构单位进行编码，降低颗粒度，从而在不改变词表规模、不增加计算时空开销的同时减少集外词的数量。由于神经网络计算量大，所以通常会将源语言和目标语言的词表规模控制在 3 万到 5 万，把词表外的罕见词、集外词统一处理为<UNK>符号，这种处理方式一方面会影响到源语言的语义信息捕获的完整性，另一方面会增加用户理解目标语言的困难程度。这就是上文中提到神经机器翻译面临的挑战之一：罕见词、集外词处理困难。为了缓解这一问题，研究者们进行了下面两种尝试：一是神经机器翻译扩大词表规模或者加装外部词典[4, 5, 9]；二是改变翻译的基本单位，由单词（word）转向字符（character）或者子词（sub-word），利用更细颗粒度的语言单位来减少集外词的数量，也就是将字词结构信息融合到神经机器翻译系统之中。不同颗粒度的词语切分如表 2 所示。

表 2 不同层级语言单位及例句

Tab. 2 Different levels of language units and sentences

语言单位	例句
词语级（word level）	秋高气爽 的 季节 ， 满山 的 红叶 飘落 。
子词级（sub-word level）	秋高@@ 气@@ 爽 的 季节 ， 满山 的 红叶 飘落 。
字符级（character level）	秋 高 气 爽 的 季 节 ， 满 山 的 红 叶 飘 落 。

采用字符作为神经机器翻译的基本语言单位，除了可以消减集外词问题之外，对于诸如汉、日、韩、泰等语言还可以避免分词带来的误差，并且受语言形态变化影响小，有助于提升形态丰富语言（德语、俄语、土耳其语等）的词语利用效率。Kim 等人、Hahn 和 Baroni（2019）等人的研究均涉及到利用神经网络将字符序列转化为词向量的方法[10, 11]。Ling 等人[12]提出在基于注意力机制的神经机器翻译模型前后两端分别增加字符到词（C2W）的组合模块和词向量到字符（V2C）的生成模块。组合模块是利用一个双向长短时记忆网络（Bidirectional LSTM）将在双语两端将字符向量组合成词向量；生成模块是将字符向量、注意力向量和目标词向量进行拼接后通过另外一个单向的长短时记忆网络（LSTM）逐字符生成目标语言的词语（见图 1）。该模型能够学习到部分前后缀在原文和译文之间的对应关系，因此可以识别和生成一些词表中不存在的词形，对于形态复杂的语言间的翻译能够起到帮助作用。但是，该方法需要在双语语料中为每一个单词和句子分别添加开始和结尾的标记，注意力机制仍作用于单词而非字符之上，实验结果与基于单词的神经机器翻译模型相比未有显著提高，但模型复杂度高，训练所需时间长。以字符为单位统计出的句长一般是以单词为单位句长的 6 到 8 倍（由于汉语字符数量多，所以不到 2 倍），造成注意力机制运算量呈平方级增长，同时增加了长距离依赖学习的难度，降低了训练速度。

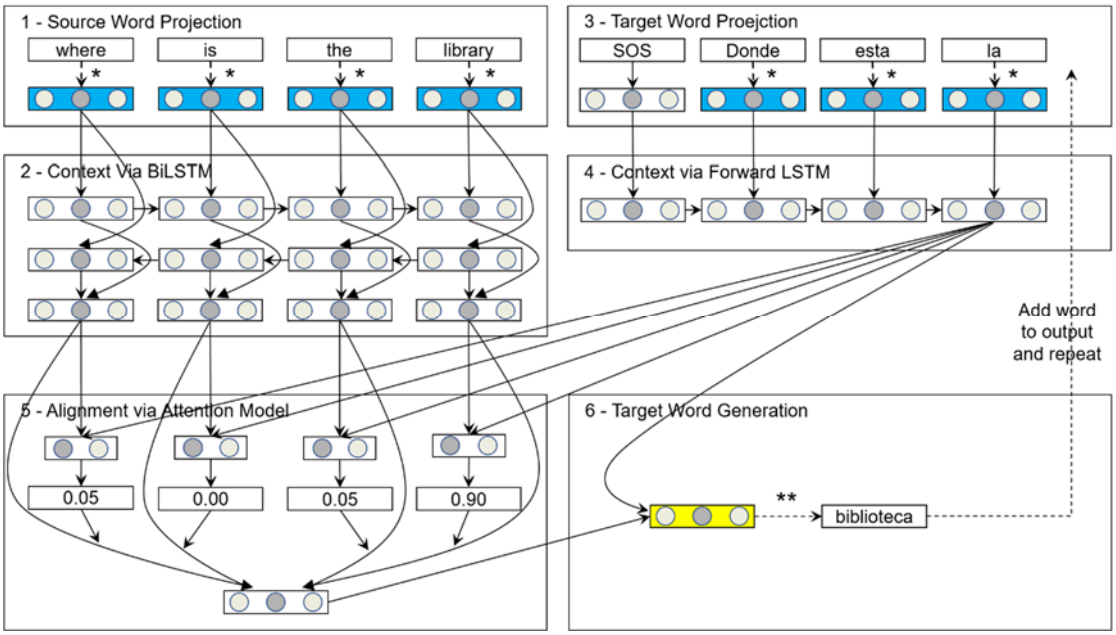


图 1 前后分别加入组合模块（C2W）和生成模块（V2C）的神经机器翻译模型示例[12]

Fig. 1 Illustration of the joint alignment (C2W) and translation (V2C) model[12]

针对这些问题，Lee 等人[13]提出了采用多层卷积（a stack of convolutional），最大池化（Max-pooling）操作与高速公路神经网络层（highway network layers）的处理方案。具体而言，先将输入字符映射为字符向量，再利用窗口大小不同的卷积层进行卷积（相当于学习到与窗口大小相同的 N 元语言模型），然后把卷积输出成分连接起来后再切分成长度固定的序列，对每个序列作最大池化操作（相当于选择最显著的特征作为分割向量（Segment Embeddings）），最后将这些分割向量（相当于具有语言学意义的结构单位）经过高速公路神经网络层和双向的门控循环单元（Bi-GRU，LSMT 的一种变体）进行编码。在解码阶段，注意力机制通过关注源语言的分割向量，并通过一个字符级的门控循环单元生成目标语言的字符序列。在德-英、捷克-英、芬兰-英和俄-英机器翻译实验结果表明，字符级神经机器翻译模型在拼写错误单词、罕见词、词形变化、临时构造词翻译处理方面具有优势，同时对于像德、捷克、芬兰这些字符相近的语言，字符级神经机器翻译模型能够学习到各语言间通用的语素，可以在不增加模型规模的条件下通过共享一个编码器实现多语言（多到一，many-to-one）机器翻译。

基于字符的神经机器翻译虽然减少了集外词的数量，缓解了词表规模受限问题，但是单个字符义项增加，更容易产生歧义，并且增大了长距离依赖问题，导致长句翻译质量下滑。为此，有学者提出采用介于词语和字符之间的单位进行编码，其中最具代表性的工作当属 Sennrich[14]提出的子词（sub-word）字节对编码（BPE，Byte Pair Encoding）方案。作者受命名实体、同源词、借词、复杂形态词（这些词大部分属于罕见词或集外词）翻译策略的启发，当专业译员遇到这些不认识的单词时往往会通过分析其组成成分预测单词的意义，认为将这些罕见词或集外词处理为子词有助于缓解神经机器翻译的词表规模受限问题。具体而言，这种方法将经常组合在一起的字符序列看作是一个单位，如英文中的词缀“er”、“ism”、“dis”，词尾“ed”、“ing”等。做法是将所有单词以字符划分，不断将频次最高的 N-gram 进行合并操作，一直迭代至词表规模大小。实验结果显示，在 WMT15 英德和英俄任务上，较之于传统的神经机器翻译模型，基于子词的模型 BLEU 值分别提升了 1.1 和 1.3。相对于基于单词的神经翻译模型和基于字符的神经翻译模型，该研究提出的子词模型在词表大小和句子长度两

方面取得了平衡。由于子词单元能够在相近或者同源语言间共享词干、词缀和词尾的信息，基于子词的神经机器翻译方法得到了广泛的应用，由最初仅用来处理罕见词或集外词，发展到全部单词均切分成子词单元再喂入神经网络模型之中进行运算，基于子词单元的神经机器模型在某些语言间（如英、法、德等）的翻译系统中逐渐成为标配，著名的谷歌神经机器翻译（GNMT）系统[15]和 Transformer 系统[16]也都采用这一设计思想和处理方式。

有的研究工作，在源语言编码和目标语言解码两端分别采用基于不同层级语言单位的建模方案。Costa-Jussa 等人[17]在源语言端通过卷积滤波器（convolution filters）和高速公路网络层（Highway Layers）实现了由字符到词向量的映射过程。字符级编码方式利用单词的内部信息，能够捕捉到源语言所有单词的全部表达形式，消减了源语言端的集外词问题。但在目标语言端仍以词语为单位进行解码与生成，因此仍受词表规模的限制。Chung[18]的主要工作是在解码端使用了一种新的名为双尺度循环神经网络（biscale RNN）的结构，可以在字符和单词两个时间尺度上进行处理，不需要进行分词，直接生成目标语言字符序列。但是该研究在源语言端采用的子词结构。与之相似的还有 Yang 等人[19]、Su 等人[20]的工作。

有的研究工作将不同层级语言单位编码后混合到同一神经机器翻译模型之中。Luong 和 Manning[21]设计了一个字符-单词混合的神经机器翻译模型。整个模型主要由单词级模块驱动，当出现< UNK >符号时，模型会调用字符级模块，将源语言中的< UNK >对应的单词转换为该单词字母构成的字符向量，把目标语言中的< UNK >恢复生成为单词（见图 2）。源语言和目标语言两端的字符级模块都是通过一个四层单向的 LSTM 训练得到的，不同之处在于，源语言端的字符级模块是上下文独立（context independent）的，因此可以进行预训练、预计算，而目标语言端的字符级模块是上下文依存（context dependent）的。不过够由于结构较为复杂，基于字符的模型训练时间长达 3 个月之久。

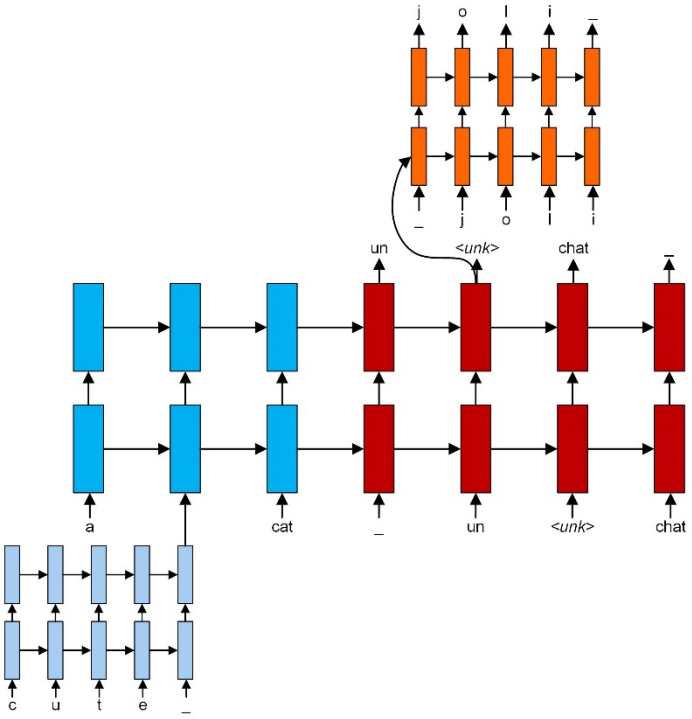


图 2 字符-单词混合神经机器翻译模型示例[21]

Fig. 2 Illustration of word-character model hybrid NMT[21]

Chen 等人[22]提出同时将字信息与词信息进行编码，即将不同颗粒度语言单位表示融合到一个神经机器翻译系统之中。在源语言端，先用两个独立的注意力模块分别学习每个单词

的词内字符向量和词外字符向量，前者可提供单词内部字符间关系信息，后者提供单词边界信息；再将学到的两个字符向量通过前馈神经网络连接后嫁接到词（或子词）向量之上，形成具有字符信息的词向量；然后将此词向量喂入循环神经网络进行计算。在目标语言端，解码器采用了一个多尺度的注意力机制（multi-scale attention mechanism）模块，既能采集到词向量蕴含的信息，也能够采集到字向量信息。实验表明，在汉英互译任务中，该模型表现优于单纯基于字符以及单纯基于单词的神经机器翻译模型；在英译德任务中，该模型优于采用 BPE 技术的子词翻译模型。实验结果还显示，这种方法不仅可用于缓解神经机器翻译的集外词问题，对于提升常见词翻译的准确性也有所帮助，原因就在于编码器中融合了由字符提供的单词的内部信息与边界信息。Wang 等人[23]的工作也是用一个混合注意力机制模型将源语言的单词信息和字符信息分别编码，两类信息具有兼容性和互补性，该方法在汉英机器翻译实验中与传统基于单词的基线模型相比取得了 1.92 个 BLEU 值的提升。

除此之外，还有研究者将目光转向的比字符颗粒度更低的语言单位：亚字（sub-character）。如果说亚词能够学到词干、构词词缀和构形词缀的信息的话，在中、日等语素文字（ideographs / logograph）体系中亚字就包含了构件（如汉字的偏旁）的语义信息。现代汉语中，形声字的比重占到 90%左右，也就是说绝大多数的汉字能够拆分为“声旁”和“形旁”，其中“形旁”相同的汉字往往在意义上有联系，如“桃、梅、梨、枝、株、棵”都与树木有关，这就为基于亚字的神经机器翻译模型提供了基础。另外在汉语和日语中，有时相同或者相近的字形表示相同的意义，如中文汉字“风景”和日文汉字“風景”写法相近，意义相同，因此在中日互译时其汉字组成成分间的信息可以互享，从而提高表示精度。Zhang 和 Komachi[24]就进行了这方面的研究。研究在中、日、英三种语言的翻译中开展，英语采用词向量，中、日文分别采用词向量、字向量、构件向量和笔画向量。除词以外的语言单位均由 BPE 切分组合而得。实验结果显示，对于中文，基于构件的表示方法能够提升模型的翻译质量，而对于日语，基于笔画的模型才是最优解。

2 融合短语结构信息的神经机器翻译研究

短语结构的意义并非都是其组成成分的简单加和，这样的例子在各种语言中都比比皆是、屡见不鲜。如英语中的“let alone（更不必说）”、“by and large（总的来说）”、“red tape（繁文缛节）”，汉语中的“网络水手”、“买面子”、“996（指每天早上 9 点上班，晚上 9 点下班，一周工作 6 天）”等等。由此可见，在翻译过程中，短语占据着举足轻重地位和作用。统计机器翻译发展历程中也证实了这一点，正是基于短语的统计机器翻译[25-27]技术走向成熟，机器翻译才算真正地走向实用。

由于统计机器翻译在短语翻译研究方面有着较长时间的积累和较为成熟的经验，所以如何利用原有研究成果与神经机器翻译模型相融合就成为研究者们自然而然的想法了。Wang 等人[28]就是在神经机器翻译的解码器上增加了一个统计机器翻译模块用于生成短语。每当解码器工作到下一步时，先通过一个名为 balancer 多层神经网络判断要生成的单词还是短语，如果要生成单词，那就还用神经机器翻译模块进行生成；如果要生成短语则调用统计机器翻译模块的结果。与之类似的研究还有[29-31]Tang 等人[29]、Dahlmann[30]、Rikters 和 Bojar[31]等人的工作，但是他们的研究都借助于外部装置提取并记忆短语翻译的结果，神经机器翻译模型本身并不能处理生成短语结构。

利用神经网络进行短语结构的翻译就需要从编码器和解码器入手，通过扩充或者改造，使其能够处理短语层级的信息。Li 等人[32]提出的模型有两个编码器和一个解码器：两个编码器分别以单词和短语为单元对源语言的句子进行编码，解码器工作时会同时考虑单词向量和短语向量中蕴含的信息。这个简单的结构在汉英翻译任务中取得了不错的成绩，较之于传统

模型平均提高了 1.13 个 BLEU 值。Ishiwatari 等人[33]提出的模型则包含两个解码器，一个用于处理短语（文中称为组块或语块 chunk）间的依赖关系，而另一个用于对短语内单词间的关系进行建模。该方法在 WAT16 英日翻译任务取得了出色的成绩。Zhou 等人[34]的工作是在解码器中引入一个额外的神经网络层，实现了从短语到单词的分层次的译文生成过程，在多种语言上进行的实验表明该方法能够显著提高翻译效果。Huang 等人[35]提出了基于短语的神经机器翻译（NPMT）。他们提出的方法不需要事先准备短语，目标语言端的短语是通过一个 Sleep-Wake 网络（SWAN）和一个调序层（reordering layer）从训练语料中自动提取到的（见图 3）。SWAN 是 Wang 等人[36]提出的一项基于分割的序列建模技术。实验结果显示，这一方法能够将目标序列切割成为具有语言学意义的短语。在 IWSLT2014 德英互译，IWSLT2015 英译越数据集上 BLUE 值结果表示，这种方法超越了基于注意力机制的神经机器翻译模型。

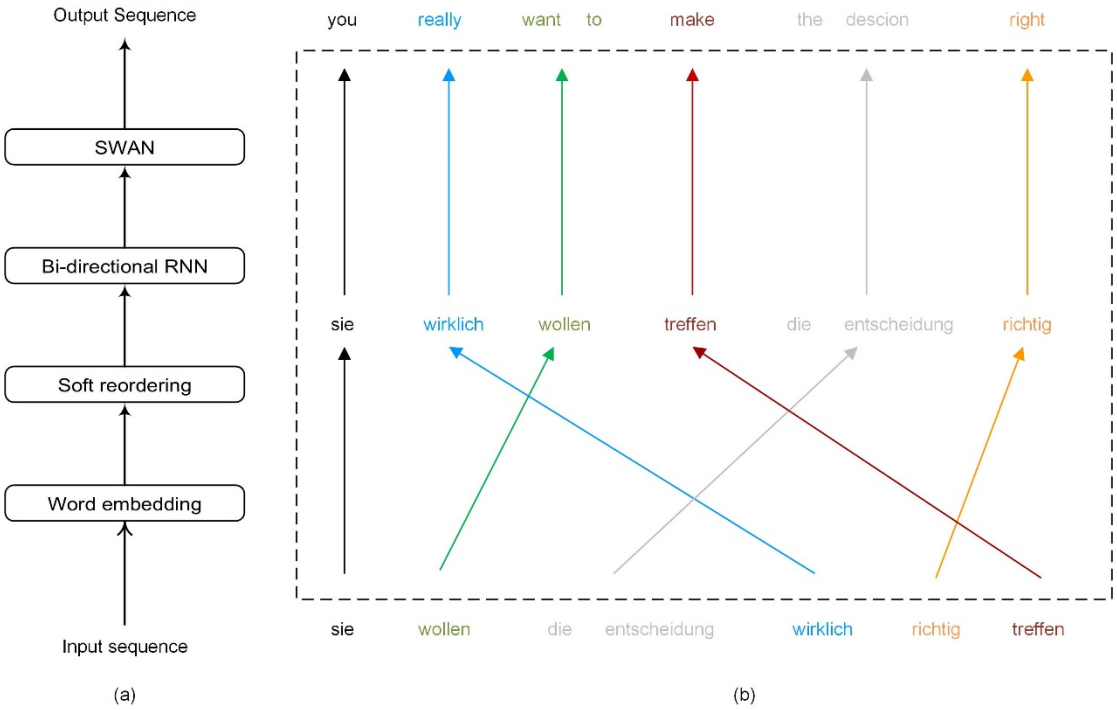


图 3 基于短语的神经机器翻译模型示例（a）及德-英翻译实例（b）[35]

Fig. 3 Illustration of phrase NMT (a) and German-English Translation (b) [35]

3 融合句法结构信息的神经机器翻译研究

句子并非单词的简单线性排列，它是有层次关系结构的。如：“关心孩子的母亲”，既可能是一个动宾结构“关心/孩子的母亲”，也可能是一个定中结构“关心孩子的/母亲”；“门把手弄坏了”，既可能是“门/把/手/弄/坏了”，也可能是“门把手/弄/坏了”。因此将句法结构信息融合至机器翻译系统中有助于消解歧义，提升准确性。早在统计机器翻译时代，句法结构信息的价值就已经得到了证明[37-42]。受上述研究的启发，学者们尝试将未被显示建模的句法结构信息融入到神经机器翻译模型之中，其中主要用到的两种句法理论分别是短语结构语法和依存语法。

在源语言端融入句法结构信息的研究有：Eriguchi 等人[43]在研究英日机器翻译时发现，两种语言在语序、句法结构方面均有较大差异，一般的注意力机制模型难以处理词与短语、短语与短语之间的对齐，为此他们提出了树到序列（tree-to-sequence）的注意力机制神经机器翻译模型，在编码阶段，利用中心语驱动的短语结构文法（HPSG）对源语言进行自底向上

的编码，从而获得了源语言的短语结构信息（见图 4）。在 WAT15 英日数据集上的测试结果证实了这种方法的有效性。

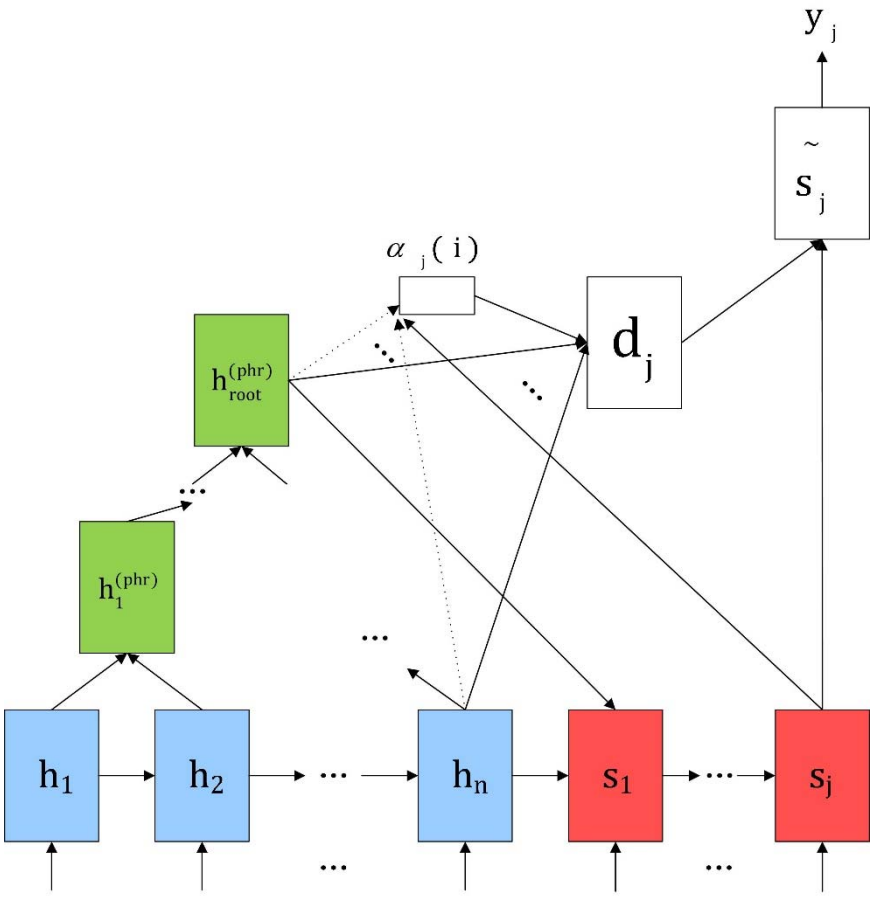


图 4 基于短语结构文法的树到序列神经机器翻译模型示例[43]

Fig. 4 Illustration of tree-to-sequence NMT model base on Phrase Structure Grammar [43]

Chen 等人[44]（2017）在此基础上，将源语言端的单向树状结构进行了强化，变成自底向上和自上而下双向编码，从一定程度上克服了 Eriguchi 等人研究中存在的顶端节点包含的句法信息多，底端节点包含的句法信息少的问题。此外，在解码端引入了基于树的覆盖率机制[45]，可以有效地将源语言上下文知识整合至注意力机制之中。研究采用宾州汉语树库作为源语言的句法剖析工具。在 NIST 英汉翻译数据集上的实验结果显示，该方法较之于基线神经机器翻译系统平均高出 3.54 个 BLEU 值，在同等条件下双向编码较之于单向编码高出 0.79-0.96 个 BLEU 值，而基于树的覆盖率机制的引入则提升了 0.4-1.13 个 BLEU 值。与 Chen 等人抛弃句法标签信息的做法不同，Li 等人[46]的工作是将句法树转化为句法标签后与词语混合成为同一个线性化序列，这种方法的好处在于避免树的复杂网络结构（见图 5）。实验结果显示，在长句翻译、词及短语对其准确率和过译三个方面均优于基线神经机器翻译模型。

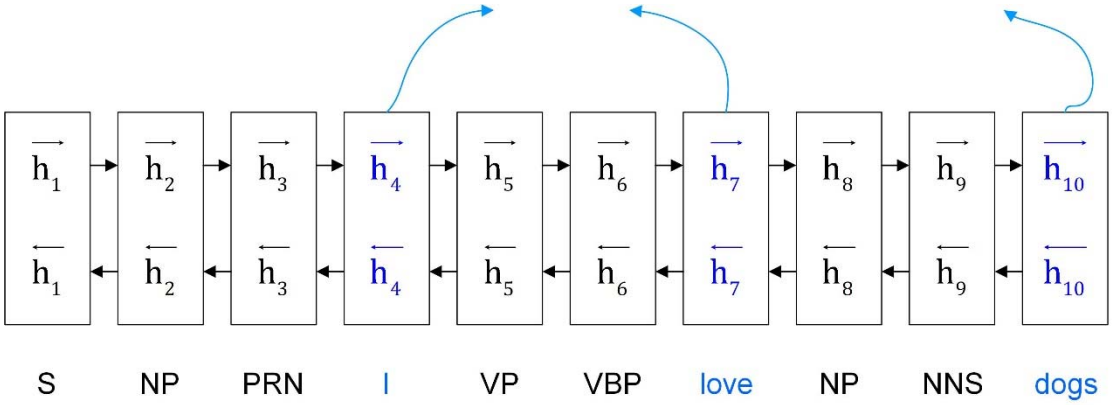
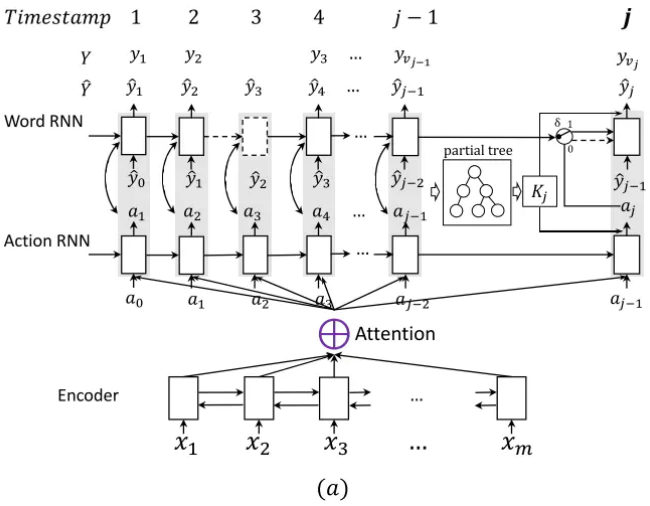


图 5 句法标签序列化示例[46]

Fig. 5 Illustration of syntax tags linearization [46]

除了短语结构文法，依存文法在源语言编码方面也有不少应用。其中代表性的工作是 Bastings 等人[47]利用图卷积网络（GCN，图神经网络 GNN 的一种）对源语言的依存结构进行编码。他们的研究将 GCN 叠加在 CNN 之上，以 CNN 编码后的隐层向量作为输入，通过依存结构信息对隐层向量进行图学习，为每一个词生成一个包含依存句法信息的向量，从而使得翻译模型获取句法知识。除此之外，在源语言融入句法知识的还有 Ma 等人[48]的森林—序列（forest-to- sequence）模型，Xu 等人[49]的图—序列（graph-to- sequence）模型等，Sennrich 和 Haddow[50]将词性还原、词性和依存句法标签向量化后与词向量进行拼接，新的词向量就包含不同层级的语言学信息。

在目标语言端融入句法知识的研究有：Nadejde 等人[51]将组合范畴文法（Combinatory Categorial Grammar，CCG）标注引入神经机器翻译的解码器端，其方法有两种：一是将句法标签与目标语言词语交叉排列，即一个词语一个对应的标签，输出序列长度增加一倍；二是借鉴多任务学习（Multi-task Learning）的思路，将句法标签序列与目标语言序列分别用一个解码器进行解码。在德语-英语和罗马尼亚语-英语的翻译实验证实了解码阶段加入句法知识的有效性，且第一种方法的结果优于多任务学习的方法。实验还显示，如果同时在源语言端也加入语言学知识的话，翻译性能会得到进一步提升。Aharoni 和 Goldberg[52]的研究思路是，在模型训练阶段，先将目标语言句子通过句法分析器转换为其句法树线性化序列，一个既包含该句子所有单词，也包含句法结构成分标签的序列，然后将这一序列代替目标语句子与源语言进行模型训练。在翻译过程中，能够同时生成目标语言和目标语言的树结构，利用目标语言树结构的约束和限制，最终可以得更为准确的目标语翻译。在 WMT16 德英新闻翻译任务数据集上的结果显示该方法能够提升 0.94 个 BLEU 值。Eriguchi 等人、Wu 等人、Le 等人的工作集中在如何在解码端利用依存文法来提升模型的翻译质量[53-55]。Eriguchi 等人[53]的思路是用 RNNG（Recurrent Neural Network Grammars）作为神经机器翻译模型的解码器；Wu 等人[54]的方法是利用两个 RNN 网络先后用以依存句法结构的生成和词语生成；Le 等人[55]的想法是将目标语言通过斯坦福依存文法分析器剖析成的句法树序列化后代替目标语言的句子进行模型训练。以上这些方法都被证明句法结构信息有助于提升机器翻译的质量。



(a)

图 5 序列到依存的神经机器翻译模型示例[54]

Fig. 6 Illustration of sequence-dependency NMT [54]

4 问题与展望

从上文所述的研究中不难看出，融合语言学知识后的神经机器翻译模型的确能够提升翻译性能。但是，也不得不承认，目前的研究还存在以下几个问题。（1）融合语言学知识后，或多或少地增加了神经机器翻译模型的复杂度，使得模型训练需要消耗更多的资源，耗费更久的时间。（2）在通过线性化的方法加入语言学知识时，不论是在编码阶段还是在解码阶段都会使得序列变得更长，从而进一步加剧了长句处理的困难程度。这就形成了一个“怪圈”：一般而言，句子越长成分越复杂，越需要句法信息的辅助，而一旦增加了句法信息，句子序列变长，又会导致模型翻译性能下降。（3）融合语言学信息的神经机器翻译模型研究受到语言学理论研究和相应工具开发的限制。在模型中引入哪种类型的语言学知识，这种知识来自于何种语言学或计算语言学理论，有没有开发出高质量的标注工具，这些问题都将与最终研究结果息息相关。以句法分析为例，目前各种类型的句法自动分析工具都会产生大量的标注错误，这就在一定程度上限制了数据使用的规模与质量。因此，我们认为今后融合语言学知识的神经机器翻译研究将着眼但不限于以下方面。（1）寻求通过一种简单易行的方式将语言学知识融入到神经机器翻译模型之中。（2）探讨句法信息和句子长度间相互作用关系，寻求二者最佳结合点。（3）理论研究、数据资源建设、工具开发与工程实践通力合作、共同发力提升翻译模型的质量与性能。（4）与最新的研究范式和模型架构向结合，不断创造最好的翻译效果。

参考文献：

[1] Koehn, P. and R. Knowles. Six Challenges for Neural Machine Translation[C]. In *Proceedings of the First Workshop on Neural Machine Translation*. 2017.

[2] 刘洋, 神经机器翻译前沿进展[J]. 计算机研究与发展, 2017. **54**(06): p. 1144-1149.

[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): p. 2734-2755.

[4] Luong, M.-T., I. Sutskever, Q. Le, et al. Addressing the Rare Word Problem in Neural Machine Translation[C]. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 2015.

[5] Jean, S., K. Cho, R. Memisevic, et al. On Using Very Large Target Vocabulary for Neural Machine Translation[C]. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 2015. Beijing, China: Association for Computational Linguistics.

[6] Bahdanau, D., K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate[C]. In *3rd International Conference on Learning Representations, ICLR 2015*. 2015.

[7] Sennrich, R., B. Haddow, and A. Birch. Improving Neural Machine Translation Models with Monolingual Data[C]. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2016.

[8] Chu, C., R. Dabre, and S. Kurohashi. An Empirical Comparison of Domain Adaptation Methods for Neural Machine Translation[C]. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. 2017.

[9] Li, X., J. Zhang, and C. Zong. Towards Zero Unknown Word in Neural Machine Translation[C]. In *IJCAI*. 2016.

[10] Kim, Y., Y. Jernite, D. Sontag, et al. Character-aware neural language models[C]. In *Thirtieth AAAI Conference on Artificial Intelligence*. 2016.

[11] Hahn, M. and M. Baroni *Tabula nearly rasa: Probing the Linguistic Knowledge of Character-Level Neural Language Models Trained on Unsegmented Text*. arXiv e-prints, 2019.

[12] Ling, W., I. Trancoso, C. Dyer, et al., Character-based neural machine translation[J]. *Computer Science*, 2015.

[13] Lee, J., K. Cho, and T. Hofmann, Fully character-level neural machine translation without explicit segmentation[J]. *Transactions of the Association for Computational Linguistics*, 2017. **5**: p. 365-378.

[14] Sennrich, R., B. Haddow, and A. Birch. Neural Machine Translation of Rare Words with Subword Units[C]. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2016.

[15] Wu, Y., M. Schuster, Z. Chen, et al., Google's neural machine translation system: Bridging the gap between human and machine translation[J]. 2016.

[16] Vaswani, A., N. Shazeer, N. Parmar, et al. Attention is all you need[C]. In *Advances in neural information processing systems*. 2017.

[17] Costa-jussà, M.R. and J.A. Fonollosa. Character-based Neural Machine Translation[C]. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. 2016.

[18] Chung, J., K. Cho, and Y. Bengio. A Character-level Decoder without Explicit Segmentation for Neural Machine Translation[C]. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2016.

[19] Yang, Z., W. Chen, F. Wang, et al. A character-aware encoder for neural machine translation[C]. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. 2016.

[20] Su, J., Z. Tan, D. Xiong, et al. Lattice-based recurrent neural network encoders for neural machine translation[C]. In *Thirty-First AAAI Conference on Artificial Intelligence*. 2017.

[21] Luong, M.T. and C.D. Manning. Achieving Open Vocabulary Neural Machine Translation with Hybrid Word-Character Models[C]. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2016.

[22] Chen, H., S. Huang, D. Chiang, et al. Combining character and word information in neural machine translation using a multi-level attention[C]. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. 2018.

[23] Wang, F., W. Chen, Z. Yang, et al., Hybrid Attention for Chinese Character-Level Neural Machine Translation[J]. 2019. **358**: p. 44-52.

[24] Zhang, L. and M. Komachi. Neural Machine Translation of Logographic Language Using Sub-character Level Information[C]. In *Proceedings of the Third Conference on Machine Translation: Research Papers*. 2018.

[25] Koehn, P., F.J. Och, and D. Marcu. Statistical phrase-based translation[C]. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*. 2003. Association for Computational Linguistics.

[26] Chiang, D., A. Lopez, N. Madnani, et al. The Hiero machine translation system: Extensions, evaluation, and analysis[C]. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*. 2005. Association for Computational Linguistics.

[27] Koehn, P., Statistical machine translation[M]. 2009: Cambridge University Press.

[28] Wang, X., Z. Tu, D. Xiong, et al. Translating Phrases in Neural Machine Translation[C]. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 2017.

[29] Tang, Y., F. Meng, Z. Lu, et al., Neural machine translation with external phrase memory[J]. arXiv preprint arXiv:.01792, 2016.

[30] Dahlmann, L., E. Matusov, P. Petrushkov, et al. Neural Machine Translation Leveraging Phrase-based Models in a Hybrid Search[C]. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 2017.

[31] Rikters, M. and O.J.a.p.a. Bojar, Paying attention to multi-word expressions in neural machine translation[J]. arXiv preprint arXiv:.06313, 2017.

[32] Li, Y., D. Xiong, and M. Zhang. Neural Machine Translation with Phrasal Attention[C]. In *Machine Translation: 13th China Workshop, CWMT 2017, Dalian, China, September 27-29, 2017, Revised Selected Papers*. 2017. Springer.

[33] Ishiwatari, S., J. Yao, S. Liu, et al. Chunk-based decoder for neural machine translation[C]. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2017.

[34] Zhou, H., Z. Tu, S. Huang, et al. Chunk-Based Bi-Scale Decoder for Neural Machine Translation[C]. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. 2017.

[35] Huang, P.-S., C. Wang, S. Huang, et al. *Towards Neural Phrase-based Machine Translation*. arXiv e-prints, 2017.

[36] Wang, C., Y. Wang, P.-S. Huang, et al. Sequence modeling via segmentations[C]. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. 2017. JMLR. org.

[37] Liu, Y., Q. Liu, and S. Lin. Tree-to-string alignment template for statistical machine translation[C]. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*. 2006. Association for Computational Linguistics.

[38] Nguyen, T.P., A. Shimazu, T.-B. Ho, et al. A tree-to-string phrase-based model for statistical machine translation[C]. In *Proceedings of the Twelfth Conference on Computational Natural Language Learning*. 2008. Association for Computational Linguistics.

[39] Marton, Y., & Resnik, P. Soft syntactic constraints for hierarchical phrased-based translation[C]. In *Proceedings of ACL-08: HLT*. 2008. Association for Computational Linguistics.

[40] Shen, L., J. Xu, and R. Weischedel. A new string-to-dependency machine translation algorithm with a target dependency language model[C]. In *Proceedings of ACL-08: HLT*. 2008.

[41] Xie, J., H. Mi, and Q. Liu. A novel dependency-to-string model for statistical machine translation[C]. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 2011. Association for Computational Linguistics.

[42] Li, P., Y. Liu, and M. Sun. Recursive autoencoders for ITG-based translation[C]. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. 2013.

[43] Eriguchi, A., K. Hashimoto, and Y. Tsuruoka. Tree-to-Sequence Attentional Neural Machine Translation[C]. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2016.

[44] Chen, H., S. Huang, D. Chiang, et al. Improved Neural Machine Translation with a Syntax-Aware Encoder and Decoder[C]. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2017.

[45] Tu, Z., Z. Lu, Y. Liu, et al. Modeling Coverage for Neural Machine Translation[C]. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2016.

[46] Li, J., D. Xiong, Z. Tu, et al. Modeling Source Syntax for Neural Machine Translation[C]. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2017. Vancouver, Canada: Association for Computational Linguistics.

[47] Bastings, J., I. Titov, W. Aziz, et al. Graph Convolutional Encoders for Syntax-aware Neural Machine Translation[C]. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 2017.

[48] Ma, C., A. Tamura, M. Utiyama, et al. Forest-based neural machine translation[C]. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2018.

[49] Xu, K., L. Wu, Z. Wang, et al., Graph2seq: Graph to sequence learning with attention-based neural networks[J]. arXiv preprint arXiv:.00823, 2018.

[50] Sennrich, R. and B. Haddow. Linguistic Input Features Improve Neural Machine Translation[C]. In *Proceedings of the First Conference on Machine Translation: Volume 1, Research Papers*. 2016.

[51] Nadejde, M., S. Reddy, R. Sennrich, et al. Predicting Target Language CCG Supertags Improves Neural Machine Translation[C]. In *Proceedings of the Second Conference on Machine Translation*. 2017.

[52] Aharoni, R. and Y. Goldberg. Towards String-To-Tree Neural Machine Translation[C]. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. 2017.

[53] Eriguchi, A., Y. Tsuruoka, and K. Cho. Learning to Parse and Translate Improves Neural Machine Translation[C]. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. 2017.

[54] Wu, S., D. Zhang, N. Yang, et al. Sequence-to-dependency neural machine translation[C]. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2017.

[55] Le, A.N., A. Martinez, A. Yoshimoto, et al. Improving sequence to sequence neural machine translation by utilizing syntactic dependency information[C]. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 2017.

Advance Research on Neural Machine Translation Integrating Linguistic Knowledge

Abstract: Although neural machine translation has become the mainstream method and paradigm in the current research and application of machine translation, there are also some problems such as the fluent but not faithful of the translation results, difficult processing of rare words, poor performance of low-resource languages, poor cross-domain adaptability, and prior knowledge utilization. Inspired by statistical machine translation research, incorporating linguistic information into neural machine translation models has become a hot topic in the field of neural machine translation research for using existing linguistic knowledge, alleviating the inherent difficulties faced by neural machine translation and improving translation quality. According to the grammatical unit's classification system, this research can be divided into three categories: neural machine translation that incorporates character or word structure information, neural machine translation that incorporates phrase structure information, and neural machine translation that incorporates syntactic structure information. First, it summarizes the main challenges and causes of neural machine translation, then highlights the current status and main achievements of neural machine translation research integrating linguistic knowledge, and finally summarizes the problems that still exist in the existing research and looks forward to future research direction.
Keywords: Neural Machine Translation; linguistic knowledge; character or word structure information; phrase structure information; syntactic structure information