层次化结构全局上下文增强的篇章级神经机器翻译

陈林卿，李军辉*，贡正仙†

（苏州大学 自然语言处理实验室，江苏 苏州 215006）

摘要

如何有效利用篇章上下文信息一直是篇章级神经机器翻译研究领域的一大挑战。本文提出利用来源于整个篇章的层次化全局上下文提高篇章级神经机器翻译性能。为了实现该目标，本文模型分别获取当前句内单词与篇章内所有句子及单词之间的依赖关系，结合不同层次的依赖关系以获取含有层次化篇章信息的全局上下文。最终源语言当前句子中的每个单词都能获取其独有的综合词和句级别依赖关系的上下文。为了充分利用平行句对语料在训练中的优势本文使用两步训练法，在句子级语料训练模型的基础上使用含有篇章信息的语料进行二次训练以获得捕获全局上下文的能力。在若干基准语料数据集上的实验表明本文提出的模型与若干强基准模型相比取得了有意义的翻译质量提升。实验进一步表明，结合层次化篇章信息的上下文比仅使用词级别上下文更具优势。除此之外，本文尝试通过不同方式将全局上下文与翻译模型结合并观察其对模型性能的影响，并初步探究篇章翻译中全局上下文在篇章中的分布情况。

关键词： 神经机器翻译；篇章上下文

Hierarchical Global Context Augmented Document-level Neural Machine Translation

CHEN Linqing, LI Junhui, GONG Zhengxian
(Natural Language Processing Laboratory, Soochow University, Suzhou, Jiangsu 215006)

Abstract

How to effectively use textual context information is always a challenge in the field of document-level neural machine translation. This paper proposes to use the hierarchical global context generated from the entire document to improve the performance of document-level neural machine translation models. In order to achieve this goal, this model obtains the dependencies between the current words in the sentence and all of the sentences and words in the document respectively, and combines the dependencies of different levels to obtain the global context containing the hierarchical contextual information, which can be use to guide translating the current sentence. Each word in the current sentence of the source language gets its own context that combines word and sentence level dependencies. In order to make full use of

基金项目：国家自然科学基金(61876120)
基金项目：国家自然科学基金(61976148)

the advantages of the parallel sentence-level corpus in training, the two-step training method is used in this paper. Based on the Transformer ,which is trained on a sentence-level corpus, the corpus containing textual information is used for secondary training to help the model gain the ability to capture and understand global context. Experiments on several benchmark corpus data sets show that the proposed model can significantly improve translation quality compared with other strong baseline models. The experiment further shows that combining hierarchical contextual information is more advantageous than word level context. In addition, this paper attempts to combine the global context with the translation model in different ways and observe its influence on the performance of the model, and studies the distribution of the global context in document-level translations.

Keywords: Neural Machine Translation , Document-level Context

1 引言

神经机器翻译近两年不断取得鼓舞人心的进展，已经成为当前机器翻译最受关注的研究领域之一。在过去几年中研究者们通过一系统模型不断提高神经机器翻译的性能(Sutskever et al., 2014; Bahdanau et al., 2015; Vaswani et al., 2017; Gehring et al., 2017)。机器翻译和人工翻译之间的质量差距被这些出色的工作不断缩小。其中Transformer模型Vaswani et al., (2017)凭借多头注意力机制在句子级神经翻译任务中达到了当前最好成绩。然而Transformer在篇章级神经机器翻译任务中的表现却差强人意，主要原因在于其忽略了篇章句子间的依赖关系也没能有效利用篇章上下文。

研究者们提出各种获取上下文的方法改善前文所述问题，(Maruf and Haffari., 2018; Wang et al., 2017; Zhang et al., 2018)通过提取前文语句帮助模型翻译当前语句，此类方法没有充分建模当前语句前后上下文有较大差异的情况。当前句若只利用前侧上下文，可能由于有效信息不完整造成负面影响甚至是当前句翻译错误。同时，当前句之后语句的翻译也可能受前句的语义偏差影响造成错误累积。Miculicich et al., (2018)提出了利用全文获取上下文的方法，但仍将注意力聚焦在篇章的一定范围内。Tan et al., (2019)等人则提出了新的方法将整个篇章作为上下文来源，通过层次化网络利用句向量之间的注意力机制获取上下文向量，并将其分配给当前句中的词以帮助模型提高篇章翻译质量。完全依赖句向量将全局上下文与当前句间接结合的方法在信息高度压缩的过程中可能造成有效信息损失，也没有直接获取并利用当前句中的词与全文中其他句子或单词的依赖关系。

不同于以上方法，本文提出利用具有层次化结构信息的全局上下文提高神经机器翻译模型的性能。如图1所示在本文提出的模型中，我们通过不同注意力层分别提取来自不同层次的上下文依赖关系并将二者结合，使得获取的上下文包含多层次篇章信息。为了尽可能多获取上下文，该模型一次性从篇章全文获取前述层次化上下文而不是只使用当前语句之前的语句。由于篇章信息获取过程基于当前句中的所有单词分别计算，使得每一个词都能差异化获取来自整个篇章的有效上下文。受(Zhang et al., 2018; Miculicich et al., 2018)启发，本文使用两步训练法进行训练，从而高效利用含有篇章信息的语料。

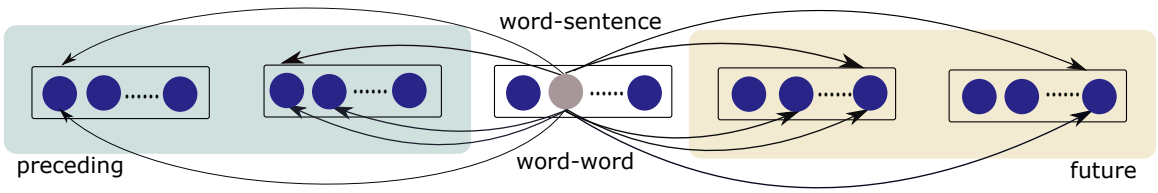

图 1: 计算当前句中的词与篇章中所有句/词的依赖关系

2 层次结构全局上下文增强的翻译模型

本文研究目标是将综合包含句子级依赖关系及单词级依赖关系的全局上下文结合进翻译模型，从而提高模型的篇章翻译质量。为了实现这个目标，我们首先利用源端编码器自注意力层

输出的词级隐藏状态获取句子向量。然后基于词级隐藏状态及句子向量分别计算当前句子中每个词与篇章中所有词及所有句子之间的依赖关系。最终，我们通过综合了两种依赖关系的权重获取具有层次化篇章结构信息的全局上下文，并利用这些上下文辅助模型进行篇章翻译。为了便于理解本文做出如下定义：含有N个语句的文档表示为：$\mathcal{X} = (X_1, \cdots, X_N)$，篇章中的句子$X_i = (x_{i,1}, \cdots, x_{i,n})$ 含有n 个词, 本文使用d_m 表示隐藏状态及词嵌入的维度。

2.1 词-句级依赖权重

为了避免之前的研究工作中仅使用当前语句前面的句子作为上下文对翻译质量造成的负面影响及错误累积。本文将篇章中的所有语句作为上下文来源。如图2(a)所示，词-句级依赖权重生成模块自下而上由编码器自注意力层，句向量嵌入层及词-句权重生成层组成。该模块将源端编码器自注意力层输出的隐藏状态以句子为单位嵌入为句向量，再通过当前句中词与全文句向量之间的注意力函数获取词-句级别的权重。

编码器自注意力层： 本文使用多头注意力函数Vaswani et al., (2017)捕获同一句子中单词间的依赖关系。篇章中的每个句子都会以词为单位被编码器编码，从而获取源端语句的词级隐藏状态:

$$S_i^{(k)} = \text{MultiHead}\left(A_i^{(k)}, A_i^{(k)}, A_i^{(k)}\right), \tag{1}$$

MultiHead表示多头注意力函数，通过将输入映射到不同子空间对输入序列之间的依赖关系进行建模。其中输出$S_i^{(k)}$ 的维度为$\mathbb{R}^{n\times d_m}$. 对于编码器的第一层来说, $A_i^{(1)} = X_i$ 而对于编码器的其他层而言$A_i^{(k)}$ 是上一层编码器的输出$A_i^{(k-1)}$。如图2所示，本文将这部分参数作为上下文生成器的共享参数。

在生成句向量之前本文使用残差网络和层标准化对编码器自注意力层的输出进行规整，得到的实际输出如下:

$$S_i^{(k)} = \text{LayerNorm}\left(S_i^{(k)} + A_i^{(k)}\right). \tag{2}$$

其中，LayerNorm是层规范化函数。出于保持模型结构图的简洁，本文在后续插图中省略了每个注意力层后的残差连接和层标准化。

句向量嵌入层： 受Lin et al., (2017)的启发，本文使用一个线性结合层获取句子向量。该层通过注意力机制将整个句子中所有单词产生的隐藏状态结合在一起从而生成句子向量。句中单词映射为句子向量的权重计算方法如下:

$$\alpha = \text{softmax}\left(W^2 \tanh\left(W^1\left(S_i^{(k)}\right)^T\right)\right), \tag{3}$$

其中$W^1 \in \mathbb{R}^{d_m\times d_m}$ ，$W^2 \in \mathbb{R}^{d_m}$是模型的参数矩阵。使用前文所述编码器自注意力层输出的词级隐藏状态及计算出的映射权重获得篇章中的句子向量:

$$v_{X_i}^{(k)} = \sum_{j=1}^{n} \alpha_{i,j} s_{i,j}^{(k)}. \tag{4}$$

其中$v_{X_i}^{(k)}$表示篇章中句子X_i经过句向量嵌入层后生成的句向量，$\alpha_{i,j}$表示句子X_i中各单词映射为句向量的权重，$s_{i,j}^{(k)}$表示句子X_i中的词经过编码器自注意力层输出的隐藏状态。

权重计算： 利用句向量嵌入层的输出计算当前句中的词与篇章中所有句子间的依赖关系，公式如下:

$$u_{i,j}^{(k)} = \text{softmax}\left(s_{i,j}^{(k)} V^{(k)} / \sqrt{d_{V^{(k)}}}\right), \tag{5}$$

其中$u_{i,j}^{(k)} \in \mathbb{R}^{1\times N}$表示句子$X_i$中单词$x_{i,j}$与篇章中所有句子的依赖权重。$V^{(k)} = (v_{X_1}^{(k)}, \cdots, v_{X_N}^{(k)})$表示一个篇章$\mathcal{X}$中所有句子向量的集合。

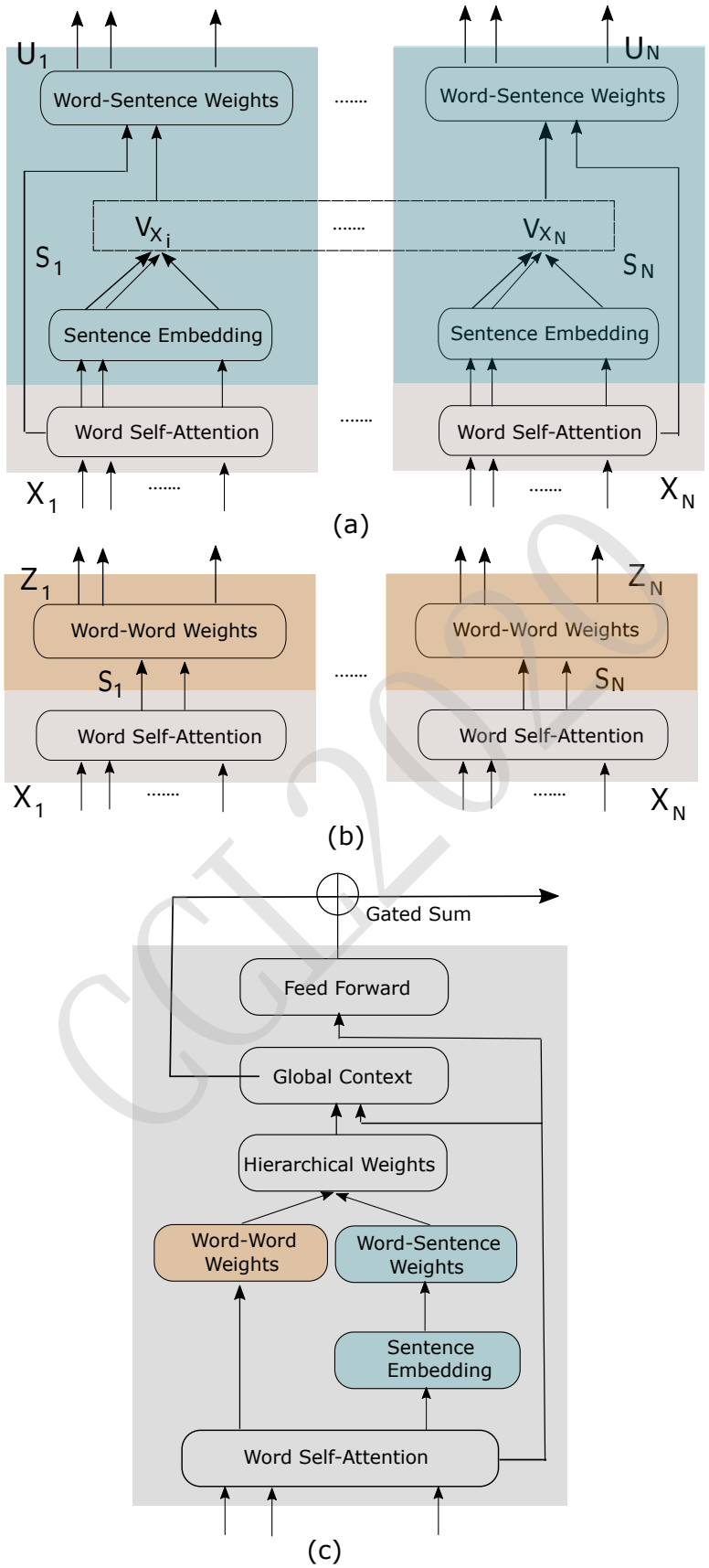

图 2: (a): 词-句级依赖权重的获取过程; (b): 词-词级依赖权重的获取过程。(c): 通过结合不同层次依赖关系获取全局上下文的过程。

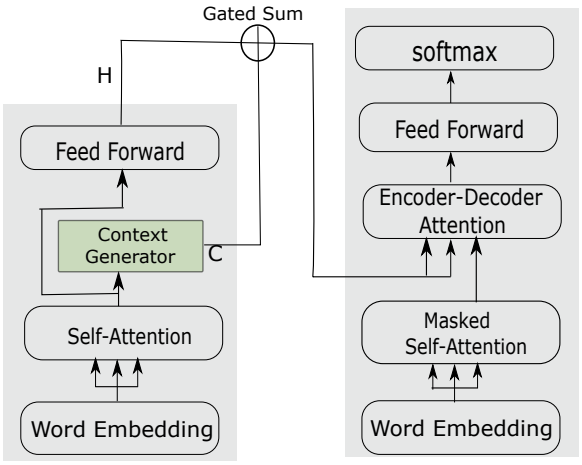

图 3: 层次化结构上下文与翻译模型的结合

2.2 词-词级依赖权重

如图2(b)所示，利用源端编码器自注意力层输出的隐藏状态，为当前句的每一个单词获取其与篇章中所有单词的依赖关系。计算公式如下：

$$z_{i,j}^{(k)} = \text{softmax}\left(s_{i,j}^{(k)}, S^{(k)}/\sqrt{d_{S^{(k)}}}\right), \tag{6}$$

其中，$M = N \times n$即篇章中所有词的个数,$z_{i,j}^{(k)} \in \mathbb{R}^{1\times M}$视为句子$X_i$中单个单词与全文单词的依赖关系权重向量。

2.3 层次化全局上下文的获取

本文所述模型使用源端篇章作为全局上下文，没有使用额外的上下文语料。为了减少计算开销，避免模型参数增加过多，我们将两个层次依赖权重的计算过程及全局上下文获取过程建立在编码器中，并共享编码器中的自注意力层参数。

如图2(c)所示，本文使用通过词-句级权重修正词-词级权重，使得最终获得的权重矩阵既含有句子层面的依赖关系，又含有整个篇章中每个单词之间的依赖关系。公式如下：

$$Q_{combine}^{i} = U_i Z_i. \tag{7}$$

其中。$U_i^{(k)} \in \mathbb{R}^{n\times N}$ 视为表示句子X_i中所有单词与篇章中其他句子之间依赖关系的权重向量，$Z_i^{(k)} \in \mathbb{R}^{n\times M}$视为篇章中句子$X_i$所有单词各自与篇章所有单词的依赖关系权重矩阵。

利用蕴含两层依赖关系的权重矩阵，将编码器自注意力层输出以篇章为单位的隐藏状态分配给当前句的每个单词。至此，当前句的每个单词都各自获取特有的蕴含自上而下不同层面依赖关系的全局上下文。

$$C_i^{hier} = Q_{combine}^{i} S^{(k)}. \tag{8}$$

其中，$C_i^{hier} \in \mathbb{R}^{n\times d_m}$即语句$X_i$获取的全局上下文。

2.4 层次化全局上下文的结合

编码器自注意力层输出的隐藏状态通过前馈全连接网络后得到翻译模型编码器输出，其表达形式如下：

$$H_i^{(k)} = \text{FNN}(S_i^{(k)}). \tag{9}$$

最终,层次化全局上下文与编码器输出通过门控单元结合。

$$H_i^{(k)} = \lambda H_i^{(k)} + (1-\lambda)\, C_i^{hier(k)}. \tag{10}$$

Set	ZH-EN		ES-EN		EN-DE	
	#SubDoc	#Sent	#SubDoc	#Sent	#SubDoc	#Sent
Training	47,758	781,524	6,531	180,853	7,491	206,126
Dev	82	1,664	33	887	326	8,967
Test	627	5,833	165	4,706	87	2,271

表 1: 训练集，开发集及测试集的统计信息

门控单元系数的计算方法如下：

$$\lambda = \text{sigmoid}\left([H_i^{(k)}; C_i^{hier(k)}]W^G\right), \tag{11}$$

其中$H_i^{(k)} \in \mathbb{R}^{n\times d_m}$是编码器经过全连接前馈神经网络层后的输出。$W^G \in \mathbb{R}^{2d_m\times d_m}$ 是模型参数矩阵。如图3所示，层次化全局上下文与编码器输出结合后进入解码器。

3 实验

本文将仅通过词-词级依赖权重获取的全局上下文称为词级上下文。将使用被词-句级权重规整过的复合权重获取的含有层次化篇章信息的全局上下文称为复合上下文。本文分别选择限定上下文获取范围及结构化上下文两类上下文结合方式中性能较好的强基线模型作为对比模型。

3.1 数据集

在中-英实验中，篇章级平行语料的训练集包括4.7万个文档中的78万个句子对[0]。我们使用NIST MT 2006数据集作为开发集，并使用MT 2002、2003、2004、2005、2008数据集作为测试集，其中测试集的合集标记为All。本文使用Jieba[1]分词将汉语句子按词切分，而英语句子则使用Moses 脚本Koehn et al., (2007)进行分词和小写处理。我们通过BPE Sennrich et al., (2016)使用3万大小的词表分别将源语言和目标语言中的单词进一步分割成子词

西班牙-英翻译任务中的训练集为IWSLT 2014和2015 Cettolo et al., (2012)，开发集为dev2010，测试集为tst2010、tst2011和tst2012。英-德翻译任务中的训练集来自IWSLT2017，本文使用tst2016和tst2017作为测试集，余下数据集作为开发集。所有数据集均使用Moses脚本进行分词和Truecasing处理。并使用3万大小的联合词表将源端及目标端语料中的单词分割成子词。由于本文词级别上下文需要计算篇章中所有词之间的依赖关系，计算开销及显存占用都十分可观。考虑到训练效率，我们将长篇章切分为最大长度为30句的段落。实验数据集的篇章数，句子数及平均篇章长度等统计信息如表1所示。

3.2 实验设置

本文基于OpneNMT[2] Klein et al., (2017)实现以平行句对为单位更新参数的基准模型Transformer，并进一步拓展为以篇章为单位更新参数的翻译模型。以篇章为单位更新使得模型可以轻易获取语料的篇章信息，从而进一步获取全局上下文。本文将模型隐藏状态的维度设为512，每个编码器解码器的层数都设置为6，多头注意力机制中的个数都设置为8，柱状搜索的大小设置为5，dropout设置为0.1。在训练过程中，我们将批大小设置为8192个字符并使用$\beta_1 = 0.1$的Adam优化器对模型进行优化Kingma and Ba., (2015)。

3.3 训练方式

受Zhang et al., (2018)启发，本文使用两步训练策略充分利用句子级平行语料在训练速度及计算开销等方面的优势。在第一步训练中使用平行句对语料对句子级别参数进行训练,在第二步训练中使用含有篇章信息的语料训练篇章级参数,该部分参数包括不同层次依赖权重的获取，含有分层结构信息全局上下文的获取，及结合上下文与编码器输出的门控等。两步训练使用的平

[0]训练集由LDC2002T01, LDC2004T07, LDC2005T06, LDC2005T10, LDC2009T02, LDC2009T15, LDC2010T03组成。

[1]https://github.com/fxsjy/jieba

[2]https://github.com/OpenNMT/OpenNMT-py

模型	MT06	MT02	MT03	MT04	MT05	MT08	All
Transformer	36.27	42.71	43.51	41.25	41.07	31.54	39.64
+ 词级上下文	37.05‡	43.79‡	44.57‡	41.98†	42.10‡	32.49‡	40.61‡
+ 复合上下文	**37.46**‡	**44.08**‡	**44.86**‡	**42.87**‡	**42.16**‡	**32.74**‡	**41.10**‡
Transformer(Zhang et al., 2018)	36.20	42.41	43.12	41.02	40.93	31.49	39.53
Transformer-DocNMT(Zhang et al., 2018)	37.12	43.29	43.70	41.42	41.84	32.36	40.22

表 2: 本文模型中-英翻译任务的性能(BLEU). † 和‡表示与Transformer基准模型相比显著性p值小于0.05/0.01

模型	西-英		英-德	
	BLEU	Meteor	BLEU	Meteor
Transformer	35.50	34.60	23.02	43.66
+ 词级上下文	37.59	36.50	24.40	45.19
+ 复合上下文	**37.75**	**36.83**	**24.98**	**45.70**
Transformer-DocNMT(Zhang et al., 2018)	37.07	36.16	24.00	44.69
HAN-DocNMT(Miculicich et al., 2018)	37.35	36.50	24.58	45.48

表 3: 本文模型在西班牙语-英语及英语-德语任务上的翻译性能(BLEU 和Meteor)

行句对语料与篇章语料是同一数据集的不同切分方式，没有引入额外语料。本文实验使用单块显存32G的Nvidia V100显卡进行训练。

3.4 评估指标

对于中-英翻译任务，本文报告了使用multi-bleu.perl脚本计算的不区分大小写的BLEU得分Papineni et al ., (2002)。对于其他翻译任务，本文报告了根据multi-bleu.perl脚本计算的区分大小写的BLEU分值和Meteor得分Lavie and Agarwal., (2007)。以上数据集和评估方法与本文比较实验的设置是一致的。我们使用paired bootstrap重采样方法评测BLEU值提升的显著性Koehn et al., (2004)。

3.5 实验结果

表2列出了汉-英翻译的性能结果。实验不但表明词级或复合上下文都能显著提高翻译性能，而且表明使用含有复合篇章信息的上下文比单独使用词级全局上下文效果更好。例如，在单一使用词级全局上下文实验中，本文方法在All测试集上的BLEU分数相比基准模型Transformer提高了0.95，在结合使用复合上下文后本文在All测试集上取得了1.36的提升。与Zhang et al., (2018)对前两句话进行建模的方法相比，在相似基准模型的前提下，本文方法(词级上下文及复合结构上下文)都取得了明显的性能提升，这表明全局上下文比当前句前两句更有助于提高文档级神经机器翻译质量。

表3列出了本文模型在西班牙-英及英-德两个篇章级翻译任务上的BLEU和Meteor得分。与中-英任务相似的是，在这两个翻译任务中利用全局上下文比只选用篇章中部分语句作为上下文更有帮助。此外，将不同层次的篇章信息结合进上下文会给翻译质量带来进一步提升。在两个翻译任务上，我们的方法相比Transformer基准模型在BLEU (Meteor)评测标准上提高了2.25(2.23)和1.96(2.04)。

4 分析与讨论

4.1 模型参数及训练时间

如表4统计数据所示，本文提出的上下文获取及结合方式由于编码器多头注意力层参数共

模型	参数（百万）
Transformer	51.3
+ 复合上下文	57.6
HAN-DocNMT(Miculicich et al., 2018)	63.0
Transformer-DocNMT(Zhang et al., 2018)	96.8

表 4: 不同模型参数比较.

享，参数增加数量较少。综合考虑了模型性能与参数及计算开销之间的平衡。同时，本文充分利用句子级平行语料在训练时间上的优势，通过两步训练法使得训练时间相比其他模型没有明显增加。

4.2 不同上下文利用方式

如图4所示，为了观察不同方式利用上下文对翻译质量的影响，我们尝试在解码器端增加专门针对全局上下文的注意力层结构。实验结果对比如表5，相比本文使用的直接融合不同层次依赖关系的方法，增加注意力层的方法增加了模型参数和计算开销，在翻译性能方面没有取得有意义的提升。

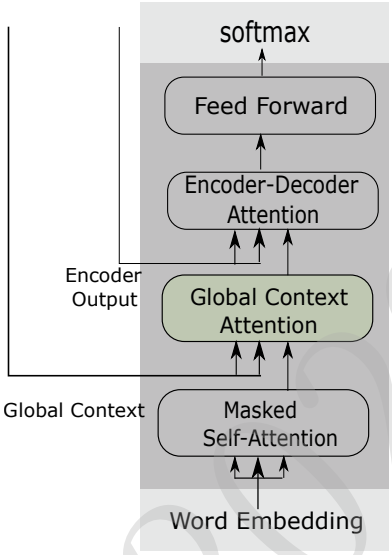

图 4: 在解码器中增加全局上下文注意力层作为上下文结合方式

结合方式	BLEU
直接融合	**41.10**
增加注意力层	41.09

表 5: 不同上下文结合方式翻译性能对比.

上下文来源	BLEU
当前句前文	40.31
当前句后文	**40.49**
无上下文	39.64

表 6: 不同来源上下文对翻译性能影响.

4.3 上下文分布

高清数字电视在美渐成主流
新华社纽约2 月12 日电(记者范小林) 随着高分辨率数字电视机(HDTV) 的售价……
刚落幕的美国“ 超级碗” 美式橄榄球赛已经成为美国广播业者迎接电视高清……。
据《金融时报》日前报道, 美国国会最近通过一项法案, 规定美国广播业者必须在……。
这意味着美国电视节目播放的全面数字化已经有了明确的时间表。
高分辨率数字电视的普及一方面要靠消费者购买电视机, 一方面……。
数字电视虽然在过去几年被一再提及, 但始终都未成为现实。
这个问题在今年出现转机。
根据美国消费电子协会的预测, 数字电视机今年在美销售量将首次超过传统电视机, ……。

表 7: 开发集篇章上下文分布样例

本文针对文中使用的全局上下文进行前后文屏蔽实验以观察当前句前文及后文各自对翻译的影响。实验结果如表6所示,二者无显著差异，屏蔽前文的翻译性能略低于屏蔽后文，这与Wong et al., (2020)的研究相符，我们推测前后文的重要性可能随语种和语料类型发生变化。本文认为该实验虽然不能直接得出当前句前后文重要性孰轻孰重的结论，但可以表明后文作为上下文对篇章翻译的重要性。

在4.2的实验中，本文使用来自整个篇章的全局上下文与翻译模型结合，其翻译质量比仅使用前文作为上下文取得了显著提升。出于探索前后上下文对篇章翻译质量影响的目的，我们对全局上下文的分布展开如下分析与实验。本文利用句子级依赖权重对开发集中的篇章句子进行统计，获取篇章中对所有句子而言都最重要的句子，并将该句使用加粗字体表示。表7中的样例可以直观表明，当前句的上下文不一定只存在于邻近语句。该现象不仅存在于本文所举样例，也存在于本文实验所使用的其他篇章语料中。

4.4 名词与代词翻译

为了观察本文提出的层次化结构全局上下文模型是如何提高翻译质量的，我们对代词和名词的翻译进行进一步的实验与分析。在代词翻译中，我们使用Miculicich et al., (2017)提出的APT度量标准评价中-英翻译实验代词翻译的准确性。如表8所示，结果表明本文提出的多层结构全局上下文模型能够更好地捕捉到每个词的全局上下文，从而提升中-英翻译实验在代词翻译的准确率。

Model	MT06	MT02	MT03	MT04	MT05	MT08	All
Transformer	69.54	73.67	68.41	65.32	67.71	71.60	68.68
+ 复合上下文	**70.24**	**74.22**	**69.02**	**65.45**	**68.29**	**71.91**	**69.40**

表 8: 代词翻译质量(APT)对比

源端	..... 今天晚上的十一二点钟左右吧。
参考翻译	..... **it** will arrive around 11 : 00 or 12 : 00 tonight .
Transformer	..... that about 11.2 pm today .
+ 词级上下文	..... **it** will be around 11.2 pm today .
+ 复合上下文	..... **it** will be around 12 : 00 tonight .

表 9: 代词翻译样例

本文在表9中列举了一个翻译例子进一步观察层次化全局上下文对代词翻译的帮助。通过实例可以看出本文提出的模型可以较好地推断出潜在代词，从而验证了该模型的代词翻译性能。对于名词翻译的分析，本文将展示另一个样例。

源端	.....一款非常优秀的基于PHP 和MySQL 数据库的社区**程序** 。
参考翻译	.....an excellent community **software** based on php and mysql database .
Transformer	.....an extremely outstanding community **procedure** based on the php..... .
+ 词级上下文	.....an extremely outstanding community **process** based on php a..... .
+ 复合上下文	.....an extremely outstanding community**programe** based on php a..... .

表 10: 名词翻译样例

表10的样例可以观察到本文提出的全局上下文比其他对比模型更好的翻译了易混淆的名词。同时不难看出相比仅使用单词级别上下文，使用层次化全局上下文对提升名词翻译质量的效果更好。

5 相关工作

(Gong et al., 2011; Hardmeier et al., 2012; Xiong et al., 2013; Tu et al., 2014; Garcia et al., 2015)在使用篇章信息提高统计翻译质量的研究领域做了大量工作。机器翻译研究热点从统计翻译转向神经翻译后不久，篇章级神经机器翻译的研究也蓬勃发展起来。根据获取上下文的范围，我们将相关研究分为两类:(1)使用部分语句作为上下文的研究;(2)使用篇章作为上下文的研究。

在第一类研究中，Tiedemann and Scherrer., (2017)基于循环神经网络(RNN)直接拼接语句作为上下文。随后(Jean et al., 2017; Wang et al., 2017; Zhang et al., 2018; Bawden et al., 2018; voita et al., 2019)的研究中，以RNNSearch和Transformer为基础使用具有不同注意力机制的多编码器提升篇章翻译质量。Miculicich et al., (2018)提出一种分层注意网络(HAN)，它通过句词的抽象表示

为当前句从前面的句子中提取上下文。Yang et al., (2019)在HAN的基础上提出一种胶囊网络将上下文信息按不同角度进行聚类。(Tu et al., 2018; Kuang et al., 2018)提出的基于缓存的方法所存储的是前面句子中的词/翻译，也归于这一类研究。

另一类研究以篇章为翻译单元，针对每个句子动态获取有用的篇章级信息。Maruf and Haffari., (2018)使用额外的存储网络将篇章转换为上下文与基于RNN的神经机器翻译模型结合。Mace and Servan., (2019)在每个源句中添加篇章标签，并将其替换为篇章级嵌入向量。Xiong et al., (2019)提出了一种二次优化策略，通过激励机制来完善第一轮翻译。Maruf et al., (2019)提出使用稀疏注意力机制选择性地捕获与当前句相关联的句子并进一步选择关键词。Tan et al., (2019)提出利用句向量之间的注意力机制获取上下文向量，并将其分配给当前句中的词。

与上述研究不同，本文提出从词-句层面和词-词层面对全局上下文进行建模从而获取复合依赖关系。使当前句的每一个词直接获取全文句子及单词中的潜在语义信息及递进关系。同时本文提出的上下文获取方式综合考量上下文范围及结合方式，既将上下文获取范围扩展至整个篇章，又没有增加额外语料或编码器。

6 总结

本文提出利用含有复合层次化篇章信息的全局上下文提升篇章级神经机器翻译质量。该模型首先为当前句中的词分别从词和句两个层面获取其篇章级依赖权重矩阵，然后通过复合权重及编码器自注意力层的输出获得全局上下文，最后将上下文与翻译模型结合。在多个基准语料数据集上的实验结果表明，与若干强基线系统相比该模型能带来显著的翻译质量提升。分析试验表明结合具有复合层次化篇章信息的全局上下文可以有助于提高篇章翻译的名词及代词翻译质量。

如何通过合理建模篇章语料中的长依赖关系获取其潜在的语义信息是一个值得不断探索的问题。我们将在未来的工作中继续对这一问题提出有意义的尝试。

参考文献

Dzmitry Bahdanau and Kyunghyun Cho and Yoshua Bengio. 2015. *Neural Machine Translation by Jointly Learning to Align and Translate*. Proceedings of ICLR.

Rachel Bawden and Rico Sennrich and Alexandra Birch and Barry Haddow. 2018. *Evaluating discourse phenomena in neural machine translation*. In Proceedings of NAACL, 1304–1313.

Matt Gardner and Joel Grus and Mark Neumann and Oyvind Tafjord and Pradeep Dasigi and Nelson Liu and Matthew Peters and Michael Schmitz and Luke Zettlemoyer. 2017. *AllenNLP: A Deep Semantic Natural Language Processing Platform*. In Proceedings of ACL Workshop for Natural Language Processing Open Source Software.

Zhengxian Gong and Min Zhang and Guodong Zhou. 2011. *Cache-based Document-level Statistical Machine Translation*. Proceedings of EMNLP, 909–919.

Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N Dauphin. 2017. *Convolutional sequence to sequence learning.*. Proceedings of the 34th International Conference on Machine Learning, 70:1243–1252.

Eva Martínez Garcia and Cristina España-Bonet and Lluís Màrquez. 2015. *Document-Level Machine Translation with Word Vector Models*. Proceedings of EAMT, 59–66.

Christian Hardmeier and Joakim Nivre and Jörg Tiedemann. 2012. *Document-Wide Decoding for Phrase-Based Statistical Machine Translation*. Proceedings of EMNLP-CoNLL, 1179–1190.

Hany Hassan and Anthony Aue and Chang Chen and Vishal Chowdhary and Jonathan Clark and Christian Federmann and Xuedong Huang and others. 2018. *Achieving Human Parity on Automatic Chinese to English News Translation*. Computing Research Repository, arXiv:1803.05567.

Sebastien Jean and Stanislas Lauly and Orhan Firat and Kyunghyun Cho. 2017. *Does neural machinetranslation benefit from larger context?*. In Computing Research Repository, arXiv:1704.05135.

Shaohui Kuang and Deyi Xiong and Weihua Luo and Guodong Zhou. 2018. *Modeling Coherence for Neural Machine Translation with Dynamic and Topic Caches*. Proceedings of COLING, 596—606.

Philipp Koehn. 2004. *Statistical significance tests for machine translation evaluation*. Proceedings of EMNLP, 388-–395.

Koehn, Philipp and Hoang, Hieu and Birch, Alexandra and Callison-Burch, Chris and Federico, Marcello and Bertoldi, Nicola and Cowan, Brooke and Shen, Wade and Moran, Christine and Zens, Richard and Dyer, Chris and Bojar, Ondřej and Constantin, Alexandra and Herbst, Evan. 2007. *Moses: Open Source Toolkit for Statistical Machine Translation*. Proceedings of ACL, (Jun):177–180.

Klein, Guillaume and Kim, Yoon and Deng, Yuntian and Senellart, Jean and Rush, Alexander. 2017. *OpenNMT: Open-Source Toolkit for Neural Machine Translation*. Proceedings of ACL, 67–72.

Diederik P. Kingma and Jimmy Ba. 2015. *Adam: A method for stochastic optimization*. Proceedings of ICLR.

Zhouhan Lin and Minwei Feng and Cicero Nogueira dos Santos and Mo Yu and Bing Xiang and Bowen Zhou and Yoshua Bengio. 2017. *A Structured Self-attentive Sentence Embedding*. Proceedings of ICLR.

Lavie, Alon and Agarwal, Abhaya. 2007. *METEOR: An Automatic Metric for MT Evaluation with High Levels of Correlation with Human Judgments*. Proceedings of WMT, (Jun):228–231.

Valentin Mace and Christophe Servan. 2019. *Using whole document context in neural machine translation*. In Proceedings of IWSLT.

Sameen Maruf and Gholamreza Haffari. 2018. *Document Context Neural Machine Translation with Memory Networks*. Proceedings of ACL, 1275–1284.

Sameen Maruf and André F. T. Martins and Gholamreza Haffari. 2019. *Selective Attention for Context-aware Neural Machine Translation*. Proceedings of NAACL, 3092–3102.

Lesly Miculicich Werlen and Andrei Popescu-Belis. 2017. *Validation of an automatic metric for the ac- curacy of pronoun translation (APT)*. Proceedings of the Third Workshop on Discourse in Machine Translation, 17–25.

Lesly Miculicich and Dhananjay Ram and Nikolaos Pappas and James Henderson. 2018. *Document-Level Neural Machine Translation with Hierarchical Attention Networks*. Proceedings of EMNLP, 2947–2954.

Mauro Cettolo and Christian Girardi and Marcello Federico. 2012. *WIT3: Web Inventory of Transcribed and Translated Talks*. Proceedings of EAMT, 261–268.

Kishore Papineni and Salim Roukos and Ward Todd and Wei-Jing Zhu. 2002. *BLEU: a method for automatic evaluation of machine translation*. Proceedings of ACL, 311–318.

Rico Sennrich and Barry Haddow and Alexandra Birch. 2016. *Neural Machine Translation of Rare Words with Subword Units*. In Proceedings of ACL, 1715–1725.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. *Sequence to sequence learning with neural networks*. Advances in neural information processing systems, 3104-–3112.

Xin Tan and Longyin Zhang and Deyi Xiong and Guodong Zhou. 2019. *Hierarchical Modeling of Global Context for Document-Level Neural Machine Translation*. In Proceedings of EMNLP-IJCNLP, 1576–1585.

Mei Tu and Yu Zhou and Chengqing Zong. 2014. *Enhancing Grammatical Cohesion: Generating Transitional Expressions for SMT*. Proceedings of ACL, 850–860.

Tiedemann, Jörg and Scherrer, Yves. 2017. *Neural Machine Translation with Extended Context*. In Proceedings of the Third Workshop on Discourse in Machine Translation, ”82–92.

Zhaopeng Tu and Yang Liu and Shuming Shi and Tong Zhang. 2018. *Transactions of the Association for Computational Linguistics*. Transactions of the Association for Computational Linguistics, (6):407–420.

Ashish Vaswani and Noam Shazeer and Niki Parmar and Jakob Uszkoreit and Llion Jones and Aidan N.Gomez and Lukasz Kaiser and Illia Polosukhin. 2017. *Attention Is All You Need*. In Proceedings of NIPS, 5998–6008.

Elena Voita and Pavel Serdyukov and Rico Sennrich and Ivan Titov. 2018. *Context-Aware Neural Machine Translation Learns Anaphora Resolution*. Proceedings of ACL, 1264–1274.

Elena Voita and Rico Sennrich and Ivan Titov. 2019. *When a Good Translation is Wrong in Context: Context-Aware Machine Translation Improves on Deixis, Ellipsis, and Lexical Cohesion*. Proceedings of ACL, 1198–1212.

KayYen Wong and Sameen Maruf and Gholamreza Haffari. 2020. *Contextual Neural Machine Translation Improves Translation of Cataphoric Pronouns*. In Proceedings of ACL, 2826–2831.

Longyue Wang and Zhaopeng Tu and Andy Way and Qun Liu. 2017. *Exploiting cross-sentence context for neural machine translation*. In Proceedings of EMNLP, 2826–2831.

Lesly Miculicich Werlen and Andrei Popescu-Belis. 2017. *Validation of an Automatic Metric for the Accuracy of Pronoun Translation (APT)*. Proceedings of Workshop on Discourse in Machine Translation, 17–25.

Deyi Xiong and Yang Ding and Min Zhang and Chew Lim Tan. 2013. *Lexical Chain Based Cohesion Models for Document-Level Statistical Machine Translation*. Proceedings of EMNLP, 1563–1573.

Hao Xiong and Zhongjun He and Hua Wu and Haifeng Wang. 2019. *Modeling coherence for discourse neural machine translation*. In Proceedings of AAAI, 7338–7345.

Zhengxin Yang and Jinchao Zhang and Fandong Meng and Shuhao Gu and Yang Feng and Jie Zhou. 2019. *Enhancing Context Modeling with a Query-Guided Capsule Network for Document-level Translation*. Proceedings of EMNLP, 1527–1537.

Jiacheng Zhang and Huanbo Luan and Maosong Sun and Feifei Zhai and Jingfang Xu and Min Zhang and Yang Liu. 2018. *Improving the Transformer Translation Model with Document-Level Context*. Proceedings of EMNLP, 533–542.

第十九届中国计算语言学大会论文集，第434页-第445页，海南，中国，2020年10月30日至11月1日。