

融合零指代识别的篇章级机器翻译

汪浩, 李军辉, 贡正仙

(苏州大学计算机科学与技术学院, 江苏苏州 215006)

摘要

在汉语等其他有省略代词习惯的语言中, 通常会删掉可从上下文信息推断出的代词。尽管以Transformer为代表的神经机器翻译模型取得了巨大的成功, 但这种省略现象依旧对神经机器翻译模型造成了很大的挑战。本文在Transformer模型基础上提出了一个融合零指代识别的翻译模型, 并引入篇章上下文来丰富指代信息。具体地, 该模型采用联合学习的框架, 在翻译模型基础上, 联合了一个分类任务, 即判别句子中省略代词在句子所表示的成分, 使得模型能够融合零指代信息辅助翻译。通过在中英对话数据集上的实验, 验证了本文提出方法的有效性, 与基准模型相比, 翻译性能提升了1.48个BLEU值。

关键词: 零指代; 篇章级机器翻译; 联合学习; Transformer

Context-aware Machine Translation Integrating Zero Pronoun Recognition

WANG Hao, LI Junhui, GONG Zhengxian

(School of Computer Science and Technology,
Soochow University, Suzhou, Jiangsu 215006)

Abstract

Pronouns are usually omitted if it can be inferred from context in pro-drop languages, such as Chinese. Although neural machine translation (NMT) models, e.g., Transformer, have achieved great success, this omission phenomenon poses a significant challenge for NMT. This paper proposes a model that integrates zero pronoun information based on Transformer, and introduces context to enrich anaphora information. This model adopts the framework of joint learning, combined a classification task to identify the omitted component of the sentence, so that model can use zero pronoun information to help translation. Through experiments on a Chinese-English dialogue dataset, proved the effectiveness of the proposed method. Compared with the benchmark model, our model obtains an improvement of 1.48 BLEU.

Keywords: Zero Pronoun, Context-aware Machine Translation, joint learning, Transformer

1 引言

代词在自然语言中占有重要地位，它蕴涵着丰富的语篇信息。在汉语、日语等语言中，如果可以从周围的语境或对话中推断出句子中的代词，这样的代词往往会被省略，使得句子更加简洁精炼。这种特殊的语法现象被称为零指代现象，这些语言被称为代词省略语言(pro-drop languages)。这种充满口语化的表达虽然不会对人类理解语言造成困难，但当计算机处理这些语言的文本时，对文本理解可能会出现极大的偏差。在机器翻译任务中，这种不适应现象尤其显著。当把代词省略语言的句子翻译成非代词省略语言的时候（例如，从汉语到英语），机器翻译系统会因为无法正确翻译出省略的代词（dropped pronouns, DP），而导致翻译性能明显下降。这个问题在非正式的文本领域中尤其严重，如对话和会谈等，因为在这些领域中为了使句子更加紧凑，代词经常被省略 (Yang et al., 2015)。Wang等 (2018)分析一个大规模中英对话语料库时表明，在该语料中文端的文本中，大约有26%的代词被省略了。由此可见，解决省略代词翻译难题对非正式文本领域的机器翻译任务具有重大的意义。

先前的研究人员在统计机器翻译模型和传统的神经机器翻译模型基础上做了一些缓解省略代词翻译错误的研究。例如，Wang等 (2016)首次提出利用平行语料库中的对齐信息自动标注出省略代词的方法，较大程度上恢复了中文对话语料的省略代词，使得翻译性能取得了非常显著的提升。Tan等 (2019)利用特殊的标注方法，将标注的省略代词的翻译作为外部词汇知识整合到神经网络翻译中。但由于零指代机器翻译领域一直受到的关注不足，导致研究人员尚未基于最为先进的机器翻译模型—Transformer (Vaswani et al., 2017)来解决省略代词翻译的难题，相关研究一直存在空白。受此启发，本文首次将机器翻译模型Transformer引入零指代机器翻译领域。Transformer虽然能够利用独特的多头注意力机制来捕获更多的语义信息，但面对看不见的省略代词也只能翻译出一些简单的部分，仍然无法准确地翻译出更复杂句子中的省略代词。特别地，由于零指代的识别往往需要篇章上下文信息，因此本文先通过往模型中引入篇章上下文信息，再进行零指代识别。同时，本文将零指代识别看作是一个分类任务，即识别零指代在句子中担当的句法成分。

本文利用联合学习的架构，将分类任务和翻译任务联合起来，在不改变翻译模型的基础上，向翻译模型隐式地加入有效的零指代信息，缓解省略代词翻译错误问题。同样，我们也期望分类任务和翻译任务可以相互作用：分类为翻译提供更多的零指代信息，翻译帮助分类解决歧义等问题。此外，先前的研究表明篇章信息能够很好地处理零指代消解问题 (Zhao and Ng, 2007; Chen and Ng, 2013)，而且篇章信息早已被证明能够显著地提高机器翻译的性能，解决译文不通顺、不连贯的现象 (Xiong et al., 2019; Zhang et al., 2018; Bawden et al., 2018)。Wang等 (2019)使用层次神经网络来总结文本中当前句子的上下文，并将其集成到预测省略代词的联合模型中，取得了不错的效果。但是Wang并未充分利用篇章信息来同时提高翻译与预测任务的性能。本文的模型更进一步，利用联合学习共享模块的结构和注意力机制，充分利用丰富的篇章信息来同时提升两个任务的效果。本文提出的方法在大规模中英对话语料库上的实验结果表明，该方法可以显著提高省略代词的翻译性能，与本文中强大的基准模型相比，翻译性能提高了1.48个BLEU值。

本文的主要贡献有：

- (1) 本文首次在零指代机器翻译任务中引入当前最为先进的机器翻译建模模型Transformer，并取得显著的效果。
- (2) 本文提出了一个简单的模型来共同学习翻译任务和省略代词的分类任务，并且证明了分类任务可以辅助提升翻译任务的性能。
- (3) 本文验证了篇章上下文信息对机器翻译任务和零指代任务的有效性。

2 相关工作

零指代机器翻译 零指代机器翻译离不开省略代词的恢复任务，前人对省略代词的恢复研究取得了很大的进展。Yang等 (2015)首次提出在中文文本信息中恢复省略代词，通过训练一个17类的最大熵分类器来判断句子中省略代词的类别。Wang等 (2016)首次提出利用平行语料的对齐关系来恢复省略代词的方法，取得了非常卓越的效果。但Wang提出的方法存在着天然的缺陷，解码过程中由于无法看到目标端的句子，因此无法利用对齐关系来恢复测试集的源端省略代词，转而利用已恢复的语料训练生成器生成省略代词，但该模型性能较低，会造成错误传播，影响到翻译任务的性能。Wang等 (2018; 2019)在联合学习的架构基础上，利用一种重构

的方式将编码器和解码器输出表征重构回恢复省略代词后的源端句子，解码时由于翻译模块参数已固定而且输入源端句子为未恢复省略代词的句子，不会受到省略代词的错误预测的影响，取得了不错的效果。本文在前人的基础上，引入了更为先进的机器翻译建模模型Transformer，在翻译任务的基础上，联合了更为有效的句法成分识别任务，除此之外，进一步加入篇章上下文，使得零指代机器翻译取得更好的效果。

篇章翻译 主流的机器翻译关注句子级别的翻译，这种逐句翻译的方式在句子层次上可能是完美的，但是忽略掉了篇章之内句子之间的上下文信息，会导致翻译句子缺乏流畅性和逻辑性，篇章级翻译的兴起有效地改善了这些问题。Wang等 (2017)提出了多层次循环神经网络，总结句子之间的篇章信息，并利用不同的策略将这些篇章信息集成到标准NMT模型中。Zhang等 (2018)采用多编码器的结构，通过额外的编码器来对上下文信息进行特征抽取，与源语言信息融合用于目标语言生成。Maruf等 (2019)在Transformer的基础上，将篇章级上下文表征和字符级注意力表征相结合，并将这种多层上下文注意力模块融入编码器和解码器中。Yang等 (2019)提出了一种新的以查询为导向的胶囊网络，该网络利用改进的动态路由算法来增强篇章级神经机器翻译的性能。考虑到主流篇章级翻译模型通常采用一个额外的上下文编码器，Chen等 (2021a)使用大规模源端单语篇章对该上下文编码器进行预训练。Chen等 (2021b)进一步使用大规模源端单词篇章和平行句对大幅度提高篇章翻译的性能。研究者在基于源端上下文建模的方式上取得了显著的成果，而基于目标端上下文建模的方式同样也受到很多关注。Voita等 (2019)和Xiong等 (2019)采用两阶段训练方法，第一阶段生成句子级目标端句子，第二阶段从生成的译文中抽取篇章信息，进行译文的二次生成，都取得了不错的提升。与之前的工作相比较，本文为了减少算力消耗和参数量，采用共享参数的多编码器结构来抽取篇章信息，由于模型的多个任务共享编码器模块，因此本文只是在编码器中融入篇章信息。

3 背景

3.1 省略代词

在代词省略语言中，如汉语和日语，零指代出现的频率明显要高于非代词省略语言如英语 (Zhao and Ng, 2007)。由于代词在语篇中包含了丰富的指代信息，而且对话领域中的句子一般都较短，零指代现象不仅导致代词的翻译缺失，而且损害了输出的句子结构甚至语义。如表1展示的是有省略代词句子的翻译实例，Source和Ref表示源端句子和参考译文，NMT表示标准Transformer的输出译文。从源端句子可以看出，例1中的主语代词“我们”和例2中的宾语代词“它”都被省略了，但在目标端翻译中这些代词是不可缺少的。对人类来说，理解源端具有省略成分的句子不是问题，因为我们可以很容易地从上下文中推理出这些省略的代词。如例2展示，代词“它”是一个指示代词，指示着前一句中的先行词“蛋糕”，人类可以轻易地从前一句推断出后一句省略了“它”，但这对机器来说是非常困难的。如例1展示，人类也能够从全句推断出句子缺少了主语，然后根据具体语义推断主语代词为“我们”，这同样考验着机器的理解能力。

具体地，省略代词的存在对代词省略语言到非代词省略语言的翻译模型造成了巨大的挑战。表1中的例子所示，NMT是使用当前最先进的机器翻译建模模型Transformer生成的译文，例子1中Transformer依旧无法成功将源端缺失的代词翻译出来，这也造成目标端译文不够完整，语义不通顺。例子2中NMT由于无法捕捉到省略代词的信息，导致将源端句子翻译成“did you do”，不仅译文翻译不完整而且出现了错误翻译的现象。

例1	Source	但 (我们) 以后就不能半夜起来促膝长谈未来了。
	Ref	we won't be able to have those long talks at night about future .
	NMT	but don't get up at night and talk to the future .
例2	Source	这顿晚餐很美味！你做的 (它) 么？
	Ref	the dinner was so delicious ! did you make it ?
	NMT	the dinner was delicious ! did you do ?

表 1: 省略代词的翻译实例；括号中代词指源端句子中的省略代词。

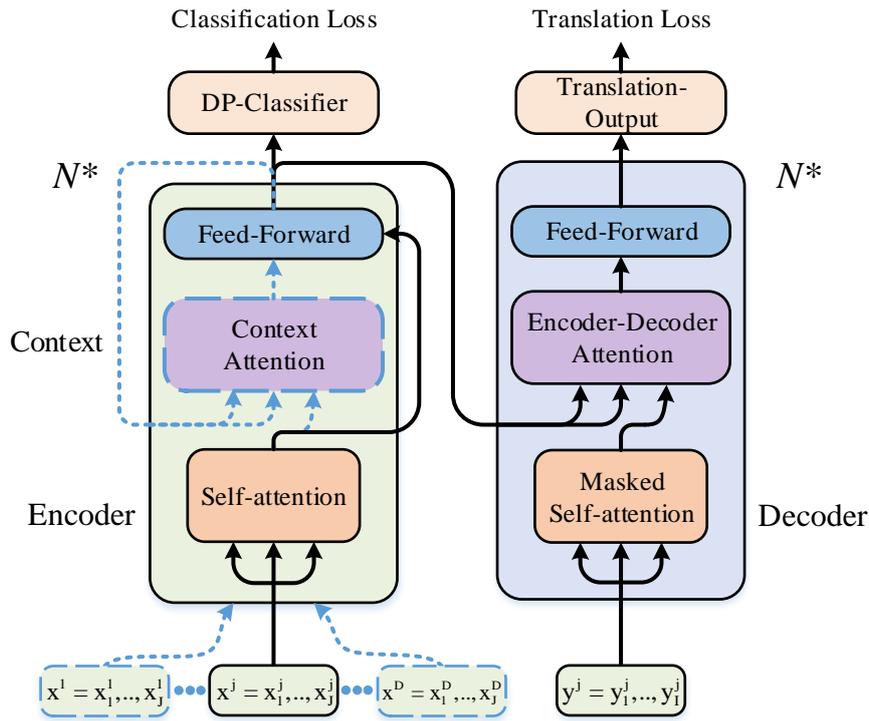


图 1: 融合零指代识别的神经机器翻译模型

3.2 神经机器翻译模型

句子级神经机器翻译模型 机器翻译的目标是将输入的源语言句子转化为目标语言句子，经典的方法是通过多分类模型对源端句子进行建模，并通过自回归模型生成目标端句子的分布。设源端句子为 $\mathbf{x} = x_1, x_2, \dots, x_J$ ，目标端句子为 $\mathbf{y} = y_1, y_2, \dots, y_I$ ，其中 $\mathbf{y}_{<t}$ 表示前 $t - 1$ 个已经预测的目标单词，条件概率 $P(\mathbf{y} | \mathbf{x})$ 代表整句话被翻译的概率，通过翻译模型对目标端句子进行建模，如式 (1) 所示：

$$P_{\theta}(\mathbf{y} | \mathbf{x}) = \prod_{t=1}^I P_{\theta}(y_t | \mathbf{y}_{<t}, \mathbf{x}) \quad (1)$$

而神经机器翻译 (NMT) 的目标是通过深度神经网络的编码器解码器结构构造这个目标函数，其中 θ 为神经网络的超参数。

篇章级神经机器翻译模型 基于篇章上下文的机器翻译任务则是在句子级翻译建模模型基础上，结合同一篇章内部分或所有上下文信息来进行建模，如式 (2) 所示：

$$P_{\theta}(\mathbf{y} | \mathbf{x}) = \prod_{j=1}^D \prod_{t=1}^I P_{\theta}(y_t^{(j)} | \mathbf{y}_{<t}^{(j)}, \mathbf{x}^{(j)}, \mathbf{D}_{-j}) \quad (2)$$

其中 \mathbf{D} 表示对应篇章中所有句子的集合， \mathbf{D}_{-j} 表示篇章中除去第 j 个句子的其他句子。

4 融合零指代识别的篇章级神经机器翻译

4.1 零指代识别任务

给定一个句子，传统的零指代识别需要识别零指代出现的位置，以及缺失的代词。一方面，考虑到零指代出现的位置非常具有歧义性，如在句子“今天要不要去电影”中，省略的代词既可以出现在“今天”前，也可以出现在其后。另一方面，考虑到省略的代词种类繁多，如实验数据训练集中省略的代词种类达28种。

而本文提出的零指代识别任务是指利用分类器模块对源端句子中省略代词所表示的句法成分进行识别。本文模型中分类器的真实标签来自于识别出省略代词后的句子。我们先识别原始句子中的省略代词，训练集使用词对齐的方式识别省略代词，测试集则是使用Wang等 (2016)提出的省略代词生成器自动识别省略代词，然后利用开源的依存句法分析库DDParser⁰分析这些省略代词的句法成分，统计并归类，最终确定了分类器分类的四个标签：缺失主语类，缺失宾语类、缺失定语类以及不缺失代词类，标签实例如表2所示。

Category	Sentences
缺失主语	Source: (你) 要不要去看电影? Ref: do you want to go see a movie ?
缺失宾语	Source: 我不能收下(它)。 Ref: i can't keep it .
缺失定语	Source: (你的) 肩膀 怎么样了? Ref: how's your shoulder ?
不缺失	Source: 我要跟我哥走了。 Ref: i'm going with my brother .

表 2: 分类任务的标签实例

4.2 模型结构

本文的出发点是如何更加有效地利用零指代信息来提升翻译模型的效果。本文提出的模型如Figure 1所示，模型框架是基于Transformer的序列到序列模型，Transformer采用多头注意力机制 (multi-head attention)，能够直接对序列中任意位置之间的关系进行建模，这很好地解决了长距离依赖问题，进行多头注意力计算时，需要将输入表征分别处理成Query (Q)，Key (K)，Value (V)，然后利用这三个表征进行注意力的计算，该过程用式 (3) ~ (5) 表示如下：

$$\text{Attention}(Q, K, V) = \text{Soft max} \left(\frac{QK^T}{\sqrt{d^k}} \right) V \quad (3)$$

其中 d^k 表示为表征K的维度。

$$\text{head}_i = \text{Attention} \left(QW_i^Q, KW_i^K, VW_i^V \right) \quad (4)$$

其中 W_i^Q ， W_i^K ， W_i^V 分别表示三个表征的权重参数矩阵。

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) W^O \quad (5)$$

其中 h 表示总头数， W^O 表示多头拼接之后进行线性变换的权重矩阵。

本文模型在Transformer的基础上进行了适应性改进，整体模型是由编码器 (Encoder)，解码器 (Decoder) 和分类器三个模块构成。

编码器 我们对标准的Transformer编码器进行了改进。编码器是由多层完全相同的子模块堆叠而成，每一个子模块是由一个多头自注意力子层，一个多头上下文注意力子层和一个全连接前馈神经网络构成。与标准的Transformer编码器不同，本文额外引入了上下文注意力子层来融合篇章上下文信息，如图1中虚线所指就是上下文信息的编码过程，编码端的输入是由当前句及其上下文组成。为了减少计算成本，本文采用共享编码器的结构，编码器先编码当前句的上下文句子，上下文注意力子层不参与此过程，第二步编码当前句时三个子层同时参与计算，上下文注意力子层的输入K，V来自于第一步中上下文的编码输出，Q来自于当前句自注意力子层的输出，当前句的编码输出结果输入到解码端进行下一步计算。

解码器 解码器与标准的Transformer解码器一致，由多个完全相同的子模块堆叠而成，每一个子模块是由一个多头掩码自注意力子层，一个多头上下文注意力子层和一个全连接前馈神

⁰<https://gitee.com/baidu/DDParser>

经网络构成。解码端输入只有源端当前句的对应目标端句子，这与编码端的输入不同。解码器的输出会被映射到目标端词表的空间，利用softmax函数计算词表中各个词对应的预测概率，最终将预测结果与真实结果计算损失。

分类器 该分类器是由两层全连接层加上sigmoid函数构成，输出是四元标签分类的结果。由于分类器的分类目标是基于句子级的向量表征，本文将编码器的输出表征通过Max-pooling和Mean-pooling操作之后拼接在一起构成分类器的输入。最终，分类器的输出与真实的标签计算损失。

4.3 联合学习

本文模型采用联合学习的训练方式，分为翻译任务和分类任务。模型的损失函数分为两部分，包括神经机器翻译模型的翻译损失和省略代词的分类损失。

翻译目标端的相关损失函数如式（6）所示：

$$\begin{aligned} L_1 &= - \sum_{n=1}^D \sum_{t=1}^{S_n} \log p \left(y^{(t)} \mid X^n, Y_{<t}^n; \theta \right) \\ &= - \sum_{n=1}^D \sum_{t=1}^{S_n} \sum_j^{w_{n,t}} \log p \left(y_j^{n,(t)} \mid x^{n,(t)}, c^{n,(t)}, y_{<j}^{n,(t)}; \theta \right) \end{aligned} \quad (6)$$

其中D表示训练集中总平行篇章对数， S_n 表示第n个平行篇章对中的句子总数， X^n 和 Y^n 表示第n个平行篇章对的源端句子和目标端句子。 $w_{n,t}$ 表示第n个平行篇章对的第t个句子的总token数， $c^{n,(t)}$ 表示引入的上下文信息， θ 表示模型的训练参数。

零指代分类损失如式（7）所示：

$$L_2 = - \frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C y_c^{(n)} \log \hat{y}_c^{(n)} \quad (7)$$

其中N为零指代分类任务的训练样例数，C表示类别标签数。 \hat{y}_c 表示模型预测类别c的概率。

最终，本文的联合学习的训练目标如式（8）所示：

$$\theta^* = \arg \min_{\theta} (L_1 + \alpha L_2) \quad (8)$$

其中 α 是零指代分类损失的权重参数，本文模型中设置 α 为1.0。

Data	Language	#Sentences	#Pronouns	#DP
Training	zh	2.15M	1.66M	0.39M
	en		2.26M	
Dev	zh	1.09K	0.76K	0.50k
	en		1.03K	
Test	zh	1.15K	0.77K	0.27k
	en		0.96K	

表 3: 对话数据集的统计情况

5 实验

5.1 实验设置

本文在大规模中英对话语料库上进行验证实验。该对话数据集是由电影或电视剧字幕构成，训练集、验证集和测试集分别包含2.15M，1.09K和1.15K个平行句对，验证集和测试集是随机选择两集完整的电视剧节目字幕挑选出来的，表3列出了语料的统计情况，该语料中句子一般较短，源语言和目标语言的代词数量差距较大，这体现了代词省略的问题，表中也展示了训练集、验证集和测试集中省略代词（DP）的数量。本文采用Jieba分词工具¹对中文句子进行

¹<https://github.com/fxsjy/jieba>

分词，而目标端英文句子则是使用Moses脚本 (Koehn et al., 2007)进行分词处理。由于原数据集中保留了篇章信息，但是没有生成独立的篇章，我们手动将数据集切割成独立的篇章，每个篇章中包含7个句子。本文对中英文数据分别采用了子词化 (Sennrich et al., 2016)，子词化操作数为30K。翻译任务的评估指标为case-insensitive 4-gram NIST BLEU (Papineni et al., 2002)，由mteval-v13a.pl测试脚本计算得出，此外本文还对模型翻译性能的BLEU值提升进行了显著性测试 (Koehn, 2004)。

本文的建模模型Transformer来源于OpenNMT²，在其基础上加以改进，将其改进为以篇章为单位更新参数的模型。本文将编码器和解码器设置为6层，多头注意力机制中含有8个头，同时设置Dropout值为0.1，隐层维度和前馈神经网络中间层单元数分别为512和2048，学习率设置为1.0。选择的优化器为Adam优化函数 (Kingma and Ba, 2015)，其中 β_1 为0.9， β_2 为0.998。训练时批处理大小为4096个token以内。进行解码时，设置beam size为5，所有其他的设置采用Vaswani系统中 (Vaswani et al., 2017)的默认设置。

5.2 实验结果

为了验证本文提出方法和模型的有效性，本文在中英翻译任务上进行实验。本文实验包括四个模型，如表4中所示分别为：（1）基准模型是由没有标注出省略代词的中英双语语料使用标准Transformer模型训练，在同样没有经过标注处理的测试集上进行测试，由表中的Baseline表示；（2）+document-level context 表示在Baseline基础上，在源端编码器中融合篇章上下文信息，本文中选取的篇章上下文为两句；（3）+DP-parse表示在Baseline基础上，联合判断省略代词在句子中句法成分的分类任务，测试集利用的省略代词信息是由Wang (2016)提出的代词生成器模型自动生成的；（4）+document-level context +DP-parse表示在联合任务的基础上，向模型加入源端上下文信息所取得的翻译效果。

本文提出模型的实验效果 表4展示出本文提出模型在中英对话测试集上的性能结果。

- 与基准模型相比，只联合省略代词的分类任务后取得了0.88的BLEU值提升，这证明了本文使用联合学习方法的有效性。
- 在进一步向模型中加入篇章上下文后，翻译任务的BLEU取得了1.48的明显提升，最终测试集BLEU值达到了37.40，分类任务的准确率也提升了约3个点，这也表明篇章上下文信息不仅能够提升翻译效果，对省略代词的分类任务也有积极的作用。

与先前研究的对比 表4中展示了之前经典的零指代机器翻译研究的模型性能。

- DPes-ref通过LSTM+CRF的标注模型，恢复了对话语料中的省略代词，并用等价的目标语言代词代替省略代词的方式，有效地提高了翻译的性能，但是由于该机器翻译建模采用为基于rnn的序列到序列模型，最终模型性能只与标准Transformer模型性能相一致，这也显示出Transformer建模模型的强大之处。
- Shared-Rec研究是通过联合重构器模块的方式，利用编码器和解码器的输出来重新解析省略代词信息。ZP-joint是在Shared-Rec的基础之上额外联合一个序列标注任务，最终取得了37.11的性能。
- 与之前模型相比较，本文提出的模型在参数量远小于其他模型的情况下，依旧取得了37.40的性能，比之前该大规模对话语料上最好效果还要提高了0.29个BLEU，证明了本文提出的模型不仅具有优异的性能还具有消耗算力少，训练时间短的优势。

5.3 代词恢复质量对翻译性能的影响

显式融合零指代信息指直接将识别出的省略代词加入到原始句子，然后利用识别后的句子训练模型，由于训练集利用词对齐识别代词方法的准确率远高于测试集利用代词生成器的识别方式，这导致使用自动识别的测试集测试时，错误的省略代词会严重影响翻译的性能。针对显式融合零指代信息方式对翻译性能的影响，本文在两种训练集和三种不同的测试集上做了对比实验。如表5中所示：（1）训练集包括未识别原始训练集中省略代词和利用词对齐自动识别原

²<https://github.com/OpenNMT/OpenNMT-py>

Model	#Params	BLEU	Accuracy(%)
Baseline	46M	35.92	/
+document-level context	52M	36.76 [†] (+0.84)	/
+DP-parse	47M	36.80 [†] (+0.88)	76.54
+document-level context+DP-parse	53M	37.40[†] (+1.48)	79.72
Previous studies			
DPEs-ref (Tan et al., 2019)	/	35.98	/
Shared-rec (Wang et al., 2018)	178.6M	36.53	/
ZP-joint (Wang et al., 2019)	143.3M	37.11	/

表 4: 中英对话翻译实验结果; †表示与Baseline相比较, BLEU值在 $p=0.01$ 时具有显著性提高。

始训练集中省略代词两种; (2) 测试集包括未识别原始测试集中的省略代词, 自动识别省略代词和人工标注省略代词三种。该对比实验使用模型都是标准Transformer训练, 参数都保持一致。

分析模型在不同数据集组合上的翻译效果后, 得到以下结论:

- 在未识别省略代词训练集的实验中, 使用自动识别和人工识别的测试集都可以提高最终的测试效果, 但自动识别的方式提高了0.93个BLEU值, 而人工识别提高了3.75个BLEU值, 这证明了识别省略代词的效果越好, 翻译性能提升越巨大, 进一步证明零指代翻译具有很大的研究前景。
- 在自动识别省略代词训练集的实验中, 由于训练数据中带有省略代词的信息, 因此使用未识别的测试集时, 训练和测试存在不一致, 性能会出现巨大的下降; 使用自动识别的测试集时, 与未识别出省略代词训练集组相比, 性能下降了0.6个BLEU值, 本文认为这是由于词对齐识别的准确率很高, 自动识别代词错误率很高, 错误的省略代词影响了句子的原本语义, 导致识别出的省略代词没有对翻译产生积极影响; 使用人工识别的测试集时, 与未识别出省略代词训练集组相比, 提升了0.7个BLEU, 达到了40.37的BLEU值性能, 本文认为这是由于训练数据和测试数据的识别质量都很高, 因此可以达到更好的效果。

上述表明, 在无平行数据条件下提高省略代词的识别质量是零指代机器翻译的重难点, 这也是我们未来研究的重点。

Training set	Test set	BLEU
未识别省略代词	未识别省略代词	35.92
	自动识别省略代词	36.85
	人工标注省略代词	39.67
自动识别省略代词	未识别省略代词	33.33
	自动识别省略代词	36.25
	人工标注省略代词	40.37

表 5: 代词识别质量对翻译性能影响的实验结果

5.4 代词翻译的性能

参考Miculicich等 (2017; 2018)的研究, 本文对测试集中代词的翻译效果进行分析实验。如表6所示, 展示的为本文提出模型在中英对话测试集上的代词翻译准确率。由实验结果可以看出, 与Baseline的代词翻译效果相比较, 提升了0.97个百分点, 证明了本文提出的融合零指代信息和篇章信息的模型能够有效地提升指代翻译的性能,

Model	Accuracy(%)
Baseline	64.78
+document-level context+DP-parse	65.75 (+0.97)

表 6: 代词翻译性能的实验结果

5.5 篇章长度对翻译性能的影响

本文通过引入篇章上下文信息提升了翻译性能。为了探究不同的篇章长度对翻译性能的影响，本文做了相关的对比实验。如表7所示，展示为本文提出模型在不同长度的篇章信息下的翻译性能。由实验结果可以看出，当篇章句子长度为2时模型取得最佳的翻译效果，使用更多的篇章句子时不仅使得翻译效果下降，而且还增加了计算代价。这与Zhang等 (2018)的结论相一致，远距离篇章上下文信息的影响是有限的。因此本文采用的篇章上下文句子长度为2句。

#Sent	BLEU
1	36.62
2	37.40
3	36.33
4	37.03

表 7: 不同的篇章长度对翻译性能影响的实验结果

5.6 有省略句子与无省略句子翻译性能对比

虽然本文实验使用的对话数据集句子中存在很多的省略代词，但是数据集中大部分句子都是完整无省略代词的句子，为了探究无省略代词和有省略代词句子的翻译性能的差距，本文做了相关的对比实验。如表8中所示，我们根据句子是否存在省略代词将原始的测试集分成了无省略测试集表示为NDP，和有省略测试集表示为DP，分别使用标准的Transformer模型和本文提出的模型进行测试。

分析模型在两种不同数据集上的测试效果后，得到以下结论：

- 在无省略测试集上，Baseline和本文提出模型得到的效果都高于全部的原始测试集，这证明了无省略代词的翻译能够取得更好的效果，本文提出模型达到的效果比Baseline提高了1.27个BLEU值，我们认为我们的模型融合的篇章信息中除了零指代信息之外，还存在其他的语义信息能够帮助提高翻译的效果。
- 在有省略测试集上，Baseline的翻译效果出现了明显的下降，而我们的模型虽然也出现了下降，但是达到的效果与Baseline相比，提高了1.8个BLEU值，这表明标准的Transformer难以很好地处理省略代词的翻译，而本文提出的模型可以有效地缓解省略代词的翻译错误问题。

Model	NDP	DP	ALL
Baseline	36.28	34.87	35.92
+document-level context+DP-parse	37.55 (+1.27)	36.67 (+1.80)	37.40 (+1.48)

表 8: 有省略句子和无省略句子翻译对比实验结果

5.7 实例分析

本章节中，本文展示出一些使用本文提出的模型识别出省略代词的例子，进一步证明本文提出方法的有效性。如表9所示，Source表示源端未识别代词的句子；Context表示源端的

上下文，本文模型采用是两句上下文，以“||”分割；Ref表示目标端参考译文；Baseline是使用Transformer模型得到译文；Our表示使用本文提出模型得到的译文。

例1和例2中，源端句子存在省略代词“你”和“我们”，Baseline由于没有融合零指代信息，所以无法成功翻译出you和we，而Our模型正确地翻译出you和we，这表明了本文提出方法模型能有效捕捉省略代词的信息。例3和例4的实例中虽然也能够捕捉到零指代信息，却没有正确地翻译出代词，我们认为这是由于有些省略代词是需要充分结合上下文才能推断得出，本文模型虽然融入了篇章信息，但是模型预测复杂的零指代关系时，还存在着一些不足，无法从篇章信息获取对应的零指代信息，这也需要进一步对篇章信息进行加工处理，融合与零指代关系更有效的篇章信息。

例1	Source Context Ref Baseline Our	(你) 要不要去看电影? 没问题, 手续都办好了 罗斯, 谢谢你 do you want to go see a movie ? want to go see a movie ? would you like to go to a movie ?
例2	Source Context Ref Baseline Our	(我们) 需要你 和 瑞秋 向 法官 说明 必须有 证人 证明 你们 当时 神智不清 没问题 we'll need you and rachel to testify . you and rachel need to talk to the judge . we need you and rachel to tell the judge .
例3	Source Context Ref Baseline Our	(我们) 就可以 搬去 一个 没人 认识 的 地方 我知道 是 很大 一笔 钱 不过 只需要 十万 we could be someplace where no one knows him . and move to a place no one knows . you can move to a place no one knows .
例4	Source Context Ref Baseline Our	(我们) 还是 会 天天 见面 这样 你 就会 记得 我 又 不是 快死了 我 只是 要 搬出去 we're gonna see each other all the time . every day i still see each other all the time .

表 9: 实例分析

6 总结

针对对话领域的零指代机器翻译如何利用好零指代信息来提升翻译效果，本文提出了一种基于联合学习的模型结构。首先，引入现今最先进的神经机器翻译建模模型Transformer来提升翻译的质量，在此基础上联合分类任务对源端句子中的省略成分进行分类，再通过融入篇章信息来进一步提升翻译和分类的性能。在大规模中英对话数据集上的实验证明了本文提出模型的有效性。

但是，本文模型还存在着一些不足。本文融合篇章信息的方式存在优化的空间，在处理比较复杂的省略代词信息时，由于模型难以从驳杂的篇章信息抽取有效的零指代信息，导致难以正确翻译出省略代词。在未来的研究中，会对如何更加有效的利用好篇章信息进行进一步研究。除此之外，如何在无平行数据的情况下提高省略代词的识别准确性也是我们研究的重点。

参考文献

- Rachel Bawden, Rico Sennrich, Alexandra Birch, and Barry Haddow. 2018. Evaluating discourse phenomena in neural machine translation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1304–1313.

- Chen Chen and Vincent Ng. 2013. Chinese zero pronoun resolution: Some recent advances. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1360–1365.
- Linqing Chen, Junhui Li, Zhengxian Gong, Boxing Chen, Weihua Luo, Min Zhang, and Guodong Zhou. 2021a. Breaking the corpus bottleneck for context-aware neural machine translation with a novel joint pre-training approach. In *Proceedings of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*.
- Linqing Chen, Junhui Li, Zhengxian Gong, Xiangyu Duan, Boxing Chen, Weihua Luo, Min Zhang, and Guodong Zhou. 2021b. Improving context-aware neural machine translation with source-side monolingual documents. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence*.
- Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations*.
- Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*.
- Philipp Koehn. 2004. Statistical significance tests for machine translation evaluation. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 388–395.
- Sameen Maruf, André F. T. Martins, and Gholamreza Haffari. 2019. Selective attention for context-aware neural machine translation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3092–3102.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318.
- Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*.
- Xin Tan, Shaohui Kuang, and Deyi Xiong. 2019. Detecting and translating dropped pronouns in neural machine translation. In *Natural Language Processing and Chinese Computing - 8th CCF International Conference*, pages 343–354.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017*, pages 5998–6008.
- Elena Voita, 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. In *Proceedings of the 57th Conference of the Association for Computational Linguistics*, pages 1198–1212.
- Longyue Wang, Zhaopeng Tu, Xiaojun Zhang, Hang Li, Andy Way, and Qun Liu. 2016. A novel approach to dropped pronoun translation. In *The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 983–993.
- Longyue Wang, Zhaopeng Tu, Andy Way, and Qun Liu. 2017. Exploiting cross-sentence context for neural machine translation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2826–2831.
- Longyue Wang, Zhaopeng Tu, Shuming Shi, Tong Zhang, Yvette Graham, and Qun Liu. 2018. Translating pro-drop languages with reconstruction models. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, pages 4937–4945.
- Longyue Wang, Zhaopeng Tu, Xing Wang, and Shuming Shi. 2019. One model to learn both: Zero pronoun prediction and translation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 921–930.

- Lesly Miculicich Werlen and Andrei Popescu-Belis. 2017. Validation of an automatic metric for the accuracy of pronoun translation (APT). In *Proceedings of the Third Workshop on Discourse in Machine Translation*, pages 17–25.
- Lesly Miculicich Werlen, Dhananjay Ram, Nikolaos Pappas, and James Henderson. 2018. Document-level neural machine translation with hierarchical attention networks. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2947–2954.
- Hao Xiong, Zhongjun He, Hua Wu, and Haifeng Wang. 2019. Modeling coherence for discourse neural machine translation. In *The Thirty-Third AAAI Conference on Artificial Intelligence*, pages 7338–7345.
- Yaqin Yang, Yalin Liu, and Nianwen Xue. 2015. Recovering dropped pronouns from chinese text messages. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing*, pages 309–313.
- Zhengxin Yang, Jinchao Zhang, Fandong Meng, Shuhao Gu, Yang Feng, and Jie Zhou. 2019. Enhancing context modeling with a query-guided capsule network for document-level translation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 1527–1537.
- Jiacheng Zhang, Huanbo Luan, Maosong Sun, Feifei Zhai, Jingfang Xu, Min Zhang, and Yang Liu. 2018. Improving the transformer translation model with document-level context. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 533–542.
- Shanheng Zhao and Hwee Tou Ng. 2007. Identification and resolution of chinese zero pronouns: A machine learning approach. In *EMNLP-CoNLL 2007, Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 541–550.