文章编号：1003-0077(2022)06-0036-08

用预定义双语对增强神经机器翻译

王 涛，熊德意

（苏州大学 计算机科学与技术学院，江苏 苏州 215006）

摘 要：将预先定义的双语对融入神经机器翻译(NMT)中一直是一项有较大应用场景，但具有挑战性的任务。受限于 NMT 的非离散特性以及逐词解码策略，想要在 NMT 中显式地融入外部双语对往往需要在解码期间修改集束搜索算法，或者对模型进行复杂修改。该文提出并探索了一种简单的将预先指定双语对融入 NMT 的方法，包括：(1)对训练数据进行适当的预处理，以添加有关预定义的双语信息；(2)使用部分共享的词向量以及额外向量增强信号，帮助模型区分预先指定的双语对和其他翻译文本。在多个语种上的实验和分析表明，该方法可以极大提高预定义短语被成功翻译的概率，达到接近 99%(中英的基准是 73.8%)的效果。

关键词：神经机器翻译；预定义双语对

中图分类号：TP391 **文献标识码**：A

Enhancing Neural Machine Translations with Pre-Defined Bilingual Pairs

WANG Tao, XIONG Deyi

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

Abstract: Integrating pre-defined bilingual pairs into Neural Machine Translation (NMT) has always been a challenging task with substantial application scenarios. Limited by the word-by-word decoding strategy, the explicit integration of external bilingual pairs into NMT often requires modifying the beam search decoding algorithm or even the model itself. This paper proposes a simple method of incorporating pre-defined bilingual pairs into NMT: (1) preprocessing the training data to add information about pre-defined bilingual pairs; (2) using partially shared embeddings help the model distinguish between pre-defined bilingual pairs and other texts. Experiments and analysis in multiple language pairs show that the method can improve the probability of successful translation of pre-defined bilingual pairs, reaching nearly 99% (the Chinese-English benchmark is 73.8%).

Keywords: neural machine translation; pre-defined bilingual pairs

0 引言

随着深度学习的发展，基于深度学习的神经机器翻译(NMT)成为机器翻译的主流方法[1-3]。与传统的基于统计的统计机器翻译不同，神经机器翻译没有特征工程、隐藏结构设计等方面的困扰，而是简单地通过训练一个大型的神经网络对输入的句子产生合适的翻译。尽管神经机器翻译在翻译质量上有着当前最好的结果，但其端到端的特性使得想要在翻译的过程中进行显式的干预是一件很困难的事情。

在许多使用场景中，我们需要神经机器翻译系统使用来自外部数据库的预先定义的翻译。例如，在跨语言电子商务场景中，许多产品的品牌名称是明确的，并且可以直接翻译成目标语言。这些品牌名称的错误翻译将导致纠纷。如表 1 所示，“舒肤佳”在第一句中被直接音译为“Shufujia”，而在第二个例子中被意译为“good for skin”，而正确的品牌名称翻译应该是“Safeguard”。这个例子说明面对这种情况，当前的神经机器翻译不仅无法保证结果的准确性，而且缺乏一致性。

通常，给定源句中一句话 $s=w_1,w_2,\cdots,w_n$，其中出现了存储在双语词典中的双语对 (p,q) 中的源端 p，其中 $p=w_k,\cdots,w_l$，p 应该被翻译系统直接翻译为 q。这对当前的神经机器翻译系统是一

收稿日期：2020-10-23 **定稿日期**：2020-11-26

个不小的挑战。一方面，神经机器翻译是在连续空间向量，而非离散空间中运行；另一方面，神经机器翻译以逐词生成的方式生成目标翻译，而双语词典中指定的翻译通常包含多个词。

表 1　品牌名称错误翻译样例

中文	谷歌翻译
舒肤佳是一个好品牌。	**Shufujia** is a good brand.
舒肤佳的肥皂我经常用。	I often use soaps that are **good for skin**.

为了解决上述问题，我们分别从数据和模型角度提出了几种方法，其中模型上的方法是为了配合数据方法的使用。数据上的方法包括标签标注、混合短语替换；模型上的方法包括部分词向量共享和额外向量增强。具体来说，在数据处理阶段，我们用特殊标签将训练数据中对来自外部词典的文本段的开始和结束的位置打上标记，让模型学习到关于特殊标签的翻译模式。同时，我们将和源端 p 等价的 q 添加到源端，让模型可以同时看到两种语言的信息。因为此时源端和目标端同时包含了 q，模型可以同时学习 q 到 q 的拷贝，以及 p 到 q 翻译，并且学习了跨语言的信息。为了增强标签以及混合短语替换的作用，我们共享了编码器和解码器词向量的标签和目标端部分。此外，我们使用了额外的向量来进一步区分预先定义的词和其他正常需要翻译的词，将在第 2 节中详细介绍。

我们在 3 个语言对上进行了实验，包括中文到英语、英语到德语，以及阿拉伯语到中文。实验结果表明，本文方法在外部词典翻译的准确率上获得了极大的提高。其中，我们在中英语言对上进行了细致的分析实验，成功翻译词典中预定义短语的概率从基准模型的 73.8%增加到 98.4%。此方法还在 45.70 的基准之上实现了 1.58 个 BLEU 的改进。在英语到德语以及阿拉伯语到中文的翻译中，使用本文方法，翻译的成功率也分别从 91.2%、95.0%（基准）提高到 99.3%、99.5%。进一步的实验分析说明了本文方法的泛化性和鲁棒性。

1　相关工作

旨在将外部定义的翻译融入神经机器翻译的方法一般通过修改模型或解码算法来实现。此外，也有一些通过数据进行学习的方法。

Stahlberg 等人[4]使用基于层次化统计机器翻译系统产生的短语作为解码器的硬解码约束，从而使神经机器翻译能够生成更多的符合语法的短语。Tang 等人[5]提出了短语网络，使得解码器可以根据外部短语表生成翻译。Wang 等人[6]尝试将存储目标短语的短语存储记忆集成到编码器-解码器框架中。Zhang 等人[7]尝试将先验知识表示为对数线性模型中的特征，并集成到神经机器翻译之中。这些工作侧重于修改神经机器翻译模型，从而支持翻译外部指定的短语。

将预定义双语词典融入神经机器翻译的另一种方法是修改解码时的集束搜索算法。Hokamp 等人[8]提出了一种基于网格的集束搜索算法，该算法允许在模型的输出中出现特定的子序列，其中子序列可以是单字或多字。Chatterjee 等人[9]进一步提出了一种“引导”机制，用于增强解码器处理带有推荐翻译文本（以 XML 注释形式存在）的能力。上面刚刚提到的几种方法，尽管它们不会改变神经机器翻译模型的结构，但必须在正常解码以及使用外部翻译中进行决策和切换，从而严重降低了解码速度。

还有几种从数据上进行增强的方法。Crego 等人[10]提出用置位符替换双语词典中定义好的词对，从而让模型学习对置位符的翻译，这样模型在进行翻译时就可以通过将源端匹配到的词组替换为置位符，翻译完成后再替换回去。使用置换符的方式简单有效，但是由于将词语替换成了无意义的置位符，源端句子丢失了一定语义，往往会造成 BLEU 下降。Song 等人[11]通过将源短语替换为目标翻译，并使用指针网络来增强对替换短语的拷贝。此方法类似于我们的混合短语替换方法，但是由于指针网络方法较为间接，短语被正确翻译的正确率相对较低。

与以前的工作相比，本文方法成功率高，且不需要引入复杂的解码算法，很容易复现。

2　用预定义双语对增强神经机器翻译

为了用外部词典中预定义的双语对 (p,q) 增强神经机器翻译，我们尝试通过标签标注以及混合短语替换来实现目标。为了进一步增强数据上的方法，我们使用了额外向量并且共享了部分词向量。几种不同的数据处理方法如表 2 所示，对模型的修改如图 1 所示。下面将详细介绍这几种方法。

表 2 几种不同的数据处理方法

原始句对	我喜欢在苏州旅游。
	I like traveling in Suzhou.
标签标注(T)	我喜欢在＜start＞苏州＜end＞旅游。
	I like traveling in ＜start＞ Suzhou ＜end＞ .
短语替换(R)	我喜欢在 Suzhou 旅游。
	I like traveling in Suzhou .
标签标注和混合短语替换(T&M)	我喜欢在＜start＞苏州＜middle＞ Suzhou ＜end＞旅游。
	I like traveling in ＜start＞ Suzhou ＜end＞ .

2.1 方法

2.1.1 标签标注

标签标注方法(缩写为“T”)十分直接。在训练数据集中,源端短语 p 及其对应的目标端短语 q 均被两个标签包围,即＜start＞和＜end＞。一个具体的例子可见表 2 中的第 2 行。随着神经机器翻译模型的训练,这两个标签将自动学习到自己的词向量,就像源端句子和目标端句子中的其他单词一样。由于 p 和 q 出现在相同的模式下,因此可以在它们之间建立连接。当我们使用共享词向量时,这种联系被进一步增强。

2.1.2 混合短语替换

短语替换(R)的方法源自于一个符合常识的直觉:对于深度神经模型来说,学习拷贝要比翻译容易得多。因此,我们提出用目标端的 q 来扩展源端的 p。如表 2 中的第 4 行所示,我们同时使用了标签标注和混合短语替换。在这种情况下,将存在第三个标记,即＜middle＞标记,在混合短语中分开 p 与 q。通过在训练数据中引入标签标注以及混合短语替换的数据,神经机器翻译模型有望学习一种模式,即将这些包含在标签中的片段翻译为其中的子片段。

混合短语替换和之前的工作[11]提出的短语替换(R)有相似之处,即用 q 直接替换掉源端的 p。表 2 中的第 3 行给出了一个样例。相较于直接替换,我们的混合短语替换方法使用了混合的源端短语和目标端短语。我们倾向于在源句中添加更多信息,而不是替换它们,因为替换可能会导致丢弃一些重要信息,包括和替换短语之间的双语信息。此外,混合短语替换对错误的替换也有一定的抗干扰能力。我们将在 3.5 节进行分析。

2.1.3 部分词向量共享

由于标签和混合短语的存在,源端和目标端都存在标签词以及目标端的词。为了增强源端和目标端标签的联系,我们共享标签和目标端词向量。如图 1 所示,编码器的词向量包含三块内容,分别是标签向量、源端词向量和目标端词向量。而解码器的词向量和最后的输出线性映射部分使用和编码器相同的标签和目标端部分向量。我们不共享全部的词向量是为了减少目标端的计算量以及出现输出错误

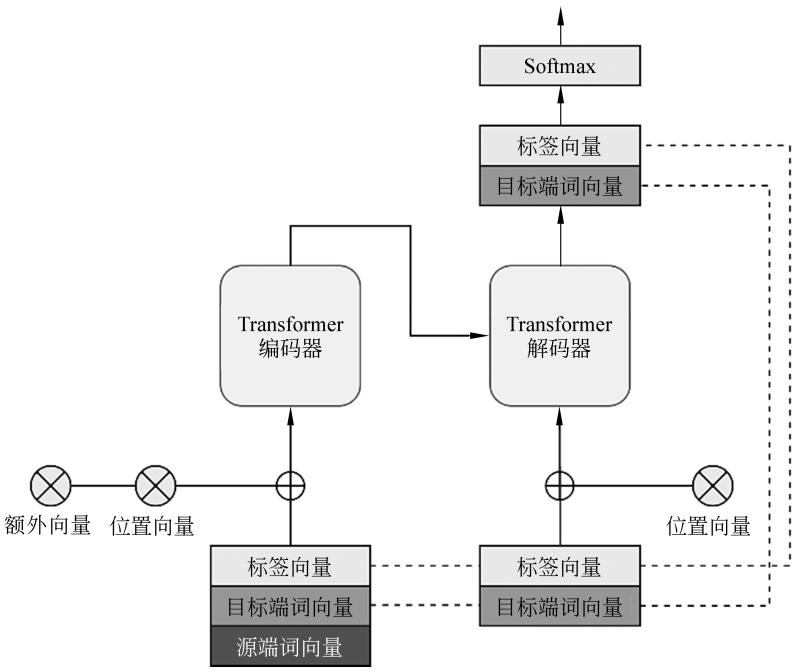

图 1 共享部分词向量并使用额外向量增强的神经机器翻译模型

语言的情况。

2.1.4　额外向量增强

为了进一步增强拷贝信号并区分 p 和 q，我们使用额外向量。对于给定的输入词，其对应的表示是通过将其词向量、位置向量[3]和我们称之为额外向量的三个向量相加得到的。同样以“我喜欢在＜start＞苏州＜middle＞ Suzhou ＜end＞旅游。”为例，其对应的标签序列就是“n n n n s n t n n”。其中“s”和“t”对应于外部双语词典中的源端和目标端，而“n”对应于其他的词。这个想法来自 BERT[12]的句子嵌入，用“A”和“B”区分单个序列中拼接在一起的句子。值得注意的是，与使用标签标注相比，使用额外向量进行增强是一种较为软的方法，因为它将信息直接集成到输入序列中，而无须更改训练文本。

2.2　从平行语料中自动挖掘双语对

预定义的双语对可以由专家总结构建，也可以从双语平行语料库中自动提取。但在训练过程中，专家总结的双语对并不能充分覆盖语料库，所以我们在本节简要介绍从双语语料库中自动挖掘双语对的方法。

整体的流程如图 2 所示。我们专注于命名实体(NE)，并使用 LTP[13]和 spaCy 工具对中文和其他语言进行实体识别。我们使用 Moses[14]生成短语表，并从短语表中查找抽取的实体词，并将大于一定概率的短语对添加到候选列表。如果一个实体词对应短语表中的多个翻译，则过滤最大概率小于 p 的。实验中设置 p 为 0.8。之后，再根据短语对的长度、重合度进行二次过滤，得到最后的预定义双语对。相较于使用词对齐工具，短语表提供了评估质量的概率度量，能抽取更高质量的预定义双语对。

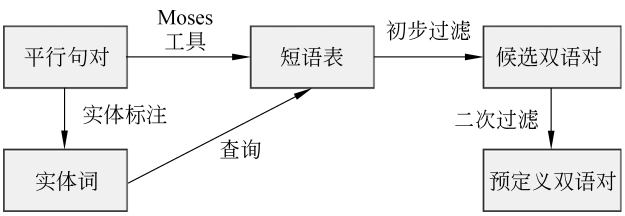

图 2　从平行语料库自动挖掘双语对的流程

3　实验

3.1　实验配置

我们采用 1.25M 大小规模的 LDC 部分语料作为中英数据集。我们选择 NIST06 作为开发集，并选择 NIST03、NIST04、NIST05 作为测试集且还在 4.5M 规模的 WMT2017 英语到德语语料库进行了实验。我们选择 newstest2014 作为开发集，选择 newstest2016 作为测试集。同时，联合国语料被用于阿拉伯语到中文的翻译。

我们使用不区分大小写的 4-元 BLEU 分数作为评价尺度，并且使用“multi-bleu.perl”脚本去计算 BLEU 分数。我们使用字节对编码(BPE)[15]处理所有这些数据，并将合并操作限制为 3 万。

我们使用了目前最主流的基于注意力机制的 Transformer 模型[3]。给定输入序列 $x_1, x_2, \cdots, x_n$，Transformer 会将其编码为一系列连续表示，然后依次生成输出序列 $y_1, y_2, \cdots, y_m$。我们设置编码器和解码器的层数都为 6，隐藏层维数设置为 512，前馈层维数设置为 2 048。我们使用了 8 头注意力机制。在训练过程中，本文使用随机梯度下降算法 Adam 来训练 NMT 模型。Adam 的 β_1 和 β_2 分别被设置为 0.9 和 0.999，学习率被设置为 0.001。训练期间，一次迭代处理 32 000 个词。解码期间，我们使用集束搜索算法并将束搜索大小设置为 6。

3.2　数据处理和质量评估

给定训练数据和预定义的双语对，对于训练数据的每一句，我们遍历其中的 n-元短语，如果匹配到预定义的双语对，则根据第 2 节中的方法进行处理。

如表 3 所示，我们从中英语料库中提取了169 142 个预定义双语对，中英训练集中有 39.2%的句子至少包含一个双语对。在测试集中，这个比例类似。对英语到德语，阿拉伯语到中文，这两个数据分别是 109 759、29.2%和 182 105、24.9%。

表 3　实验数据统计

语料库	训练集大小/M	双语对大小	替换短语占比/%
中英	1.25	169 142	39.2
英德	4.5	109 759	29.2
阿中	16	182 105	24.9

我们可以看到，尽管中英 LDC 语料库规模较小，但由于其新闻领域特性，所以包含更多的专有名词，训练集以及测试集中也包含更多的可替换短语。

在解码阶段，我们对源端进行相同的处理后解码。除了使用 BLEU 对翻译质量进行评估，我们同样评估被替换短语被成功翻译的概率。

3.3 整体实验结果

表 4 展示了作为基准的 Transformer 模型和我们的方法在三个语料库上的结果，表 5 展示了不同方法的组合在中英数据集上的结果。模型对应的两列数据分别表示 BLEU 值和预定义短语被成功翻译对的概率(句子级别)。表格中的 T、M、E 分别对应于第 2 节中所描述的标签标记、混合短语替换、额外向量增强。我们将 Song[11] 等人之前的工作用“R”表示，与我们的混合短语替换(M)方法进行对比。在 Song[11] 等人的论文中，已经论述其方法基本优于之前的方法，所以本文仅和其方法进行比较。部分词向量共享在使用了“M”或者“R”的所有方法上都进行了使用。“T&M&E”表示结合了所有提出的方法的结果。

表 4 总体实验结果

数据集	基准		+T&M&E	
中英	45.70	73.8%	**47.28**	**98.4%**
英德	33.17	91.2%	**33.23**	**99.3%**
阿中	41.10	95.0%	41.10	**99.5%**

从表 4 的结果我们可以看到，我们的方法在不同的语种上都有一定的性能提升。特别是对于中英 LDC 语料，在基准模型上 BLEU 提高了 1.58，预定义双语对翻译的成功率提高了 24.6%。

相较于中英，英德和阿中提升相对较小。英德和阿中的 BLEU 几乎没有变化，预定义双语对翻译的成功率分别提高了 8.1%和 4.5%。一方面，英德和阿中的语料规模较大，训练出的模型对于特定的短语翻译成功率较高；另一方面，英德语料和阿中语料中包含预定义短语的句子占比较少。如表 3 所示，在中英数据集中，大约 40%的句子包含至少一个预定义短语，而在英德和阿中数据集上，占比仅有 30%和 25%左右。如果仅计算包含预定义短语的句子，英德数据集的 BLEU 提高了 0.36。

3.4 不同方法组合的结果对比

表 5 展示了不同方法的组合在中英数据集上的具体结果。我们同时将不同方法组合的结果和基准模型以及在语料中加入短语表的模型进行对比。其中基准模型是标准的 Transformer 模型，加入短语表的模型则是在原有的语料的基础上，把提取的预定义双语对直接加入到语料中一起训练。

表 5 不同方法组合在中英数据集上的结果

方法	NIST06		NIST03		NIST04		NIST05		Avg	
基准	45.55	74.4%	45.12	69.5%	46.36	74.2%	45.61	77.8%	45.70	73.8%
+短语表	45.46	75.0%	45.59	68.8%	46.48	73.5%	45.47	75.1%	45.85	72.5%
+T	45.81	76.5%	46.49	69.3%	46.98	73.4%	45.48	78.6%	46.32	73.8%
+R[11]	46.01	94.0%	46.56	92.1%	47.00	92.1%	46.15	93.9%	46.57	92.7%
+T&R	46.09	98.0%	46.69	98.4%	46.70	**98.7%**	46.26	98.3%	46.55	**98.4%**
+T&M	**46.74**	**98.7%**	46.37	**99.1%**	**47.23**	98.0%	46.82	98.0%	46.81	**98.4%**
+T&M&E	46.38	98.3%	**47.33**	98.2%	47.18	98.6%	**47.34**	**98.4%**	**47.28**	**98.4%**

我们可以看到，在训练语料中加入双语对并不会对 BLEU 和短语翻译成功率带来明显的影响。这是由于短语本身就是从语料中抽取获得的，因此语料包含了相关信息。

单独使用标签标注方法(T)不能给预定义短语翻译成功率带来明显收益，单独使用短语替换方法(R)将平均成功率提高到了 92.7%，相对来说是一个很大的提升。但是当标签标注方法(T)和短语替换或者混合短语替换方法结合时，可以获得最好的效果，成功率达到了 98.4%。我们认为这是因为当使用 M 或者 R 方法时，我们共享了部分词向量，从而为句子源端和目标端的标签标注提供了更强的连接。结合了所有方法的 T&M&E 同样达到了最好效果。

从 BLEU 值来看，T&M 方法相较于 T&R 有一定的优势，而结合了 E 方法后性能得到了进一步的提升。正如我们在第 2 节中描述的，相较于直接替换(R)，混合短语替换(M)保留了原始的预定义短语信息，增强模型对跨语言信息的学习，并带来了一定的抗噪能力。额外向量增强(E)在增强了对替换短语拷贝信号的同时，提供了对源端语块的区分，从而帮助模型更好区分正常翻译部分、被替换短语

部分，以及可以直接拷贝的部分。

3.5 词典外短语以及错误替换短语的翻译

为了进一步评估本文方法的作用以及运作机制，我们针对预定义双语词典外的短语以及错误替换短语两种情况进行分析。

同样以“我喜欢在＜start＞苏州＜middle＞Suzhou ＜end＞旅游。”为例，使用预定义双语词典外的短语可以将“苏州”和“Suzhou”替换为不在词典中的其他双语对，这测试了模型的泛化能力，同时也和实际场景中实时扩充词典的需求符合；而错误替换短语指“苏州”保留不变，“Suzhou”修改为其他的词，这在一定程度上可以检验保留的源端部分对翻译产生的影响。针对这两种情况，我们分别人工构造了 200 个样例，并进行评估，结果如表 6 所示。

表 6　词表外短语及错误替换短语的翻译结果

情况	结果	百分比/%
词表外短语	成功翻译	98
	错误翻译	2
错误替换	翻译源端	6
	拷贝替换	90
	都不出现	4

从表 6 中可以看到，模型对词表外的短语有很好的泛化性能，能达到 98%成功率，错误的往往是那些替换了无意义稀有词的句子。这说明我们的模型在临时扩充的双语对上依然有良好的适应能力。翻译错误替换句子时，模型在大部分情况下仍然倾向于直接拷贝源端的替换部分，特别是替换的是类似的词，如不同代词替换、不同人名替换。令人惊讶的是有 6%的错误替换样例会被正确的翻译源端部分代替，而丢弃错误的替换。经过人工分析，我们发现翻译模型会尽量让拷贝的结果合理出现在翻译句子中。当替换为一些不可能出现的词（特别是稀有词），模型经常会丢弃掉那部分错误替换，退化为翻译被替换部分。还有 4%的句子会被错误替换严重影响，翻译出随机的词。

4 分析

4.1 针对词向量的分析

由于使用了混合短语替换方法，部分共享的词向量在训练中学习到了跨语言的信息。这一点我们可以通过提取翻译模型中的词向量并计算其中一些词的邻近词来观察到。

表 7 中展示了在词向量空间中和“india”/“印度”以及“beijing”/“北京”最为邻近的 5 个词。距离通过计算 cosine 距离获得。可以看到，和“beijing”最为靠近的词是其对应的中文“北京”，同时“beijing”也是其对应中文的最邻近的词。混合短语替换方法中的混合短语，成为了词向量空间中的锚点，让翻译模型的词向量学习到了更多跨语言的信息。同时，跨语言的词向量可以进一步帮助提高翻译的质量[16-17]。

表 7　词向量空间中的邻近词

india	印度	beijing	北京
印度	印	北京	beijing
印	印@@	京	京
japan	印度@@	china	北京
russia	india	shanghai	来京
namibia	印中	京@@	京@@

4.2 针对注意力机制的分析

Transformer 模型使用多头注意力机制[3]，让不同的头注意到不同语义空间的信息。图 3 是句子“＜start＞丹麦＜middle＞danish＜end＞首相＜start＞拉斯@@ 穆@@ 森＜middle＞ras@@ mus@@ sen＜end＞星期二在首@@ 相@@ 府举行新闻发布会。”解码过程中的交叉注意力矩阵图，即解码器对编码器的注意力图，颜色越深注意力权重越大。我们使用最后一层的输出并对多头注意力取平均。

当关注被标签标记的预定义短语时，我们可以看到无论是“danish”还是“ras@@ mus@@ sen”都同时对源端的中文和英语部分有一定的注意力权重，这展示了混合短语替换的作用，即同时为目标端提供翻译和拷贝的信息。同时，我们注意到像“danish”这种在语料中经常出现的词，注意力矩阵在中文部分的权重会更大，因为模型较为确认这种翻译。而对于“ras@@ mus@@ sen”这种人名，由于数据中出现次数较少，翻译模型不能充分学习到，所以倾向于直接拷贝，从而对源端的英文部分有更高的权重。

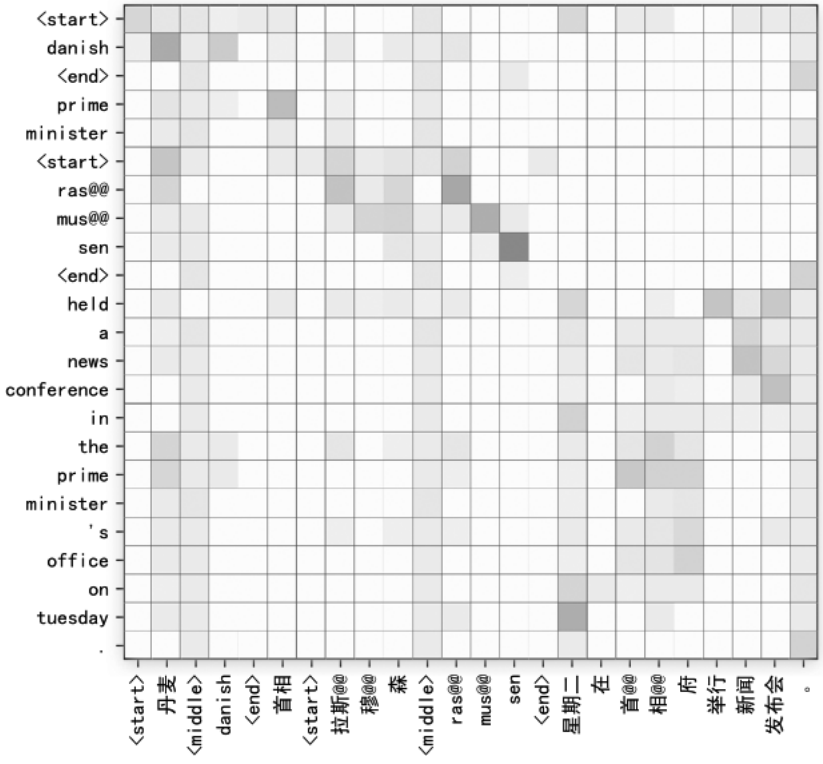

图 3 句子样例的交叉注意力示意图

4.3 针对具体样例的分析

表 8 中的样例体现了本文方法的优势。由于数据存在偏置，我们的字典中“南韩”对应“korea”，所以原文中的“南韩”被默认替换为了“korea”。在直接替换的＋R 方法下，翻译模型倾向于直接拷贝，翻译为“the korea and north korea”。但是考虑到“north korea”也在上下文中，“南韩”翻译为“korea”不够准确。我们的 T&M&E 方法可以依靠保留的“南韩”，同时参考了原始信息以及替换信息，输出更为准确的翻译“south korea and north korea”。从这个例子看，本文方法对类似情况有更好的鲁棒性。

表 8 翻译样例比较

源端句子	冷战时期相互为敌的**南韩**与北韩，双方代表队在进场仪式中携手共同入场。
参考译文	the delegations from **south and north korea**, the two cold war foes, marched into the stadium together for the entrance formalities.
＋R	冷战时期相互为敌的 **korea** 与北韩，双方代表队在进场仪式中携手共同入场。
翻译	**the korea** and north korea, which are hostile to eachother during the cold war period, joined hands at the arrival ceremony.
＋T&M&E	冷战时期相互为敌的＜start＞南韩＜middle＞ **korea** ＜end＞与北韩，双方代表队在进场仪式中携手共同入场。
翻译	**south korea** and north korea, which are hostile to each other during the cold war era, jointly attended the opening ceremony.

5 总结

本文提出使用简单的数据预处理，包含标签标注、混合短语替换，以及对应的模型修改，包括共享部分词向量和额外向量增强，从而将外部的预定义双语对融入神经机器翻译。三个语对上的实验证明了本文方法的有效性。各种方法的组合对比实验说明了不同方法的作用。通过进一步分析，我们从词向量和注意力角度分析了方法为何有效。在未来的

工作中，我们希望考虑一词多义的情况，即如何将控制不同含义的外部词融入神经机器翻译中。

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 3rd International Conference on Learning Representations, 2015.

[2] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems, 2014: 3104-3112.

[3] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017: 5998-6008.

[4] Stahlberg F, Hasler E, Waite A, et al. Syntactically guided neural machine translation[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016: 299-305.

[5] Tang Y, Meng F, Lu Z, et al. Neural machine translation with external phrase memory[J]. arXiv preprint arXiv:1606.01792, 2016.

[6] Wang X, Tu Z, Xiong D, et al. Translating phrases in neural machine translation[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2017: 1421-1431.

[7] Zhang J, Liu Y, Luan H, et al. Prior knowledge integration for neural machine translation using posterior regularization[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017: 1514-1523.

[8] Hokamp C, Liu Q. Lexically constrained decoding for sequence generation using grid beam search[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017: 1535-1546.

[9] Chatterjee R, Negri M, Turchi M, et al. Guiding neural machine translation decoding with external knowledge[C]//Proceedings of the 2nd Conference on Machine Translation, 2017: 157-168.

[10] Crego J, Kim J, Klein G, et al. Systran's pure neural machine translation systems[J]. arXiv preprint arXiv:1610.05540, 2016.

[11] Song K, Zhang Y, Yu H, et al. Code-switching for enhancing NMT with prespecified translation[C]//Proceedings of the NAACL-HLT (1), 2019:449-459.

[12] Devlin J, Chang M W, Lee K, et al. Bert: pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.

[13] Che W, Li Z, Liu T. LTP: A Chinese language technology platform[C]//Proceedings of the 27th International Conference on Computational Linguistics: Demonstrations, 2010: 13-16.

[14] Koehn P, Hoang H, Birch A, et al. Moses: open source toolkit for statistical machine translation[C]//Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions. Association for Computational Linguistics, 2007: 177-180.

[15] 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, 2016: 1715-1725.

[16] Lample G, Ott M, Conneau A, et al. Phrase-based & neural unsupervised machine translation[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2018: 5039-5049.

[17] Artetxe M, Labaka G, Agirre E, et al. Unsupervised neural machine translation[C]//Proceedings of the International Conference on Learning Representations, 2018: 1-12.

王涛(1996—)，硕士研究生，主要研究领域为自然语言处理、机器翻译。
E-mail: rgwt1234@gmail.com

熊德意(1979—)，通信作者，博士，教授，主要研究领域为自然语言处理、人工智能。
E-mail: deyi.xiong@gmail.com