

# 基于句子级上下文内容的神经机器翻译方法

杨 娇

(商洛学院 人文学院, 陕西 商洛 726000)

**摘要:** 传统的神经机器翻译方法在忽略句子层面语境的情况下, 利用词层面语境来预测目标语的翻译, 这有利于统计机器翻译中的翻译预测; 但由于词的主题往往是根据句意和上下文动态变化的, 存在一词多义等复杂情况; 这会导致翻译的不准确; 因此, 利用卷积神经网络将句子级上下文表示为潜在的主题表示, 并设计了一个主题关注度模型, 将源句子级主题上下文信息集成到基于 Attention 和基于 Transformer 的神经机器翻译方法中; 实验结果表明, 提出的方法的 BLEU 评分最高大约等于 40。

**关键词:** 机器学习; 卷积神经网络; 机器翻译; 潜在主题表示; 上下文内容

## Neural Machine Translation Method Based on Sentence-level Contextual Content

Yang Jiao

(School of Humanities, Shangluo University, Shangluo 726000, China)

**Abstract:** Traditional neural machine translation methods use word-level context to predict target language translation while ignoring sentence-level context, which is conducive to statistical machine translation prediction. However, because the theme of words often changes dynamically according to sentence meaning and context, there are complicated situations such as polysemy. This will lead to inaccurate translation. Therefore, the convolutional neural network is used to represent the sentence-level context as a potential topic representation, and a topic attention model is designed to integrate the source sentence-level topic context information into neural machine translation methods based on Attention and Transformer. Experimental results show that the maximum BLEU score of the proposed method is approximately 40.

**Keywords:** machine learning; convolutional neural network; machine translation; latent topic representation; contextual content

### 0 引言

基于短语的统计机器翻译 (phrase-based statistical machine translation, PBSMT)<sup>[1-2]</sup> 和神经机器翻译 (neural machine translation, NMT)<sup>[3-4]</sup> 在翻译语境中学习主题信息引起了广泛的关注。尽管取得了这样的成功, 现有的方法大多数还是围绕着提前学习每个单词的固定主题分布, 以在看不见的源语句中模拟单词主题。在实践中, 词的主题往往根据句子的上下文动态变化, 而不是静态的预先训练的主题分布。换言之, 单词主题在很大程度上依赖于它们的句子级上下文, 甚至一个单词在一个句子中可以有多个主题。然而, 现有的 NMT 结构只关注与下一个目标词相关的源词级上下文信息, 而忽略了基于句子级上下文的主题信息。

因此本研究以句子级上下文为研究对象, 对源话题信息进行建模, 并设计了一个话题关注度模型, 将学习到的潜在话题表示融入到已有的 NMT 结构中, 以提高目标词的预测能力。为此, 首先通过卷积神经网络 (convolutional neural networks, CNNs) 的一个变体, 将源语句上的源主

题信息表示为潜在主题表示 (latent topic representations, LTRs)<sup>[5]</sup>。然后根据单词上下文和主题上下文学习, 用于计算用于预测目标单词的额外主题上下文向量。本文的创新点在于本研究的方法是动态而不是静态地学习每一个源语句的 LTR, 并通过主题注意而不是简单的向量连接将 LTR 集成到解码器中, 同时对源词和译文进行联合而不是单独的修改。

### 1 CNN 模型变体

在本研究中, 主要采用了一种 CNN 变体, 它基于句子级上下文捕获源主题信息, 本节将具体介绍该 CNN 的网络结构, 如图 1 所示。

考虑到 CNN 将话题信息隐式地映射到话题向量中, 将这些向量称为 LTR。与传统 CNN 的句子分类任务不同, 在这里使用卷积层来学习 2 MD 维特征向量, 因此, 通过最大值得合并行向量来提取 MD 维向量。提出的 CNN 网络结构如下。

#### 1.1 网络结构中的输入层

$v_j \in \mathbb{R}^D$  是与句子中第  $j$ -th 个词相对应的  $D$  维词向

收稿日期: 2020-06-05; 修回日期: 2020-07-06。

基金项目: 陕西省自然科学基金(2020JM-630); 陕西省社科基金(2019M039)。

作者简介: 杨 娇(1987-), 女, 陕西商洛人, 讲师, 硕士, 主要从事翻译理论和信息化方向的研究。

引用格式: 杨 娇. 基于句子级上下文内容的神经机器翻译方法[J]. 计算机测量与控制, 2021, 29(1): 194-199.

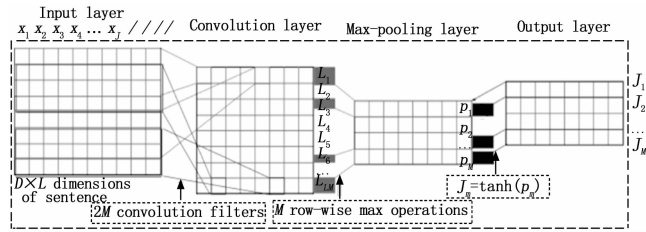


图 1 所提出的 CNN 模型结构

量。因此，长度为  $J$  的句子被表示为  $D \times J$  向量矩阵  $M$ 。这里，将最大句子长度  $L$  设置为 50。较短的句子用尾随的零填充。

### 1.2 网络结构中的卷积层

由  $2M$  个滤波器组成，每个滤波器  $W_m \in \mathbb{R}^D$  ( $1 \leq m \leq 2M$ ) 应用于  $M$  的  $t$  个连续行的窗口，产生特征  $L_d^m \in \mathbb{R}$ ，如式 (1) 所示：

$$L_d^m = Relu(W_m \cdot M_{d,d+t-1} + b_m), 1 \leq d \leq D \quad (1)$$

其中： $b_m \in \mathbb{R}$  是一个偏倚项， $M_{d,d+t-1}$  是所有单词的  $d$ -th 维到  $\{d+t-1\}$ -th 维的连接。每个滤波器应用于输入矩阵  $M$  的每个可能窗口，生成特征向量  $L_m = \{L_1^m, \dots, L_D^m\} \in \mathbb{R}^D$ ，类似于句子分类任务中编码源主题信息的特征向量  $L_m$ 。依次使用  $2M$  滤波器遍历  $M$  生成特征映射  $L$ ，其中  $L = \{L_1, L_2, \dots, L_{2M}\}$ 。

注意，在建议的 CNN 的句子分类任务中，每个过滤器对所有词向量的组成行执行，而不是对局部词向量执行。在此期望学习每个特征值时考虑所有词的部分信息，期望最终的潜在主题表示依赖于句子级上下文而不是局部词级上下文。换句话说，一个特征值是从所有单词的特定向量空间而不是从局部单词向量中学习的。此外，本研究还设计了一个类似于句子分类任务的 CNN 结构。

与图 2 (a) 中的输入矩阵相比，长度  $J$  的句子被表示为  $J \times D$  向量矩阵，而不是原来的  $D \times J$  向量矩阵，如图 2 (b) 所示。因此，每个过滤器都从本地单词嵌入中提取。

### 1.3 使用的最大池层

在  $L$  的连续行对上取行最大值，如式 (2) 所示：

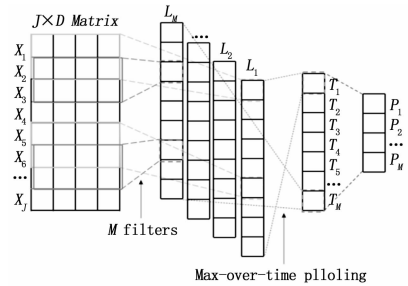
$$P_m = \max(L_{2m-1}, L_{2m}), 1 \leq m \leq M \quad (2)$$

结果输出特征映射为  $P$ ， $P = \{P_1, P_2, \dots, P_M\}$ 。注意，对最大池化操作应用于  $L_{2m-1}$  和  $L_{2m}$  以获得  $D$  维主题特征向量  $P_m$ 。与句子分类任务的最大池相比，在这里使用  $D$  维主题特征向量  $P_m$  来表示输入句子中的主题，而不是主题特征值。众所周知，矢量表示具有更好的编码单词或主题信息的能力。同时，主题的矢量表示也更容易集成到现有的 NMT 架构中。

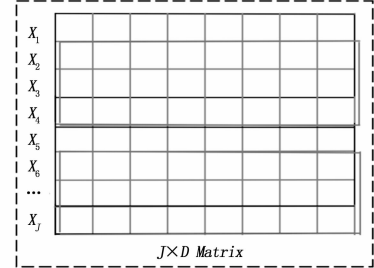
### 1.4 网络结构中的输出层

这将  $\tanh$  函数应用于  $P_m$  以获得 LTR  $T_m$ ，具体如式 (3) 所示：

$$T_m = \tanh(P_m), T = \{T_1, T_2, \dots, T_M\} \quad (3)$$



(a) 句子分类任务中的 CNN 模型



(b)  $J \times D$  输入矩阵

图 2 不同输入矩阵的对比

$T$  是提出的 LTRs，稍后将用于学习 NMT 的主题上下文向量。学习  $T$  的模型参数如式 (4) 所示：

$$\eta = \{W_1, W_2, \dots, W_{2M}; b_1, b_2, \dots, b_{2M}\} \quad (4)$$

## 2 句子级主题语境下的 NMT

### 2.1 基于 Attention 机器翻译

在这一节中，引入了一个基于 Attention 的 NMT 模型 (ANMT) 来学习用于翻译预测的额外句子级主题上下文向量。根据文献 [6] 可知，首先使用递归单元  $REC_1$ ，通过使用先前的解码器隐藏状态  $S_{i-1}$  和先前发射的目标字  $E[y_{i-1}]$ ，通过式 (5) 计算隐藏状态建议  $S'_i$ ：

$$S'_i = REC_1(S_{i-1}, E[y_{i-1}]) \quad (5)$$

其中： $S_{i-1}$  是根据原始单词上下文和建议的主题上下文计算的。

第二，根据文献 [7] 可知，机制利用隐藏状态建议  $S'_i$  来计算每个  $T_n$  在时间步骤  $i$  的额外主题对齐权重  $\epsilon_m^T$ ，从而用 LTRs 加权和求和得到主题上下文向量建议  $c_j^T$ ，如式 (6) 所示：

$$\begin{cases} \epsilon_m^T = v_m^T \tanh(u_m^T S'_i + W_a^T T_n) \\ A_m^T = \frac{\exp(\epsilon_m^T)}{\sum_{k=1}^N \exp(\epsilon_k^T)} \\ c_j^T = \sum_{n=1}^N A_m^T T_n \end{cases} \quad (6)$$

直观地说，NMT 的目的是产生一个与源句意义相同的目标词序列，而不是产生相同的目标话题序列。换言之，在翻译预测过程中，主题信息可以起到辅助作用。因此，计算  $\lambda_i \in [0, 1]$ ，这是一个门控标量<sup>[8]</sup>，用于在时间步骤  $i$  为下一个目标词加权源主题上下文的预期重要性：

$$\lambda_i = \text{sigmoid}(\omega_\lambda s'_i + u_\lambda c_i^T) \quad (7)$$

在式 (7) 中， $\omega_\lambda$  和  $u_\lambda$  是模型参数。它依次用于在时

间步骤  $i$  计算主题上下文向量  $c_i^T$ ，其等于  $\lambda_i c_i^T$ 。

然后，使用主题上下文向量  $c_i^T$  作为 REC<sub>2</sub> 的修改版本的附加输入，该修改版本现在基于主题上下文向量  $c_i^T$ ，根据文献 [7] 可知，可以用原始上下文向量  $c_i$  和隐藏状态建议  $s'_i$  通过式 (8) 计算当前解码器隐藏状态  $s_i$ ：

$$\begin{cases} z_i = \sigma(W_z c_i + W_z^T c_i^T + u_z s'_i) \\ R_i = \sigma(W_r c_i + W_r^T c_i^T + u_r s'_i) \\ s_i = \tanh(W_c + W^T c_i^T + R_i \odot (u s'_i)) \\ s_i = (1 - z_i) \odot s_i + z_i \odot s'_i \end{cases} \quad (8)$$

最后，新的解码器隐藏状态  $S_i$  用于计算下一个目标词  $y_i$  的概率，同时嵌入先前省略的目标词  $E[y_{i-1}]$  和两个上下文向量  $c_i$  和  $c_i^T$ ，如式 (9) 所示：

$$\hat{P}(y_i | y < i, x) \propto \exp(L_o \tanh(L_s E[y_{i-1}] + L_w c_i + L_T c_i^T + L_s s_i)) \quad (9)$$

其中： $L_o, L_s, L_w, L_T$  和  $L_s$  是投影矩阵。注意，除了额外的主题上下文向量  $c_i^T$  之外，需要使用新的解码器隐藏状态  $s_i$  替换文献 [7] 中的原始隐藏状态  $s_i$ 。

### 2.2 基于 Transformer 的 NMT 中的主题

在本节中，将介绍如何将提出的 CNN 集成到现有的 NMT 中，以共同学习 LTR 和翻译。本研究使用的 NMT 为基于 Transformer 的 NMT<sup>[9]</sup>。

首先，提出的 CNN 作为编码器的附加模块，从输入源语句中学习 LTR 序列。其次，使用额外的多头注意力 (multi-head attention) 模块，学习基于前一解码器层的目标查询的新主题上下文表示。与原有的词级上下文表示方法相比，新的主题上下文表示方法侧重于获取句子级的主题信息，用于翻译预测。最后，主题上下文向量和原始单词上下文表示一起用于预测目标转换，如图 3 所示。

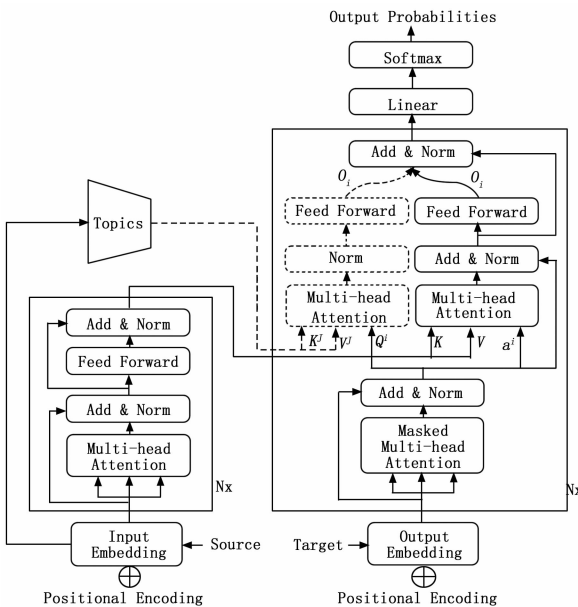


图 3 基于 Transformer 的 NMT 网络结构

形式上，编码器的 CNN 模块首先从输入源  $T$  语句学习 LTR 序列。然后  $T$  被映射到一组键值对  $\{K, V\} = \{(K_1, V_1), (K_2, V_2), \dots, (K_M, V_M)\}$ 。在解码器中，多头自关注将先前解码器层  $Q^j$ ， $K$  和  $V$  的目标查询转换为  $H$  次：

$$\{Q_i^h, K^h, V^h\} = \{Q_i W_h^Q, K W_h^K, V W_h^V\} \quad (10)$$

在这里  $\{Q_i^h, K^h, V^h\}$  是第  $h$ -th 个主题头的查询，键和值向量。 $\{W_h^Q, W_h^K, W_h^V\} \in \mathbb{R}^{d_{model} \times d_i}$  是模型参数。根据文献 [7] 可知，每个主题子空间的主题上下文可用式 (11) 表示：

$$O_i^h = \text{softmax}\left(\frac{Q_i^h K^{h^T}}{\sqrt{d_k}}\right) V^h \quad (11)$$

最后，将  $H$  子空间中的主题上下文向量连接为当前时间步主题向量  $Q_i$ 。根据文献可知，新主题上下文向量  $Q_i$  和原始单词上下文向量  $Q_i$  都用于通过线性，潜在的多层函数计算下一个目标单词的翻译概率，如式 (12)：

$$P(y_i | y_{<i}, x) \propto \exp(L_o \tanh(L_w O_i + L_T O_i)) \quad (12)$$

在这里  $L_o, L_w$  和  $L_T$  代表投影矩阵。

### 2.3 显式主题表示

与传统主题方法相比，本文提出的 LTRs 通过神经网络对源主题进行隐式编码，这可能很难显示这些学习到的 LTRs 编码的主题信息。为了进一步了解源主题的有效性，设计了一种通过术语频率逆文档频率 (TF-IDF) 的显式主题表示方法<sup>[10]</sup>。具体地说，长度为  $J_g$  的输入句子被视为文档  $X_g$ ，并为  $X_g$  中的每个单词  $x_j$  计算 TF-IDF  $TI_j$ ，如式 (13) 所示：

$$TI_j = \frac{n_{j,g}}{J_g} \times \log \frac{|G|}{|g: x_i \in X_g|} \quad (13)$$

在这里中， $n_{j,g}$  表示输入句子  $d_g$  中的第  $j$ -th 个单词的出现次数； $|G|$  是训练数据中源语言句子的总数； $|g: x_i \in X_g|$  是训练数据中包含单词  $x_i$  的源句子的数目。然后选取固定百分比 (实验中为 40%) 的高 TD-IDF 词，将其转换成词向量，形成一个主题 T 序列。

与以往的 LTR 相比，T 是基于 TF-IDF 方法显式提取的，因此称为显式主题表示 (explicit topic representations, ETRs)。最后，利用 ETRs 序列代替以往的 LTRs 序列，并将其集成到现有的 NMT 体系结构中，以增强翻译预测的注意对齐。

### 2.4 训练模型

提出的对主题关注的 NMT 是一个不需要任何预先训练的主题信息的集成架构。换言之，它同时学习 LTRs 和翻译，而不是单独学习。具体地说，让  $\theta$  表示与翻译相关的参数，而  $\eta$  表示用于学习 LTRs 的参数。一组训练示例  $\{[x^g, y^g]\}_{g=1}^G$  上的损失函数为等式 (14)：

$$T(\theta, \eta) = \sum_{g=1}^G \sum_{i=1}^{|y|} \log \hat{P}((y_i | y_{<i}, x; \theta, \eta)) \quad (14)$$

因此，训练反对意见的重点是寻找一个组参数 ( $\theta^*$ ,  $\eta^*$ )，以便在一组训练示例  $\{[x^g, y^g]\}_{g=1}^G$ ，具体可以用式

(15) 表示。

$$T(\theta^*, \eta^*) = \operatorname{argmax}_{\theta, \eta} T(\theta, \eta) \quad (15)$$

### 3 实验与评估

#### 3.1 数据集与设置

在这里使用两个翻译数据集对所提出的方法进行评估：一个用于 LDC 中英翻译 (ZH-EN)，另一个用于 WMT' 14 英德翻译 (EN-DE)。ZH-EN 训练集包括 118 万个来自 LDC 语料库的双语句子对，而 MT06 和 MT02/MT03/MT04/MT05 数据集分别用作开发集和测试集。EN-DE 训练集包括 409 万个 WMT' 14 个语料库的双语句子对，而 newstest2012 和 newstest2013/newstest2014/newstest2015 分别用作开发集和测试集。

#### 3.2 结果与评估

##### 3.2.1 基于 RNN 的 NMT 的主题数和训练效率

图 4 显示了针对不同源话题的 MT02 上所提出的 NMT 模型的翻译性能和训练速度。随着话题数的增加，训练速度在开始时 (从 0 到 10) 显著下降，然后 (10 之后) 略有下降。对于 NMT 性能，当主题数从 0 增加到 40 时，BLEU 从 37.78 增加到 38.73；40 后，BLEU 开始随着主题数的增加而减少。很明显，当源主题数为 40 时，所提出的 NMT 在速度下降可接受的情况下表现最好。

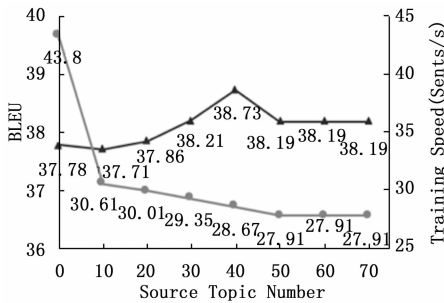


图 4 源话题数与 BLEU 和训练时间之间的关系

##### 3.2.2 基于 Attention 的 NMT 的性能比较

在这一节中，评估了基于 Attention 的 NMT (ANMT) 翻译性能结果，并与 PBSMT<sup>[11]</sup>，TiNMT<sup>[12]</sup>，SFLTR<sup>[13]</sup>，DWCont<sup>[14]</sup>，LTR 进行了比较。

表 1 显示了不同方法在测试集上的翻译性能。ANMT 的平均成绩比 PBSMT 高出 3.87 个 BLEU 点，表明 ANMT 是一个很强的 baseline。

表 1 基于 ZH-EN 翻译结果

ZH-EN	Dev(MT02)	MT03	MT04	MT05	MT06
PBSMT	33.53	31.22	33.19	29.63	29.65
TiNMT	38.11	35.98	38.18	33.89	32.82
ANMT	37.78	35.64	38.03	33.11	32.42
+SFLTR	37.65	35.79	38.14	33.37	32.76
+DWCont	38.01	36.12	38.46	33.94	33.15
+LTR	38.73	36.67	38.94	34.87	33.66

此外，表 2 列出了 EN-DE 任务的翻译结果以及 40 个源话题 (与 ZH-EN 任务相同)。

表 2 EN-DE 翻译结果

EN-DE	Dev(tst2012)	tst2013	tst2014	tst2015
PBSMT	15.21	17.13	15.43	17.67
TiNMT	16.96	20.28	19.12	20.81
ANMT	17.03	20.23	18.60	20.68
+SFLTR	17.16	20.48	18.54	20.92
+DWCont	17.64	21.12	19.71	21.16
+LTR	18.06	21.67	20.29	21.85

从表 2 中可知，+LTR 对 ANMT 和 TiNMT 进行了类似的改进，因此表明本研究的工作是改进其他语言对翻译的可靠方法。

##### 3.2.3 基于不同构型的 NMT 网络的结果比较

在这一节中，在一个强大的 baseline 下，进一步评估了提出的方法。这个 baseline 是由一个标准的训练配置训练的。例如，采用字节对编码 (BPE) 算法<sup>[16]</sup>，字节大小设置为 32 k，所有输入输出层的维度设置为 512，内部 FFN (feedforward neural network) 层的维度设置为 2 048。一方面，对于基于 RNN 的 NMT 模型，另一方面，与 3.2.2 节的设置相同；另一方面，对于基于 Transformer 的 NMT 模型，在 8 000 个预热步骤下，学习率有所变化，模型 (基础) 被训练了大约 20 万个批次。所有模型都在一个 P100 GPU 上进行训练和评估。SacreBLEU<sup>[17]</sup> 被用作 EN-DE 任务的评估指标，并且 -bleu.perl 语言作为 ZH-EN 任务的评价指标。

这些转换结果如表 3 和表 4 所示。新的 baseline，包括 ANMT 和 Transformer，优于表 1 和表 2；特别是新的 Transformer (base) 的 BLEU 得分与文献 [9] 中展示的结果相同。这表明本研究的 baseline 是强大的比较系统。在表 3 和表 4 中，+LTR 和 +ETR 都优于 Transformer (base)，这表明所提出的句子级主题信息有利于 NMT。+LTR 取得了与表 3 和表 4 中的 +ETR 相当的性能，这意味着这些学习到的 LTR 能够捕获源主题信息。+LTR 优于 +SFLTR，说明动态注意主题语境比单一固定主题更能有效地预测目标词。+LTR 的性能优于 +DWCont，并且它们都有相似的模型参数。这表明改进来自句子级的主题信息，而不是更大的模型参数。+ETR 和 +LTR 的 BLEU 评分均优于文献 [9] 在 Transformer (big) 模型中。这表明提出的方法提高了翻译性能。

##### 3.2.4 输出样本研究

如图 5 中示例 1，在此将所提出的方法与 ANMT 对淡色源汉语单词 “xinchun” 的翻译进行了比较。直观地说，本研究提出的 LTR 和 ETR 利用句子层面的语境来编码源话题，如经济、仪式、中国和其他潜在话题。通过注意机制，

表 3 基于 Transformer 的 NMT 翻译结果

名称	架构	测试集		# Speed1	# Speed2	# Param
		newstest	newstest			
		2014	2015			
Existing NMT Systems						
文献[11]	GNMT	26.3	N/A	N/A	N/A	N/A
文献[15]	CONVS2S	26.43	N/A	N/A	N/A	N/A
文献[9]	Transformer (base)	27.3	N/A	N/A	N/A	65.0M
	Transformer (big)	28.4	N/A	N/A	N/A	213.0M

Our NMT systems

本文模型	Transformer (base)	27.25	29.08	9910	181	97.6M
	+DWCnt	27.81	29.57	9021	175	116.9M
	+SFLTR	27.51	29.39	9837	181	99.8M
	+ETR	28.10	29.71	9412	179	109.4M
	+LTR	28.18	29.81	8956	175	116.4M
	Transformer (big)	28.34	30.02	4345	154	272.8M
	+SFLTR	28.31	30.21	4015	151	274.8M
	+ETR	28.94	30.79	3875	149	290.6M
	+LTR	29.06	30.86	3819	149	292.6M

表 4 基于 RNN 的 NMT 翻译结果

名称	架构	Test Set		# Speed1	# Speed2	# Param
		newstest	newstest			
		2014	2015			
本文模型	ANMT	20.4	22.6	17451	328	54.3M
	+DWCnt	21.1	23.18	17232	321	56.9M
	+SFLTR	20.76	22.88	17396	323	54.8M
	+ETR	21.61	23.74	16943	318	55.6M
	+LTR	21.68	23.69	16879	315	56.1M

本研究的方法学习了一个额外的话题语境向量，该向量聚焦于源话题，仪式和中国，以正确地将“xinchun”翻译为“springfestival”，而不是“newyear”。此外，通过图 6 中的显式主题对齐可视化验证了这些观察结果。例如，译文“spring”和“festival”有着相似的主题词，如“nianchuwu”，“xinchun”，“tuanbai”。

图 5 的示例 2 显示了所提出的方法与 ANMT 之间的其他源汉语单词“neidi”的翻译。同样，可能有多种来源的话题，如经济、中国、政治和其他潜在话题。直观地说，当不考虑源语话题（或句子层面的语境）时，源语“neidi”可以译为“大陆（mainland）”或“内地（inland）”。总的来说，翻译“内地”与话题地理密切相关，而翻译“大陆”则与话题政治密切相关。在提议的 LTR 和 ETR 中，主题关注的焦点是中国和政治，选择翻译“大陆”而不是“内地”。

Example 1

Src: 今天是年初五，我很高兴获金银业贸易场邀请，出席你们的新春团拜。  
(pinyin) jintian shi nianchuwu, wo hen gaoxing huo jinyin ye maoyichang yaoqing, chuxi nimen de xinchun tuanbai.  
并主持新一年的开市仪式。  
bing zhuchi xin yi nian jinshi de kaishi yishi.  
Ref: today is january 5th of the lunar new year, i am glad to be invited by the chinese gold & silver exchange society to participate in your spring festival celebration and preside over the opening ceremony of the new year gold market.  
ANMT: today, i am very pleased to receive the invitation of the gold and silver trade to attend the ceremony of your new year and preside over the ceremony for the new year.  
+ETR: today, i am glad to be invited by the gold and silver trade association to attend the spring festival and preside a ceremony for the new year.  
+LTR: today, i am glad to be invited by the gold and silver trade association to attend your spring festival and preside over a ceremony for the new year.

Example 2

Src: 广东是中国内地经济外向度最高的省份，已经率先感受到  
(pinyin) guangdong shi zhongguo neidi jingji waixiangdu zui gao de shengfen, yijing shuixian ganshou dao  
前所未有的发展机遇和严峻挑战。  
qiansuoweyou de fazhan jiyu he yanjun tiaozhan.  
Ref: guangdong boasts the most outward-oriented economy among all the provinces in the chinese mainland, it is the first to see the unprecedented opportunities of development and serious challenges alike.  
ANMT: guangdong is the largest province in china: a inland economy, and has taken the lead to feel the unprecedented development opportunities and severe challenges.  
+ETR: guangdong is the largest province in chinese' s mainland economy, and has taken the lead to see unprecedented development opportunities and severe challenges.  
+LTR: guangdong is the largest province in china: s mainland economy, and has taken the lead to see unprecedented development opportunities and severe challenges.

图 5 具体翻译示例

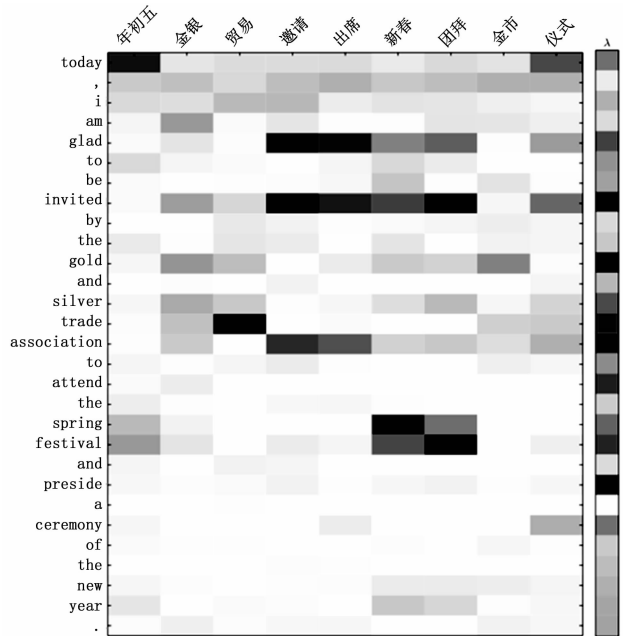


图 6 结果对齐可视化

#### 4 结束语

本研究探讨了源话题信息对句子层面语境的依赖性，并提出了一种话题注意方法，将潜在话题表征整合到现有的 NMT 结构中，以提高翻译预测能力。在未来的工作中，将利用显式的源主题信息来增强 NMT 中的目标词预测。

#### 参考文献:

[1] 朱晓宁. 基于中轴语的统计机器翻译研究 [D]. 哈尔滨: 哈尔滨工业大学, 2019.  
[2] Cao Z, Li S, Liu Y, et al. A Novel Neural Topic Model and Its Supervised Extension [A]. Twenty-ninth AAAI Conference on Artificial Intelligence [C]. AAAI Press, 2015: 2210-2216.  
[3] 薛擎天, 李军辉, 贡正仙. 多语言的无监督神经机器翻译 [J]. 厦门大学学报 (自然科学版), 2020, 59 (2): 192-197.  
[4] Chen W, Matusov E, Khadivi S, et al. Guided Alignment

Training for Topic-Aware Neural Machine Translation [A]. Proc. 38th Annu. Meeting Symp. Antenna Meas. Techn [C]. Assoc., Austin, TX, USA, 2016: 121-134.

[5] 李 平, 张路遥, 曹 霞, 等. 基于潜在主题的混合上下文推荐算法 [J]. 电子与信息学报, 2018, 40 (4): 957-963.

[6] Chen K, Wang W, Utiyama U, et al. Neural Machine Translation With Sentence-Level Topic Context [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2019, 27 (12): 1970-1984.

[7] Cho K, Van Merriënboer B, Gulcehre C, et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation [J]. Computer Science, 2014: 1724-1734.

[8] Calixto I, Liu Q, Campbell N. Doubly-Attentive Decoder for Multi-modal Neural Machine Translation [A]. Proc. 55th Annu. Meeting Assoc. Comput. Linguistics [C]. 2017: 1913-1924.

[9] Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need [A]. Proceeding of the 31st Conference on Neural Information Processing Systems (NIPS) [C]. Long Beach, CA, USA, 2017: 998-1013.

[10] 叶雪梅, 毛雪岷, 夏锦春, 等. 文本分类 TF-IDF 算法的改进研究 [J]. 计算机工程与应用, 2019, 55 (2): 104-109, 161.

[11] Koehn P, Hoang H, Birch A, et al. Moses: Open source tool-

kit for statistical machine translation [A]. Proc. 45th Annu. Meeting Assoc. Computat. Linguistics Companion Volume Proc. Demo Poster Sessions [C]. 2007: 177-180.

[12] Zhang J, Li L, Way A, et al. Topic-informed neural machine translation [A]. Proc. 26th Int. Conf. Comput. Linguistics [C]. 2016: 1807-1817.

[13] 闫铁珊. 基于关键字和类型信息的实体检索方法研究 [D]. 西安: 西安理工大学, 2019.

[14] Zhang J, Liu Q, Zhou J. ME-MD: An Effective Framework for Neural Machine Translation with Multiple Encoders and Decoders [A]. Twenty-sixth International Joint Conference on Artificial Intelligence [C]. 2017: 3392-3398.

[15] Collins M, Koehn P, and Kucerová I. Clause restructuring for statistical machine translation [A]. Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (ACL' 05) [C]. Association for Computational Linguistics, USA, 2005: 531-540.

[16] Sennrich R, Haddow B, Birch A. Neural Machine Translation of Rare Words with Subword Units [A]. Proc. 54th Annu [C]. Meeting Assoc. Comput. Linguistics, 2016: 1715-1725.

[17] Post M. A Call for Clarity in Reporting BLEU Scores [A]. Proc. 3rd Conf. Mach. Transl.: Res. Papers, Association for Computational Linguistics [C]. Belgium, Brussels, 2018: 186-191.

(上接第 139 页)

发, 使用该系统进行晶圆测试方案开发调试的开发流程与原有开发流程对比, 有效地缩短了测试方案开发周期, 并且在探针台管理、实时监控等方面也更具优势, 具体见表 2。

表 2 本软件开发模式与原有开发模式应用性对比

模式	开发周期 /天	探针台管理	监控方式	客户调试
原有测试开发模式	7~14	人工查看	人工监控	需要工程师协助
探针台远程控制模式	5~7	通过软件实时管理	软件实时反馈探针台状况	客户独立调试

## 5 结束语

该软件从 2019 年投入使用后, 截至 2020 年 4 月, 已经参与近 15 个晶圆测试项目的开发流程, 软件运行良好、操作方便且效果显著, 具有良好的可靠性、可移植性、可交互性, 其中内网穿透效果良好, 穿透率达到 95%。

本文结合晶圆测试与探针台设备的实际情况, 基于现有探针台设备以及 GPIB 等硬件资源的基础上设计开发了一

个探针台的内外网 NAT 穿透远程控制管理软件, 由于该软件与不同类型探针台的标准指令相关度不高, 在软件开发过程中使用模块化开发, 可以在后续使用中针对新的设备进行推广<sup>[8]</sup>。软件对多站点同步自动化管理、测试数据的统计分析还需要进一步的研究应用。

## 参考文献:

[1] 裴万里, 郝先人, 袁远东, 等. 一种低成本 UHF RFID 芯片集中 CP 测试方法 [J]. 宇航计测技术, 2013 (6): 79-83.

[2] 童志义. 探针测试设备简介 [J]. 电子工业专用设备, 2008 (2): 32-35.

[3] 赵立军. 基于 VISA 接口的可编程电源输出控制程序 [J]. 计量与测试技术, 2015, 42 (12): 1-2.

[4] 王红艺, 李冠峰, 吴庆涛, 等. 一种人机界面组件组装方法的研究 [J]. 河南科技大学学报 (自然科学版), 2011 (1): 43-46.

[5] 梅 黎. 基于软交换的 NAT/防火墙穿透技术研究 [J]. 现代电信科技, 2005 (1): 30-32.

[6] Ping Y, Run-Zhong Y I, Liang T. Server Internet/Intranet Dynamic Mapping Based on DDNS and NAT [J]. Computer Engineering, 2008, 34 (20): 136-137.

[7] 蔡勇华, 邓 招, 宋 滔, 等. 基于通用总线的频谱监测设备设计与实现 [J]. 通信技术, 2018, 51 (2): 501-505.