

文章编号: 1000-5862(2020)01-0046-05

基于 QE 的机器翻译重排序方法研究

翟煜锦, 李培芸, 项青宇, 李茂西*, 袁白莲, 钟茂生, 王明文

(江西师范大学计算机信息工程学院, 江西 南昌 330022)

摘要: 该文提出了一种融合 BERT 语境向量的多模型集成的翻译质量估计方法, 以及基于译文质量估计的多候选译文重排序方法, 实验结果表明, 这 2 种方法均取得了较好的实验效果.

关键词: 机器翻译; 机器翻译质量估计; 重排序; 编码器-解码器模型; 卷积神经网络

中图分类号: TP 391 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2020.01.08

0 引言

第 15 届全国机器翻译学术会议(CCMT2019)组织的机器翻译评测项目一共包括 4 个子任务: 双语翻译评测任务、多语翻译评测任务、语音翻译评测任务和翻译质量估计评测任务. 江西师范大学作为参评单位之一, 参加了其中 2 个子任务: 机器翻译评测任务和句子级别翻译质量估计评测任务. 该文介绍了笔者参加这 2 个评测任务中所使用的参评系统和相关技术, 以及在评测方发布的测试集上系统性能.

1 模型

针对机器翻译子任务和机器翻译质量估计子任务, 本节将分别介绍提交的参评系统以及采用的主要方法.

1.1 机器翻译质量估计

句子级别翻译质量估计(Quality Estimation, QE)一般视为机器学习中的回归问题, 其任务目标是利用从源语言句子和机器译文中抽取出描述译文质量的特征(包括反映翻译复杂度、流利度和忠实度的特征等)来估计译文质量^[1]. 在本次翻译质量估计评测中, 采用了多模型融合的策略, 即融合联合神经网络模型的结果和基于卷积神经网络的质量估计器的结果来提高机器评分与人工评分之间的相关性.

1.1.1 联合神经网络模型 联合神经网络模型(UNQE)^[2-4]由 2 个子网络构成: 基于双向循环神经

网络(Recurrent Neural Network, RNN)的编码器-解码器模型^[5](特征提取模型)和计算 QE 得分的 RNN 模型(估计器). 特征提取模型通过翻译上下文信息提取机器译文质量向量, 随后将其提供给估计器来预测机器译文的质量得分. UNQE 将上述 2 部分子网络组合成一个整体网络, 并通过质量估计训练集联合训练这 2 部分神经网络中的权重.

双向 RNN 编码器-解码器模型是一个序列到序列的模型^[6], 模型的编码器部分将输入的源语言句子 $\mathbf{x} = (x_1, x_2, \dots, x_m)$ 映射到一个固定长度的向量. 该向量输入到解码器端, 解码器端按时序输出翻译结果 $\mathbf{y} = (y_1, y_2, \dots, y_n)$. 通过下式计算出每个词的条件概率:

$$p(y_i | \{y_1, y_2, \dots, y_{i-1}\}, \mathbf{x}) = g(y_{i-1}, s_{i-1}, \mathbf{c}_i) = \exp(\mathbf{y}_i^T \mathbf{W}_0 \mathbf{t}_i) / \sum_{k=1}^{K_y} \exp(\mathbf{y}_k^T \mathbf{W}_0 \mathbf{t}_i), \quad (1)$$

其中 g 是预测 y_i 的非线性函数, \mathbf{c}_i 是源语言的上下文向量, s_{i-1} 是 RNN 的隐藏层状态, $y_i \in \mathbf{R}^{k_y \times 1}$ 是以 one-hot 表示的目标语言词或子词, K_y 是目标语言的词表大小, $\mathbf{W}_0 \in \mathbf{R}^{k_y \times d}$ 是 softmax 方法的权重矩阵, $\mathbf{t}_i \in \mathbf{R}^{d \times 1}$ 是中间表示, 它包含了机器翻译目标语言在第 i 位上可能出现词的所有信息, d 是目标语言的词向量维数.

预测词的方法是利用最大似然估计方法计算目标语言端下一个词的概率^[7], 但是在本文的模型中, 双向 RNN 编码器-解码器模型^[8-9]被用来提取描述机器译文质量的特征, 即“质量向量”^[10]. 目标词 y_i ($i = 1, 2, \dots, m$) 的质量向量表示机器译文中

收稿日期: 2019-09-12

基金项目: 国家自然科学基金(61662031, 61462044, 61877031, 61876074)资助项目.

通信作者: 李茂西(1977-), 男, 湖北黄梅人, 教授, 博士, 主要从事自然语言处理和机器翻译的研究. E-mail: mosesli@jxnu.edu.cn

词语 y_i 是否是源语言中词语正确翻译的信息,因此将(1)式做了修改,它可以通过下式计算:

$$q_{y_i} = ((y_i^T W_0) \odot t_i^T)^T, i = 1, 2, \dots, m, \quad (2)$$

其中 $W_0 \in \mathbf{R}^{k_y \times d}$ 是一个权重矩阵, $t_i \in \mathbf{R}^{d \times 1}$ 为中间表示,它包含了机器翻译目标语言在第 i 位上可能出现词的所有信息,将该信息输入到解码器端可以得到 y_i 在目标语言第 i 位上的质量向量. 运算符 \odot 表示 2 个矩阵对应元素的乘积.

特征提取模块通过翻译上下文向量以及循环神经网络的隐含层状态,计算出每个单词的质量向量. 将得到的质量向量输入到估计器模型中,通过下式计算得出整句译文的质量估计得分:

$$v_i = g(v_{i-1}, q_{y_i}), i = 1, 2, \dots, m, Q_{E_{score}} = W_{QE} v_n,$$

其中 v_i 为 RNN 第 i 个时刻的隐含层状态,函数 g 为非线性激活函数, W_{QE} 为权重矩阵.

1.1.2 基于卷积神经网络的质量估计器 通过 UNQE 特征提取模型提取的译文中每个词语的质量向量,描述了其与源语言句子的映射关系,反映了机器译文的翻译忠实度. 但是,除了忠实度,机器译文质量的优劣还与其流利度相关. J. Devlin 等^[11] 提出的 BERT 语言模型在大规模的单语语料上进行训

练,给定具体的机器译文(设输入长度为 c 的目标词序列: t_1, t_2, \dots, t_c),由它推导的语境向量(长度为 c 的目标词序列对应生成 c 个 BERT 语境向量)描述了译文的流利度. 兼顾译文翻译忠实度与流利度,提出了基于卷积神经网络的融合 BERT 语境词向量和质量向量的翻译质量估计方法.

为了结合 BERT 语境词向量与质量向量,将每个目标词对应的 2 种向量分别输入卷积神经网络中,计算出描述整个译文的忠实度向量和流利度向量,然后将忠实度向量和流利度向量连接,输入单个节点的全连接层,使用 sigmoid 激活函数输出译文质量得分. 模型的整体结构如图 1 所示,图 1 中左下、右下部分分别描述了如何通过词语的质量向量计算出译文的忠实度向量、如何通过词语的 BERT 语境词向量计算出译文的流利度向量. 基于卷积神经网络的质量估计器主要由 2 部分子网络组成,2 个子网络均包含 2 层卷积层、池化层和 1 层压平层 (Flatten),其中池化层采用最大池化操作. 为捕捉更加丰富的语言特征,对 2 个子网络均设置了 (2,3,4) 3 种不同的卷积窗口,并且卷积核个数均为 100 个.

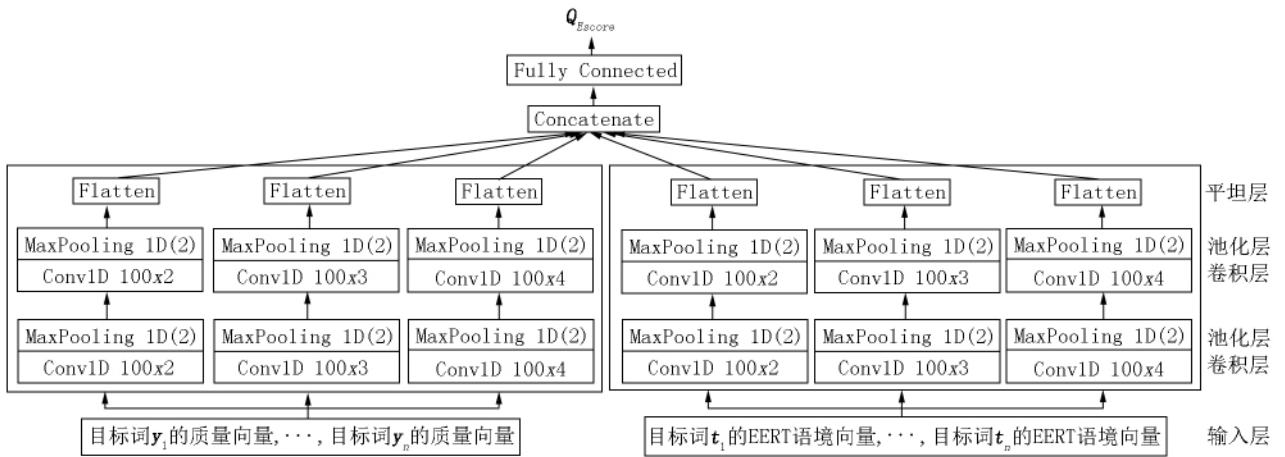


图 1 CCNN 模型结构图

将图 1 所示的模型称为结合上下文信息的卷积神经网络模型 (ContextualConvolution Neural Networks, CCNN). 网络中词语的质量向量由(2)式计算,BERT 语境词向量使用谷歌 Devlin 等发布的 BERT 语言模型计算. BERT 语境词向量包含 12 个不同隐含层的输出,根据谷歌发布的提取 BERT 语境词向量方法,提取 4 个隐含层(-1, -2, -3, -4)的语境词向量分别引入 CCNN 方法中,简称为 $CCNN^{-1}$, $CCNN^{-2}$, $CCNN^{-3}$ 和 $CCNN^{-4}$; 同时,为了充分利用 BERT 语言模型中每个隐含层蕴含的上下文信息,对上述提取的语境向量进行了平均池化运算,并将其结果作为特征引入到译文质量估计中,简称

为 $CCNN^{AVG}$.

经过压平层处理后,得到表征译文整体的质量向量和语境向量,将其相连后的结果使用 sigmoid 激活函数输出得到译文质量估计得分 $Q_{E_{score}}$. 给定 Q_E 任务的训练集,在训练过程中,CCNN 的优化目标为在训练集上最小化质量估计得分 $Q_{E_{score}}$ 与人工评价得分 H_{TER} 的平方差.

1.2 基于译文质量估计的机器翻译重排序模型

在新闻领域机器翻译评测任务中,开源的神经翻译模型 Transformer 被使用对源语言句子进行解码,解码生成源语言句子的多个候选译文. 使用译文

质量估计方法,对多个候选译文进行评分.以译文的评分结果作为依据,对候选译文进行重排序,将排序后的最优译文作为最终的输出,通过该方法提高输出译文的质量.

Transformer 是 A. Vaswani 等^[12]提出的完全基于自注意力机制的神经机器翻译方法.与基于 RNN 的编码器-解码器模型相比,Transformer 不依赖任何循环操作,并行处理序列中的所有单词或符号,同时利用自注意力机制将上下文与较远的单词结合起来,大大加快了模型训练速度,同时也提升了翻译性能.利用 Transformer 模型,采取 2 种不同的方式生成多个候选译文:

(i) 指定最后 n 次保存的模型进行解码 (n -checkpoint). Transformer 默认保存最后 20 次训练模型,尝试使用最后 10 次保存的模型,对源语言句子逐个利用每个模型的参数进行解码,集合每个模型的最优输出结果组成多个候选译文.

(ii) 指定最好的模型输出 n -best 翻译假设 (n -best). 在解码时,利用最后一次保存的模型并使用解码选项“return_beams = True”指定对源语言句子解码,输出前 n 个最大概率的候选译文组成多个候选译文.

尝试使用上述 2 种方式生成源语言句子的多个候选机器译文,根据其在开发集上的性能来选择采用何种方式来生成多个候选译文.

2 实验

2.1 实验设置

评测中使用的计算机软硬件环境配置情况如表 1 所示.在译文质量估计评测任务中,除了使用 CCMT 官方提供的训练数据,还使用了 CWMT18 评测方提供的用于新闻翻译任务的双语平行语料.其规模统计如表 2 所示.

表 1 计算机操作系统与硬件配置

操作系统	CPU	内存	GPU
	Intel(R) Core(TM)		
Ubuntu16.04 LTS	i5-6500	32G	GeForce GTX 1080
	CPU 3.20GHz		

表 2 译文质量估计评测数据集统计(句子数)

	方向	英汉	汉英
训练集	CCMT	14 789	10 070
	WMT	6.0 M	6.0 M
	总计	6.1 M	6.1 M
开发集	CCMT	1 381	1 381

在新闻领域机器翻译评测任务中,使用的语料完全来自于 CCMT2019 官方提供的训练数据.数据规模统计如表 3 所示.

表 3 机器翻译评测数据集统计(句子数)

方向	英汉	汉英
训练集	9 M	9 M
开发集	1 500	2 481

为了防止数据稀疏,对上述数据分别进行了预处理.对英文数据的预处理包括:标记化、大小写转换和句子 BPE 子词切分^[13];在对中文语料进行预处理时采用斯坦福中文分词工具进行分词.另外,在译文质量估计评测任务中,剔除了长度超过 70 的句子,挑选出源语言句子与目标语言句子长度比值在 $[1/3, 3]$ 范围内的句子.在提取句子中词语的 BERT 语境词向量时,使用谷歌预训练好的语言模型.

在翻译质量估计模型训练过程中,固定质量向量维数为 700, BERT 语境词向量维数为 768,使用谷歌发布的 BERT-Base、Uncased 语言模型提取英中方向的语境词向量,使用 BERT-Base、Chinese 语言模型提取中英分方向的语境词向量.在训练 CCNN 质量估计器时,隐含层单元设置为 100,训练过程使用 rmsprop 优化器,训练批处理大小为 108,并使用 dropout($r_{ate}=0.2$) 机制防止过拟合.

在机器翻译模型训练过程中,使用标准的 Transformer 框架^[14]进行模型训练.在训练时使用单个 GPU,批处理大小为 2 048,训练 50 万次.每训练 1 000 次对模型进行一次保存.Checkpoint 文件中保存最后 20 个模型的文件路径,其余参数均为默认参数设置.

2.2 实验结果

2.2.1 句子级别译文质量估计任务 在 UNQE 的实验中,将质量向量分别设定为 500 维和 700 维进行训练,得到不同的译文质量估计结果(分别记为 $UNQE_{500}$ 和 $UNQE_{700}$),最后将上述结果以算术平均的方式进行融合,记为 $Ave_{500+700}$.

对于基于卷积神经网络的质量估计器 CCNN,采用 700 维质量向量和 BERT 语境词向量作为输入. CCNN 模型左子网络以 700 维质量向量为输入,右子网络则分别以 BERT 语言模型的 4 个隐含层 ($-1, -2, -3, -4$) 语境词向量及其平均池化结果为输入,将上述结果分别记为 $CCNN^{-1}$ 、 $CCNN^{-2}$ 、 $CCNN^{-3}$ 、 $CCNN^{-4}$ 和 $CCNN^{AVG}$.最后,将上述结果以算术平均的方式进行融合,记为 $CCNN^{ENSEMBLE}$. CCNN 系统和 UNQE 系统在开发集上的实验结果如表 4 所示.

表 4 CCNN 系统、UNQE 系统在开发集上的性能得分

模型		英汉		汉英	
		Pearson r	Spearman	Pearson r	Spearman
CCNN	CCNN ⁻¹	0.48	0.40	0.49	0.44
	CCNN ⁻²	0.53	0.40	0.51	0.45
	CCNN ⁻³	0.54	0.41	0.50	0.45
	CCNN ⁻⁴	0.53	0.38	0.49	0.45
	CCNN ^{AVG}	0.53	0.41	0.50	0.45
	CCNN ^{ENSEMBLE}	0.54	0.40	0.52	0.44
UNQE	UNQE ₅₀₀	0.42	0.30	0.53	0.47
	UNQE ₇₀₀	0.46	0.30	0.52	0.49
	Ave ₅₀₀₊₇₀₀	0.48	0.33	0.54	0.49
	ENSEMBEL	0.55	0.42	0.56	0.49

在该方法基础上,将 Ave₅₀₀₊₇₀₀ 和 CCNN^{ENSEMBLE} 2 种方法的结果以算术平均方式进行融合. 融合后的系统在开发集上的性能如 “ENSEMBEL”行所示. 实验结果表明: 将 UNQE 系统与 CCNN 系统进行集成,有效地提高了机器译文质量估计与人工评价评分的相关性. 因此,将该集成方法在测试集上对机器译文的评分作为最终评分进行提交. CCMT 官方给出的在测试集上的最终评测结果如表 5 所示.

表 5 多模型集成的质量估计方法在测试集上的性能

方向	Pearson r	Spearman
英汉	0.36	0.32
汉英	0.51	0.42

2.2.2 新闻领域机器翻译任务 在机器翻译评测任务中,首先通过训练得到基线系统 Baseline. 然后,采用 2 种不同的方式来生成多个候选译文,第 1 种候选译文生成方法通过指定最后 10 个模型参数进行解码,生成候选机器译文,记为 10-checkpoint; 第 2 种方法使用在开发集上表现最好的模型,通过设置解码选项直接生成 10-best 候选译文. 再在生成候选译文后,使用译文质量估计模型分别对这 10 个候选译文进行句子级别评分. 最后,以该评分为排序依据,对于每个源语言句子的所有候选译文,均选择排序后的最优译文进行最终输出. 该方法在开发集上的实验结果如表 6 所示(该得分使用 Transformer 的 BLEU 脚本计算得到).

表 6 新闻领域翻译任务在开发集上的系统性能

模型	英汉	汉英
	BLEU(character)	BLEU(Uncased)
Baseline	32.18	22.86
Rerank _{10-checkpoint}	31.86	21.06
Rerank _{10-best}	33.24	23.91

实验结果表明: 与基线系统相比,在不借助人工参考译文的情况下,仅利用源语言句子的信息对 n -best 候选译文进行质量估计并重排序,可以提高机器译文的输出质量. 然而,使用 n -checkpoint 方法生成的候选译文,经过重排序后并没有提高机器译文的输出质量. 因此,在测试集上使用第 2 种方法生成多个候选译文,并在其基础上利用译文质量估计方法对其重排序,挑选最优译文作为最终输出译文.

CCMT 官方最终公布的在测试集上的评测结果如表 7 所示.

表 7 新闻领域翻译任务在测试集上的系统性能

模型	英汉	汉英
	BLEU(character)	BLEU(Uncased)
Baseline	45.9	41.9
Rerank _{10-best}	46.6	42.8

2.3 实验分析

为了揭示基于译文质量估计的重排序方法提高最终输出译文质量的原因,随机挑选一组结果进行分析,表 8 给出了 Transformer 模型输出的 n -best 候选译文原始排序(排序列),以及每一个源语言句子的 BLEU^[15] 得分(BLEU 列) 和 QE 得分(QE 列). 表 8 结果表明: Transformer 模型在解码时输出条件概率最大的候选译文,即 1-best. 但对比所有 10-best 候选译文的 BLEU 和 QE 得分结果可以看出,翻译系统给出的 1-best 候选译文的翻译质量不一定是最优的. 同时发现, QE 的评价结果与 BLEU 的评价结果在一定程度上具有一致性,即 BLEU 值高的句子,对应的 QE 值较低. 因此,在缺乏人工参考译文的情况下,利用 QE 对 n -best 候选译文进行质量估计,是 BLEU 的一种有效替代方案. 因此,以其评分结果为依据从多个候选译文中挑选最优译文,显著提高了机器输出译文的质量.

表 8 不同方法对 n -best 候选译文的打分(排序)结果对比

源语言句子	摩洛哥足球协会去年 8 月向国际足联递交了承办 2026 年世界杯的正式申请.			
人工参考译文	The Moroccan Football Association submitted a formal application to FIFA last August to host the 2026 World Cup.			
Transformer 模型输出的 10-best 译文		排序	BLEU	QE
Transformer 模型输出的 10-best 译文	Moroccan Football Association submitted its official application to FIFA last August for hosting the World Cup of 2026.	1	0.40	0.151
	Moroccan Football Association submitted its official application to FIFA last August to host the World Cup of 2026.	2	0.58	0.147
	Moroccan Football Association submitted its official application to FIFA last August for hosting the World Cup in 2026.	3	0.40	0.155
	Moroccan Football Association submitted its official application to FIFA last August to host the World Cup in 2026.	4	0.58	0.152
	Moroccan Football Association submitted its official application to FIFA in August last year to host the World Cup of 2026.	5	0.30	0.167
	The Moroccan Football Association submitted its official application to FIFA last August for hosting the World Cup of 2026.	6	0.44	0.134
	The Moroccan Football Association submitted its official application to FIFA last August to host the World Cup of 2026.	7	0.60	0.130
	Moroccan Football Association submitted its official application to FIFA in August last year for hosting the World Cup of 2026.	8	0.26	0.170
	The Moroccan Football Association submitted its official application to FIFA in August last year to host the World Cup of 2026.	9	0.35	0.145
	The Moroccan Football Association submitted its official application to FIFA in August last year to host the World Cup in 2026.	10	0.35	0.151
原始输出译文 (1-best)	Moroccan Football Association submitted its official application to FIFA last August for hosting the World Cup of 2026.	1	0.40	0.151
重排序后输出译文	The Moroccan Football Association submitted its official application to FIFA last August to host the World Cup of 2026.	7	0.60	0.130

3 总结

本文主要介绍了江西师范大学参加 CCMT2019 机器翻译相关评测的情况. 在句子级译文质量估计任务中,提出了多模型集成的方式:将 UNQE 和 CCNN 模型进行集成,可以有效地提高译文质量估计的效果;在新闻领域机器翻译任务(汉英方向和英汉方向)中,提出了基于译文质量估计的多候选译文重排序方法,实验结果表明该方法能更准确地挑选出最优译文,特别是在汉英翻译方向,使用重排序的方法显著提高了翻译的质量.

4 参考文献

- [1] Specia L, Shah K, De Souza J G C, et al. QuEst: a translation quality estimation framework [EB/OL]. [2019-05-05]. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.386.3135>.
- [2] Li Maoxi, Xiang Qingyu, Chen Zhiming, et al. A unified

neural network for quality estimation of machine translation [J]. *IEEE Transactions on Information and Systems*, 2018, 101(9): 2417-2421.

- [3] Shah K, Bougares F, Barrault L, et al. SHEF-LIUM-NN: sentence level quality estimation with neural network features [EB/OL]. [2019-03-06]. <https://www.aclweb.org/anthology/W16-2392.pdf>.
- [4] 陈志明, 李茂西, 王明文. 基于神经网络特征的句子级译文质量估计 [J]. *计算机研究与发展*, 2017, 54(8): 1804-1812.
- [5] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate [EB/OL]. [2019-03-09]. <https://arxiv.org/abs/1409.0473>.
- [6] Sennrich R, Firat O, Cho K, et al. Nemo: a toolkit for neural machine translation [EB/OL]. [2019-03-13]. <https://arxiv.org/abs/1703.04357>.
- [7] 宗成庆. 统计自然语言处理 [M]. 北京: 清华大学出版社, 2008.
- [8] 刘洋. 神经机器翻译前沿进展 [J]. *计算机研究与发展*, 2017, 54(6): 1144-1149.
- [9] 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): 2734-2755. (下转第 88 页)

The Spatio-Temporal Characteristics of Carbon, Nitrogen and Phosphorus in Surface Soil of Tideland in Poyang Lake Wetlands

XIE Dongming¹, YI Qing¹, ZHOU Guohong¹, HUANG Qinghua¹, ZHOU Yangming², JIA Junsong^{2*},
QIANian Haiyan³, ZHU Zaiyu^{1*}, YE Zhanghuang¹

(1. School of Tourism, Jiangxi Science and Technology Normal University, Nanchang Jiangxi 330038, China; 2. College of Geography and Environment, Key Laboratory of Poyang Lake Wetland and Watershed Research, Ministry of Education, Jiangxi Normal University, Nanchang Jiangxi 330022, China; 3. School of Earth Sciences, East China University of Technology, Nanchang Jiangxi 330013, China)

Abstract: Carbon, nitrogen and phosphorus are important nutrient elements in soil of wetland tideland, and their distribution characteristics directly affect the productivity and ecosystem service function of wetland ecosystem. Based on the observation and analysis of organic carbon, total nitrogen and total phosphorus in surface soil (0 ~ 20 cm) of tideland in Poyang Lake wetlands with sampling from 2014 to 2017, the results show that the interannual variation characteristics of organic carbon, total nitrogen and total phosphorus in surface soil in tideland of Poyang Lake wetlands are different. The change of organic carbon is not significant, but the change of total nitrogen and total phosphorus is significant. The changes of organic carbon, total phosphorus and total nitrogen in surface soil are very significant when gradient is various. The interannual changes of C/N ratio and C/P ratio in surface soil are very significant, but the N/P ratio is not significant. The changes of C/N ratio in surface soil is not significant, but the changes of C/P ratio and N/P ratio are very significant when gradient is various. The content of nitrogen and phosphorus is higher than that in other regions, however, the ratio of C/N, C/P and N/P in surface soil are lower than those in other regions in China with interannual changes various gradient changes. The factors that affected carbon, nitrogen and phosphorus in surface soil of tideland in Poyang Lake Wetlands are complicated and perennial. The changes of vegetation cover, flooding time and water quality of Poyang Lake wetlands have influence on the spatial and temporal distribution of soil carbon, nitrogen and phosphorus.

Key words: organic carbon; total nitrogen; total phosphorus; tideland; Poyang Lake wetlands (责任编辑: 刘显亮)

(上接第 50 页)

- [10] Kim H, Jung H-Y, Kwon H, et al. Predictor-estimator: neural quality estimation based on target word prediction for machine translation [J]. ACM Transactions on Asian and Low-Resource Language Information Processing, 2017, 17 (1): 1-22.
- [11] Devlin J, Chang M W, Lee K, et al. Bert: pre-training of deep bidirectional transformers for language understanding [EB/OL]. [2019-03-20]. <https://arxiv.org/abs/1810.04805?context=cs>.
- [12] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [EB/OL]. [2019-04-06]. <https://arxiv.org/abs/1706.03762>.
- [13] Sennrich R, Haddow B, Birch A. Neural machine translation of rare words with subword units [EB/OL]. [2019-03-17]. <https://arxiv.org/abs/1508.07909>.
- [14] Vaswani A, Bengio S, Brevdo E, et al. Tensor2tensor for neural machine translation [EB/OL]. [2019-03-16]. <https://arxiv.org/abs/1803.07416>.
- [15] Papineni K, Roukos S, Ward T, et al. BLEU: a method for automatic evaluation of machine translation [EB/OL]. [2019-03-19]. <https://dl.acm.org/citation.cfm?id=1073135>.

The Study on the Method of Machine Translation Reordering Based on QE

ZHAI Yujin, LI Peiyun, XIANG Qingyu, LI Maoxi*, QIU Bailian, ZHONG Maosheng, WANG Mingwen
(College of Computer Information Engineering, Jiangxi Normal University, Nanchang Jiangxi 330022, China)

Abstract: A multi-model ensemble quality estimation method is proposed, which integrates the BERT context vectors. Based on quality estimation, a multi-candidate translation reordering method has been proposed. The experimental results show that both methods have achieved good results.

Key words: machine translation; machine translation quality estimation; reordering; encoder-decoder model; convolutional neural network (责任编辑: 冉小晓)