计算机科学
COMPUTER SCIENCE

翻译错误类分布加权的专利译文自动后编辑集成模型

赵三元, 王裴岩, 叶娜, 赵欣瑜, 蔡东风, 张桂平

引用本文

赵三元, 王裴岩, 叶娜, 赵欣瑜, 蔡东风, 张桂平. 翻译错误类分布加权的专利译文自动后编辑集成模型[J]. 计算机科学, 2023, 50(11A): 230300072-8.

ZHAO Sanyuan, WANG Peiyan, YE Na, ZHAO Xinyu, CAI Dongfeng, ZHANG Guiping. Automatic Post-editing Ensemble Model of Patent Translation Based on Weighted Distribution of Translation Errors [J]. Computer Science, 2023, 50(11A): 230300072-8.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.230300072

翻译错误类分布加权的专利译文自动后编辑集成模型

赵三元　王裴岩　叶　娜　赵欣瑜　蔡东风　张桂平
沈阳航空航天大学人机智能研究中心　沈阳 110136
(17393890485@163.com)

摘　要　自动后编辑(APE)是一种自动修改机器译文错误的方法,能够改善机器翻译系统的译文质量。目前,APE研究主要集中于通用领域,然而对于专业性强和译文质量要求较高的专利译文的APE则鲜有研究。文中研究了专利译文自动后编辑,提出了翻译错误类分布加权的专利译文自动后编辑集成模型。首先,提出术语加权翻译编辑率(WTER)计算方法,在翻译编辑率(TER)中加入了每个词的术语概率因子,使术语错误较多的样本WTER值较高。然后,通过WTER从3个机器翻译系统构造的训练数据中选择错译、漏译、增译与移位错误样本子集分别构建错误修正偏向性APE子模型。最后,通过翻译错误类分布加权错误修正偏向性APE子模型。该方法针对专利专业性、强术语较多的特点,每个子模型分别面向一类错误,考虑了错误修正的偏向性,通过模型集成兼顾了译文错误多样性,在英中专利摘要数据集上的实验结果表明,相比3个基线系统,所提方法的BLEU值分别平均提升了2.52,2.28和2.27。

关键词:自动后编辑;专利译文;翻译错误类分布;集成;翻译编辑率
中图法分类号　TP391

Automatic Post-editing Ensemble Model of Patent Translation Based on Weighted Distribution of Translation Errors

ZHAO Sanyuan, WANG Peiyan, YE Na, ZHAO Xinyu, CAI Dongfeng and ZHANG Guiping
Human-Computer Intelligence Research Center, Shenyang Aerospace University, Shenyang 110136, China

Abstract　Automatic post-editing(APE) is a method of automatically modifying errors in machine translation, which can improve the quality of machine translation system. Currently, APE research mainly focuses on general domains. However, there is little research on APE for patent translations, which requires high translation quality due to their strong professionalism. This paper proposes an ensemble model of APE of patent translation based on the weighted distribution of translation errors. Firstly, the term weighted translation edit rate(WTER) calculation method is proposed, which introduces the concept of term probability factor in translation edit rate(TER), and improves the WTER value of samples with more term errors. Then, the proposed WTER model is used to select subsets of mistranslation, missing translation, additional tralslation and shift error samples from the training data constructed by the three machine translation systems to construct the error correction biased APE sub-model, respectively. Finally, the biased APE sub-model is corrected by the weighted distribution of translation errors. The proposed method considers the strong professionalism and numerous technical terms in patent translations. Based on the consideration of error-correction bias, it integrates multiple sub-models to balance the diversity of translation errors. Experimental results on an English-Chinese patent abstract dataset show that, compared with the three baseline systems, the proposed method improves the BLEU values by an average of 2.52, 2.28, and 2.27, respectively.

Keywords　Automatic post-editing, Patent translation, Distribution of translation errors, Ensemble, Translation edit rate

1　引言

现阶段机器翻译(Machine Translation, MT)具有良好的翻译表现,但是对于专业性强的专利翻译,仍然需要人工校对来保证译文质量,与机器翻译相适应的译后编辑工作模式在全球语言服务行业发挥着越来越重要的作用[1]。自动后编辑(Automatic Post-Editing, APE)[2]是一种可实现自动修改机器译文的方法,可让机器进一步分担译后工作量[3]。然而,目前APE研究主要集中于通用领域,而对于专利译文这类需要较多后编辑的APE任务却鲜有研究。

Guan等[1]指出,在专业术语要求非常高的专利翻译领域,机器翻译目前无法达到令人满意的质量。Dong等[4]指出术语翻译效果仍存在较大的提升空间,其翻译结果存在着大量术语漏翻、错翻的现象。Xu等[5]指出专利文献中术语的使用非常密集,凸显出专利术语翻译工作的重要性。本文对6 610句专利机器译文错误进行了统计,其中40.95%为术语引

基金项目:国家自然科学基金(U1908216);教育部人文社会科学研究青年基金(19YJC740107);沈阳市科学技术计划(20-202-1-28)
This work was supported by the National Natural Science Foundation of China(U1908216), Education of Humanities and Social Science Research on Youth Fund Project(19YJC740107) and Shenyang Science and Technology Plan(20-202-1-28).
通信作者:王裴岩(wangpy@sau.edu.cn)

起的翻译错误，此外，机器译文主要有 4 类错误[6]，错译占 51.02%、漏译占 21.56%、增译占 16.16%、移位占 11.26%。研究表明，对词的后编辑是 APE 主要的纠错方向[7]，因此对于专利 APE 模型，修正术语翻译错误是关键问题。

现有的 APE 模型采用基于 Transformer[8] 的编码器-解码器[9]架构，模型训练需要大量的训练数据。APE 模型训练数据是由源文、机器译文与后编辑译文构成的三方数据。其中，后编辑译文是对机器译文进行人工后编辑的结果，不易获得较多的人工编辑数据，专利 APE 数据尤为缺乏。现有研究从训练策略和训练数据构造两个方面解决 APE 模型训练问题。

在训练策略方面，“预训练＋微调”是基本策略，以 Correia[10] 为代表的研究中，首先利用预训练模型（如多语言 BERT[11]）来初始化模型的参数，然后利用三方数据对参数进行微调。该方法在专利中可以利用少量的专利 APE 训练数据达到和大规模数据训练接近的效果。然而微调模式下模型性能依赖于专利 APE 微调样本，导致模型不易捕获专利机器译文的普遍性错误。

在训练数据构造方面，研究人员提出了两种自动产生三方数据的方法。一方面，利用数据增强[12-13]产生伪机器译文，即加入噪声数据来模拟机器译文错误。虽然该方法可以产生较多的数据，但是伪机器译文中的错误并不能完全反映真实机器译文错误的特点，使模型不易修正专利机器译文的真实性错误。

另一方面，将双语句对中的源文与目标译文作为 APE 三方数据的源文和后编辑译文，然后利用 MT 系统翻译源文得到三方数据的机器译文。虽然双语句对中的目标译文不是对机器译文人工后编辑的结果，但也是一种正确译文，能够反映机器译文的错误情况，如 eSCAPE[14] 数据集。然而，单一 MT 系统产生的专利机器译文具有错误偏向性，使得 APE 模型易过拟合于产生数据的 MT 系统特定翻译错误，难以捕获专利机器译文普遍性错误。Meng 等[3] 指出受控训练学习到的纠错规则不一定适用于所有场景，有时反而会降低 MT 译文的质量。

基于以上分析，本文提出了一种翻译错误类分布加权的专利自动后编辑集成模型。首先，改进翻译编辑率（Translation Edit Rate，TER）[6]，提出了术语加权 TER（term Weighted TER，WTER），加入了每个词的术语概率因子，增大了术语翻译错误对 TER 值的影响，使得术语翻译错误较多的样本 TER 值较高。然后，通过 WTER 从 3 个机器翻译系统构造的训练数据中分别选择错译、漏译、增译与移位错误样本子集，构建错误修正偏向性 APE 子模型。最后，通过翻译错误类分布比例加权集成上述各子模型。

上述方法在构建子模型时考虑了术语翻译错误因素，针对专利专业性强、术语较多的特点，每个子模型分别面向一类错误，考虑了错误修正的针对性，通过模型集成兼顾了错误多样性。在 3 个机器翻译系统上的实验结果显示，所提方法能够更好地捕获专利机器翻译系统的普遍性错误并加以修正，没有过拟合于某一机器翻译系统。在 3 个机器翻译系统上，平均 BLEU 值提升了 5.55，平均 TER 值降低了 3.46。相比基线 APE 模型平均 BLEU 值分别提升了 2.52，2.28 和2.27，平均 TER 值降低了 2.6，3.84 和 3.22。

本文第 2 章介绍了所提模型；第 3 章介绍了实验方法与结果分析；最后总结全文并展望未来。

2 APE 集成模型

2.1 翻译错误类分布加权的专利 APE 集成模型

集成学习是一种联合多个学习器进行协同决策的机器学习方法[15]，可以有效结合多个学习器提升模型预测的准确性。通常有模型平均组合[16] 和模型加权组合等方法。为了使专利 APE 模型能兼顾专利机器译文的多样性错误和针对性错误，本文提出了一种翻译错误类分布加权的专利 APE 集成模型。该方法的优点是，通过专利机器译文翻译错误类分布来确定各子模型权重，可以更好地结合各子模型修正过程的贡献度，从而综合考虑不同子模型的决策结果，提高修正模型的准确性。本节将具体介绍翻译错误类分布加权的专利 APE 集成模型，其整体架构如图 1 所示。

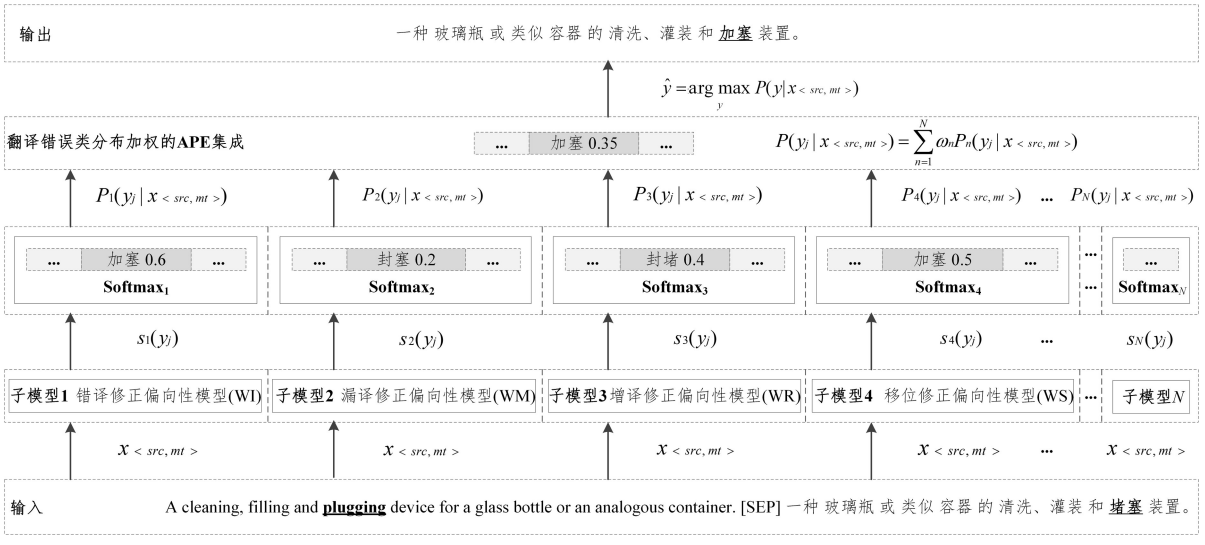

图 1 翻译错误类分布加权的专利 APE 集成模型图

Fig. 1 Diagram of APE ensemble model of patent translation based on weighted distribution of translation errors

与神经机器翻译相比，本文的 APE 模型修正过程是：首先将源文（src）、机器译文（mt）与连接符“[SEP]”进行连接，形式为〈src [SEP] mt〉，作为输入；其次，经过多个错误修正偏向性 APE 子模型进行加权集成；最后输出修正后的译文。如图 1 所示，英文短语“plugging device”的中文为“加塞装置”，“加塞”特指给瓶子充填加塞和封口的专业术语[17]，假如

该词的机器译文为“堵塞”，该译文并不准确，经过本文的APE模型得到修正后的译文为“加塞”。即输入表示为 $x_{\langle src,mt\rangle}$，得到预测概率最大的修正后的目标译文 $\hat{y}$，该过程可用建模公式(1)描述。

$$\hat{y}=\arg\max_{y} P(y|x_{\langle src,mt\rangle}) \tag{1}$$

将 N 个子模型的预测概率进行加权集成，进一步将其转换为求目标译文中每个词的预测概率，该过程用式(2)描述。

$$P(y_j|x_{\langle src,mt\rangle})=\sum_{n=1}^{N}\omega_n P_n(y_j|x_{\langle src,mt\rangle}) \tag{2}$$

其中，$P_n(y_j|x_{\langle src,mt\rangle})$代表第 n 个子模型所预测的目标译文 $\hat{y}$ 中第 j 个词 y_j 的预测概率，ω_n 代表权重，经过 N 个子模型的加权集成得到对 y_j 的预测概率 $P(y_j|x_{\langle src,mt\rangle})$。若定义第 n 个子模型对第 j 个词 y_j 的解码输出向量为 $\boldsymbol{s}_n(y_j)$，经过 Softmax 映射到词表大小为 V 的分布上。其计算式如式(3)所示：

$$P_n(y_j|x_{\langle src,mt\rangle})=\mathrm{Softmax}_n(\boldsymbol{s}_n(y_j)\cdot\boldsymbol{W}_o) \tag{3}$$

其中，$\boldsymbol{W}_o$ 是线性变换矩阵，大小是 $d\times V$，d 为解码向量输出维度。ω_n 通过翻译错误类分布来确定各个子模型的集成权重。

2.2 错误修正偏向性 APE 子模型

为了构建翻译错误类分布加权的专利 APE 集成模型，需要合理地选择不同翻译错误类的训练数据来训练 APE 子模型作为集成的基模型。本文基于专利机器译文的特点，提出了一种错误修正偏向性 APE 子模型的训练方法。

2.2.1 模型结构

通过 N 类译文错误可得到 N 个子模型。对于机器译文中主要存在的错译(Incorrect，I)、漏译(Miss，M)、增译(Redundant，R)和移位(Shift，S) 4 类基本错误，融入每个词的术语概率因子，分别训练错译修正偏向性(Weighted Incorrect，WI)、漏译修正偏向性(Weighted Miss，WM)、增译修正偏向性(Weighted Redundant，WR)和移位修正偏向性(Weighted Shift，WS) 4 个 APE 子模型，子模型结构基于 Transformer，如图 2 所示。

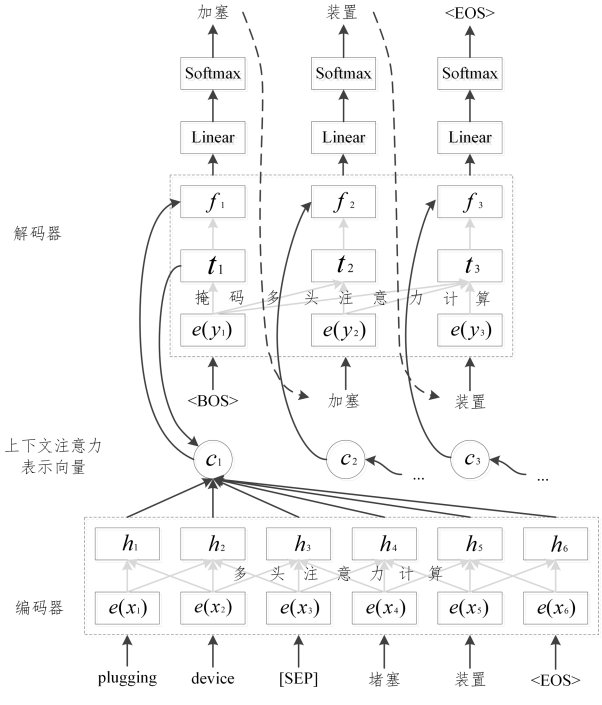

图 2 APE 子模型 Transformer 结构图

Fig. 2 Structure of APE sub-model Transformer

Transformer 是由 6 层相同的编码器和解码器堆叠而成，图 2 中以 1 层为例展示子模型 Transformer 结构图。即源文 $x_{\langle src,mt\rangle}$ 中的词经过编码得到 $e(x_i)$，进入编码器后经过多头注意力计算得到隐层向量 $\boldsymbol{h}_i$ 并送入解码器；解码器中，第 n 个子模型对目标译文的序列预测计算用建模公式(4)表示。

$$P_n(y|x_{\langle src,mt\rangle})=\prod_{j=1}^{l}P_n(y_j|y_{<j},x_{\langle src,mt\rangle}) \tag{4}$$

其中，l 指要生成目标译文的长度，$y_{<j}$指目标译文第 j 个位置前生成的译文序列，表达式的含义是通过结合源文 $x_{\langle src,mt\rangle}$ 和前 j 个位置已生成的序列 $y_{<j}$，来预测当前第 j 个词 y_j 的生成概率。因此解码过程中，$P_n(y_j|y_{<j},x_{\langle src,mt\rangle})$的计算进一步可转换为式(5)。

$$P_n(y_j|y_{<j},x_{\langle src,mt\rangle})=P_n(y_j|t_{j-1},c_{j-1}) \tag{5}$$

其中，t_{j-1}表示在预测第 j 个词时，前 $j-1$ 个词在解码器的输入表示 $e(y_{j-1})$进行多头注意力计算的表示结果与编码器多头注意力计算不同，解码器需要加入掩码矩阵来防止透露未来信息。c_{j-1}表示对编码器的输出和解码器的输入 t_{j-1}进行上下文注意力计算的结果。

此外，〈BOS〉是句子开始标志符，〈EOS〉是句子结束标识符。以生成第一个词的过程为例，解码过程可以描述为：通过对〈BOS〉的多头注意力表示结果 t_1 和编码器多头注意力表示结果进行计算得到上下文注意力结果 c_1，并将其送入前馈神经网络层 f_1，再经过线性层，最后经 Softmax 转换得到第一个译文单词，随后依次生成每个单词，直到遇到句子结束标识符。

2.2.2 基于术语翻译错误类的训练数据选择

本文以术语翻译错误类为基础分别选择具有各类代表性错误的训练数据。机器译文错误类可定义为，N 类译文错误类集合 $Type=\{type_1,\cdots,type_n,\cdots,type_N\}$，其中 $type_n$ 为某一类错误。则机器译文中存在的错译(I)、漏译(M)、增译(R)和移位(S) 4 类错误类集合即为 $Type=\{I,M,R,S\}$。仍以“plugging device”的译文“加塞装置”为例，其可能出现的 4 类错误如表 1 所列。

表 1 专利机器译文错误示例

Table 1 Examples of patent machine translation errors

错误类型	示例
错译(I)	堵塞 装置
漏译(M)	__塞 装置
增译(R)	加塞塞 装置
移位(S)	装置 加塞

本文利用了多个翻译系统，为每条源文产生机器译文，由于不同翻译系统的错误具有一定的偏向性，因此需要选择不同错误分布的样本。即对于 M 个机器翻译系统，则有机器译文集合 $MT=\{mt_1,mt_2,\cdots,mt_m,\cdots,mt_M\}$，从 MT 中选择一条 mt_m 构建三方数据。

本文基于 TER 来选择训练数据。TER 是一种衡量译文质量的评价指标，它指将候选译文修改为参考译文所需要的最少编辑次数，对应于 APE 中是将机器译文修改为后编辑译文，其计算公式如式(6)所示：

$$TER=\frac{edit(c_{\mathrm{mt}},r_{\mathrm{pe}})}{l_{\mathrm{pe}}} \tag{6}$$

其中，$edit(c_{\mathrm{mt}},r_{\mathrm{pe}})$指从机器译文 c_{mt} 修改为后编辑译文 r_{pe} 所

需要的编辑操作数，操作包含替换、删除、增加和移位，对应于I，M，R和S 4类错误，l_{pe}指后编辑译文的长度。虽然该方法可以得到准确的整体错误编辑操作求和的结果，但是该方法没有区分4类翻译错误分布的结果，且该方法将所有词的编辑代价均视为1，并没有考虑术语翻译错误因素在训练数据选择中的作用。

基于以上分析，为了考虑错误分布，同时增加对术语翻译错误样本的选择权重，本文对句子中的每个词计算该词为术语的术语概率因子，这样就使得越接近术语的词的编辑代价也越大，增大了选择术语翻译错误样本的可能性，并按照错误类分别计算TER值，其改进后的计算公式如式(7)所示：

$$TER_{type_n}(k)=\frac{\sum_{i=1,(w_i(k)\in Err_{type_n}(k))}(p_i+\tau)}{\sum_{i=1}^{l_{pe}}(p_i+\tau)} \tag{7}$$

其中，$TER_{type_n}(k)$代表第k个句子某一类错误$type_n$的翻译编辑率，分母为后编辑译文中每个词的术语概率因子p_i的代价和，分子为机器译文相比后编辑译文错误词$w_i(k)$的代价和，$w_i(k)$为第k个句子中错误类为$type_n$词的集合$Err_{type_n}(k)$，l_{pe}代表后编辑译文的长度。考虑每个词存在0概率值的情形，在计算术语概率因子时加入常数τ对其进行平滑。

为了计算术语概率因子p_i，本文首先基于Word2Vec[18]来得到机器译文和后编辑译文中每个词的词向量；其次，利用单语术语表计算术语词向量；最后，分别对机器译文和后编辑译文句子中每个词的词向量$\boldsymbol{e}_i$与术语表中每个术语词向量$\boldsymbol{e}_j$计算余弦相似度，第i个词语与第j个术语的余弦相似度计算过程如式(8)所示：

$$sim_i(\boldsymbol{e}_i,\boldsymbol{e}_j)=\frac{\boldsymbol{e}_i\cdot\boldsymbol{e}_j}{|e_i|\times|e_j|} \tag{8}$$

若术语词表大小为V_t，则第i个词计算得到V_t个余弦相似值，即第i个词的余弦相似度向量$\boldsymbol{Sim}_i$用式(9)描述。

$$\boldsymbol{Sim}_i=(\boldsymbol{sim}_i(\boldsymbol{e}_i,\boldsymbol{e}_j)_1,\cdots,\boldsymbol{sim}_i(\boldsymbol{e}_i,\boldsymbol{e}_j)_{V_t}) \tag{9}$$

从中取前K个最大的余弦值概率进行平均作为第i个词的术语可能性概率，则p_i的计算式如式(10)所示：

$$p_i=\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{sim}_i(\boldsymbol{e}_i,\boldsymbol{e}_j)_k \tag{10}$$

改进方法在数据选择时相比传统的方法有两个优点：(1)按错误分布对存在术语或术语翻译错误较多的句子赋予更高的权重，并将其作为训练样本，使模型更好地学习术语翻译纠错知识；(2)对未知样本句子中术语概率因子的计算具有更好的泛化能力。

3 实验与分析

3.1 实验数据

3.1.1 专利英中双语句对数据

本文实验数据集是欧洲专利局质检过的英中专利摘要数据集，包含机械、计算机、医学、化学和生物等领域，通过句对齐得到共计8470句，按照78%，11%与11%的比例随机分为训练集、验证集及测试集。各数据集的句子数(“句”)、句平均字数(“字-中”)与句平均词数(“词-中”与“词-英”)分布情况如表2所列。

表2 英中专利数据集介绍

Table 2 Introduction to English-Chinese patent datasets

统计量	训练集	验证集	测试集	总量
句	6610	930	930	8470
词-英	38.2	38.2	37.9	323399
字-中	49.3	48.6	48.9	416628
词-中	28.2	27.9	28.0	238599

3.1.2 APE数据

本文基于专利英中双语句对构造专利APE数据。专利APE数据中的源文(英)为专利英中双语句对训练数据的源文，目标句作为后编辑译文(中)。机器译文则通过课题组开发的PTS(Patent Translation System)[1)]、百度翻译系统(Baidu)[2)]和谷歌翻译系统(Google)[3)]得到。

PTS利用专利英中双语摘要语料(该语料相比构建APE的训练数据，虽然未经欧洲专利局质检，但是经过了人工校对)和WMT(Workshop on Machine Translation，WMT)语料，共计3998万句，训练了面向专利的MT系统。

PTS，Google和Baidu 3个翻译系统得到机器译文句平均词数与字数分布如表3所列。

表3 机器译文统计信息

Table 3 Machine translation statistics

系统	统计量	训练集	验证集	测试集
PTS	字	46.4	46.0	45.9
	词	26.4	26.1	26.3
Google	字	49.3	49.0	49.2
	词	28.0	27.9	27.9
Baidu	字	50.4	49.8	50.2
	词	28.7	28.5	28.5

由表3可见，Baidu系统机器译文句平均词数与字数最大，而PTS系统的则最小。Google系统的机器译文句平均词数及字数与目标译文最接近(见表2)。

3.2 对比方法

为了验证本文方法的有效性，在训练数据选择时按照源文数据的数量即按照源文数据的100%选择数据训练APE模型。将本文提出的翻译错误类分布加权的专利APE集成模型(WIMRS)与以下方法进行对比。

机器翻译(MT)：为了验证本文方法能否提升机器译文质量，本文对比了PTS系统、Baidu和Google翻译系统。

APE方法[10]：本文对比了Correia等[10]的APE方法[4)]，该方法利用了多语言Bert预训练模型，可以在少量的训练数据上进行微调达到和大规模数据训练接近的效果。本文利用构造的有限三方数据集(PTS、Baidu和Google)训练了APE_{PTS}，APE_{Baidu}和APE_{Google} 3个模型。

随机抽样训练数据选择方法(Rand)[19]：从同一源文对应的3个机器译文中随机选择一个机器译文构造三方数据训练模型，以验证训练数据选择的必要性。

数据增强方法(Data Augmentation，DA)[12]：为了探究选择真实机器译文构建训练数据的必要性，将本文方法与DA方法进行了对比，即在后编辑译文中插入噪声，构造出源文、伪机器译文和后编辑译文的三方数据训练APE模型。

1) https://www.koalatrans.com

2) https://fanyi.baidu.com/translate

3) https://translate.google.cn

4) https://github.com/deep-spin/OpenNMT-APE

基于余弦相似度的训练数据选择方法(Cosine)[13]：为了探究选择 APE 训练数据时考虑机器译文质量因素的必要性，将本文方法与 Cosine 方法进行了对比。通过计算同一后编辑译文对应 3 个机器译文的余弦值，选择余弦值较小的机器译文，即从 3 个机器译文中选择与后编辑译文最不相似的机器译文来构建三方数据训练 APE 模型。

基于翻译编辑率训练数据选择方法(TER)[6]：为了探究基于术语翻译错误类选择子数据集构建并集成子模型的必要性，将本文方法与 TER 方法进行了对比，主要的实现过程是计算同一后编辑译文对应的 3 个机器译文的 TER 值，选择 TER 值较大的机器译文，构建三方数据。

3.3 模型参数设置

在数据选择时，利用了 Word2Vec 中基于 Skip Gram 实现的腾讯词向量[20]计算余弦相似度，实验中 K 值为 5 得到的概率值较为准确，此外，实验的中文单语术语表大小为 46 万，平滑值 τ 取值为 0.1。训练模型时发现专利数据集上的训练步数在 100 000 步以内能达到较好的结果，以 10 000 步作为模型检查点保存。模型中的部分参数设置如表 4 所列。

表 4 超参数设置

Table 4 Hyperparameter setting

超参数	取值范围或类型
暂退率(Dropout Rate)	0.1
标签平滑值(Label Smoothing)	0.1
学习率(Learning Rate)	0.000 05
优化器(Optimizer)	Adam
训练批量大小(Batch Size)	512
验证批量大小(Valid Batch Size)	8
预热步数(Warm up Steps)	5 000

3.4 评价指标

本文使用机器翻译自动后编辑任务中常用的评价指标 BLEU(Bilingual Evaluation Understudy，BLEU)[21]值和 TER 值来评价自动后编辑的效果。BLEU 值越大或者 TER 值越小，代表对机器译文的自动后编辑效果越好。

3.5 实验结果和分析

3.5.1 与现有方法的对比实验

表 5 列出了对比实验结果。与其他方法相比，本文方法显著优于其他对比方法，表明了其优越性。利用本文方法对原始机器译文进行后编辑，使 3 种机器译文的 BLEU 值(MT)都有提升，分别提高 19.71%，15.36%，10.22%。平均 BLEU 值提升了 5.55，TER 值降低了 3.46。

表 5 不同 APE 方法的对比实验结果

Table 5 Comparative experimental results of different APE methods

模型	PTS		Baidu		Google		平均	
	BLEU	*TER*	*BLEU*	*TER*	*BLEU*	*TER*	*BLEU*	*TER*
WIMRS	**42.33**	**40.19**	**42.74**	**40.49**	**43.12**	**39.63**	**42.73**	**40.10**
MT	35.36	44.37	37.05	44.16	39.12	42.15	37.18	43.56
APE_{PTS}	41.46	41.24	39.36	43.91	39.82	42.96	40.21	42.70
APE_{Baidu}	38.75	48.64	42.17	41.19	40.44	41.99	40.45	43.94
APE_{Google}	38.65	46.86	40.52	42.31	42.20	40.79	40.46	43.32
Rand	40.37	41.79	40.83	41.92	41.34	41.28	40.85	41.66
DA	36.23	44.18	36.75	45.53	38.40	43.78	37.13	44.50
Cosine	40.57	42.04	40.83	41.97	40.84	41.33	40.75	41.78
TER	40.63	41.69	40.75	41.76	41.15	41.29	40.84	41.58

通过进一步分析实验结果可得出如下结论：

(1)本文提出的方法能够更好地捕获专利机器翻译系统的普遍性错误并加以修正，没有过拟合于某一机器翻译系统。然而，针对特定机器翻译系统训练的 APE 模型具有一定的偏向性，过拟合于各自系统。在 3 个机器翻译系统上，所提方法皆优于利用文献[10]的方法得到的 3 个 APE 模型(APE_{PTS}，APE_{Baidu} 与 APE_{Google})，平均 BLEU 值分别提升了 2.52，2.28 和 2.27，平均 TER 值也均有所降低。虽然 3 种机器译文上分别训练的 APE 模型都可以显著地提升原始机器译文的质量，使 BLEU 值分别提高了 6.1，5.12 和 3.08，但是这 3 个模型对于其他机器翻译系统产生的译文纠错效果皆不如在各自系统上的效果。

(2)所提出的翻译错误类分布加权的专利自动后编辑集成模型优于构建单一模型。各子模型对于不同类错误更具有针对性。模型集成在确保有效性的同时兼顾错误针对性及多样性，而单一模型较难控制。WIMRS 与 TER 相比，BLEU 提升 1.99，TER 值降低 1.48。

(3)在 DA 数据上较难训练出针对真实机器译文错误情况的 APE 模型。所有数据选择方法皆优于 DA，原因是 DA 所构造的伪机器译文与真实机器译文的错误分布不一致。

(4)在选择 APE 模型训练数据时考虑机器译文质量因素是必要的。并且，考虑质量因素时，具有专利机器翻译错误针对性的数据评价指标更具有优势，能够选择具有专利机器翻译模型错误代表性的训练数据，从而训练更为有效的 APE 模型。从实验结果可知，WIMRS 优于随机选择。并且，考虑错误分布的方法优于词匹配度的 Cosine 方法。

3.5.2 数据选择量验证实验

本文进一步实验了训练数据选择量对 WIMRS 方法的影响。实验结果如图 3、图 4 所示。其中，100%指对于每个源文从 3 个机器翻译系统生成的译文集中仅选择一条机器翻译译文；200%指对于每个源文从 3 个机器翻译系统生成的译文集中选择出两条机器翻译译文；300%指 3 个机器翻译系统生成的译文集都加入训练数据，采用随机排序[22]策略训练模型。从图 3、图 4 中可见，相比 100%的结果，数据选择量提升至 200%时，各模型的效果皆有提升。当数据量选择达到 300%，模型效果远远不如数据选择后的模型。实验结果反映出 WIMRS 能够有效地挑选出训练专利 APE 模型的数据。

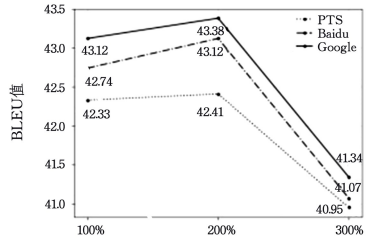

图 3 不同数据量下 WIMRS 方法的 BLEU 值

Fig. 3 BLEU values of WIMRS method with different data volumes

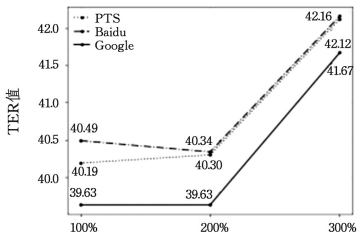

图 4 不同数据量下 WIMRS 方法的 TER 值

Fig. 4 TER values of WIMRS method with different data volumes

3.5.3 子模型集成方法验证实验

本文还比较了基于术语翻译错误分布的多个子模型集成方法的实验结果，如表 6 所列。其中 WI(错译)、WM(漏译)、WR(增译)和 WS(移位)是分别基于 $WTER_I$，$WTER_M$，$WTER_R$ 与 $WTER_S$ 选择训练数据训练的 4 个子模型。$WIMRS_{Avg}$ 指上述 4 个子模型采用平均加权[16]的集成方法。从表 6 可知，WIMRS 的 BLEU 值比 $WIMRS_{Avg}$ 高 0.11，TER 则降低 0.16，这表明基于错误分布子模型加权策略要优于平均策略。WIMRS 与 $WIMRS_{Avg}$ 都优于 4 个子模型。其原因是，机器译文的错误分布不同而产生的偏向性问题，进一步验证了在专利 APE 模型上应通过模型加权策略来兼顾针对性与多样性。WM(漏译)在 PTS 系统上的效果显著优于其他子模型。WI(错译)在 Baidu 系统上的效果较好。Google 系统上 4 个子模型的效果差异较小。

表 6 子模型集成实验结果

Table 6 Experimental results of sub-model ensemble

模型	PTS		Baidu		Google		平均	
	BLEU	*TER*	*BLEU*	*TER*	*BLEU*	*TER*	*BLEU*	*TER*
WIMRS	**42.33**	**40.19**	42.74	**40.49**	**43.12**	**39.63**	**42.73**	**40.10**
$WIMRS_{Avg}$	41.98	40.43	**42.77**	40.68	43.10	39.67	42.62	40.26
$WTER_I$	40.33	41.68	41.34	41.78	41.50	41.03	41.06	41.50
$WTER_M$	41.36	41.07	40.80	42.39	41.27	41.49	41.14	41.65
$WTER_R$	40.11	42.21	41.26	41.79	41.51	40.92	40.96	41.64
$WTER_S$	40.54	41.77	40.93	41.80	41.52	41.21	41.00	41.59

图 5 给出了 WIMRS 模型面向不同机器翻译系统的子模型加权参数。

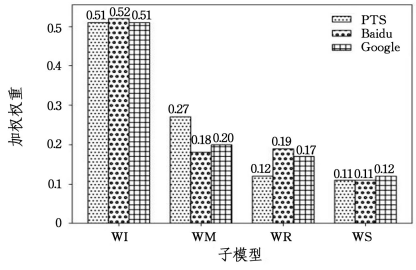

图 5 各子模型加权权重图

Fig. 5 Weighted weight diagram of each submodel

WI 模型具有最高的加权参数值，由于加权参数是通过错误分布比例计算，因此反映出词错译是专利机器译文中主要的错误。PTS 翻译系统的 WM 加权参数值相较于 Baidu 和 Google 系统更高，而 WR 相对较小，表明 PTS 存在更多的漏译错误，增译错误较少。结合表 6 的实验结果，可一步反映出不同机器翻译系统的错误偏向性。在子模型加权参数设置时需要考虑各子系统错误分布。

3.5.4 术语翻译错误类加权有效性验证实验

为了验证所提出的术语翻译错误类加权方法的有效性，将 WIMRS 和 $WIMRS_{Avg}$ 与 IMRS 和 $IMRS_{Avg}$ 进行了对比。IMRS 和 $IMRS_{Avg}$ 与 WIMRS 和 $WIMRS_{Avg}$ 的子模型集成策略一致，分别采用错误分布加权及平均加权策略。其差别是子模型构建时训练数据选择策略采用了 TER 而非考虑了词术语可能性的 WTER。实验结果如表 7 所列。

表 7 术语翻译错误类加权方法有效性验证实验结果

Table 7 Experimental results of the validation of weighted method of term translation errors

模型	PTS		Baidu		Google		平均	
	BLEU	*TER*	*BLEU*	*TER*	*BLEU*	*TER*	*BLEU*	*TER*
WIMRS	**42.33**	40.19	42.74	**40.49**	**43.12**	**39.63**	**42.73**	**40.10**
$WIMRS_{Avg}$	41.98	40.43	**42.77**	40.68	43.10	39.67	42.62	40.26
IMRS	42.32	**40.08**	42.14	40.85	42.68	40.25	42.38	40.39
$IMRS_{Avg}$	42.25	40.28	42.20	40.77	42.63	40.04	42.36	40.36

相比 WIMRS 与 $WIMRS_{Avg}$，IMRS 的 BLEU 值降低 0.35，$IMRS_{Avg}$ 的 BLEU 值降低 0.26。这表明，面向专利文本术语较多和难翻译的特点，数据选择中需要考虑术语翻译质量。

本文提出的 WTER 及在此基础上提出的 WIMRS 更好地考虑了专利特点，更适于面向专利 APE 模型的数据选择与数据利用。

3.6 实例分析

本文给出了一个实例(见表 8)，从而体现 WIMRS 的自动后编辑应用效果。其中，SRC 代表源句，PE 表示后编辑译文，PTS，Baidu 和 Google 分别是 3 个 MT 系统产生的机器译文，APE_{PTS}，APE_{Baidu} 和 APE_{Google} 分别为在 3 种机器译文上训练得到的 APE 模型的 APE 译文。表 8 中，加粗字体为译文中出现错误的部分。可以发现，本文方法可以

将其修改正确，APE_{PTS}模型虽然也能将其修改正确，却引入了漏译等新问题。

表 8 实例分析

Table 8 Case analysis

SRC	The present invention discloses a cleaning, filling and **plugging** device for a glass bottle or an analogous container, which comprises a bottle feeding device, a bottle cleaning machine, a filling and **plugging** machine and a bottle discharging device which are orderly connected, wherein the bottle feeding device is connected with a bottle feeding mechanism in the bottle cleaning machine, and a bottle group conveying device is arranged between the bottle cleaning machine and the filling and **plugging** machine.
PE	本发明公开了一种玻璃瓶或类似容器的清洗、灌装和**加塞**装置，它包括依次相连的送瓶装置、洗瓶机、灌装**加塞**机以及出瓶装置，送瓶装置与洗瓶机中的进瓶机构相连，洗瓶机与灌装**加塞**机之间设有成组传送装置。
WIMRS	本发明公开了一种玻璃瓶或类似容器的清洗、灌装和加塞装置，它包括依次连接的进瓶装置、洗瓶机、灌装加塞机和出瓶装置，送瓶装置与洗瓶机中的送瓶机构相连，在洗瓶机和灌装加塞机之间设有瓶组输送装置。
PTS (2022.12.13)	本发明公开了一种玻璃瓶或类似容器的清洗、灌装、**封塞**装置包括依次连接的进瓶装置、洗瓶机、灌装**封堵**机和出瓶装置、与洗瓶机中的进瓶机构连接的进瓶装置、与灌装**封堵**机之间设置的瓶组输送装置。
APE_{PTS}	本发明公开了一种玻璃瓶或类似容器的清洗、灌装和加塞装置，包括依次连接的进瓶装置、洗瓶机、灌装加塞机和**加塞**装置，与洗瓶机中的进瓶机构相连的进瓶装置，在[**漏掉洗瓶机**]灌装加塞机之间设有瓶组输送装置。
Baidu (2022.12.13)	本发明公开了一种玻璃瓶或类似容器的清洁、灌装、**堵塞**装置，包括依次连接的进瓶装置、洗瓶机、灌装、**堵塞**机和出瓶装置，其中进瓶装置与洗瓶机中的进瓶机构连接，在洗瓶机与灌装**堵漏**机之间设有瓶组输送装置。
APE_{Baidu}	本发明公开了一种玻璃瓶或类似容器的清洗、灌装、加塞装置，包括顺序连接的进瓶装置、洗瓶机、灌装**堵塞**机和出瓶装置，将送瓶装置与洗瓶机中的送瓶机构相连，在洗瓶机与灌装**堵塞**机之间设有瓶组输送设备。
Google (2022.12.13)	本发明公开了一种玻璃瓶或类似容器的清洗灌装**塞**装置，包括依次连接的进瓶装置、洗瓶机、灌装**塞**机和卸瓶装置。送料装置与洗瓶机中的送瓶机构连接，洗瓶机与灌装**塞**之间设有瓶组输送装置。
APE_{Google}	本发明公开了一种玻璃瓶或类似容器的清洗、灌装和**塞塞**装置，它包括依次连接的进瓶装置、洗瓶机、灌装**塞塞**机和出瓶装置，所述送瓶装置与洗瓶机中的送瓶机构相连，在洗瓶机与灌装**塞**机之间设有瓶组输送装置。

结束语 针对专利机器译文的高质量翻译需求，本文提出了一种翻译错误类分布加权的专利 APE 集成模型，即通过专利机器译文的错误类分布作为权重来集成 APE 子模型；通过术语翻译错误类选择训练数据构造 APE 子模型；此外，利用多个翻译系统产生机器译文构建了一个面向专利 APE 的三方数据语料库。与多种方法相比，本文方法在专利的术语翻译错误修正方面有较大的提升。本文的研究结果表明：(1)选择 APE 模型训练数据时考虑机器译文的质量因素是必要的。并且，考虑质量因素时，具有专利翻译错误针对性的数据评价指标更具有优势；(2)模型集成有助于提升 APE 的性能。并且，专利中基于翻译错误类分布加权的集成在确保有效性的同时兼顾了错误针对性及多样性，而单一模型较难控制。

本文证明了基于翻译错误类分布加权集成的有效性，并且研究发现术语翻译错误和错译仍然是专利机器翻译和 APE 要重点解决的问题。未来将继续探究其他多模型集成的策略，如引入门控机制和强化学习等方法，有效利用错误类分布来优化权重以充分关注纠错针对性，提升 APE 集成模型对术语错误和专利错译等质量问题的修正能力。

参 考 文 献

[1] GUAN F X, FEI Y N. Prospect Analysis of Patent Translation in Man-Machine Age [J]. China Invention & Patent, 2019, 16(11): 64-67.

[2] SIMARD M, UEFFIFING N, ISABELLE P, et al. Rule-Based Translation with Statistical Phrase-Based Post-Editing [C]// Proceedings of the Second Workshop on Statistical Machine Translation. 2007: 203-206.

[3] MENG F Y, TANG X R. Efficiency First: Reviewing Technologies of Machine Translation Post-Editing [J]. Computer Engineering and Applications, 2020, 56(22): 25-32.

[4] DONG Z H, REN W P, YOU X D, et al. Machine Translation Method Integrating New Energy Terminology Knowledge [J]. Computer Science, 2022, 49(6): 305-312.

[5] XU P W, LENG B B. Common Difficulties and Practical Strategies in English Translation of Patent Terms [J]. Chinese Science & Technology Translators Journal, 2019, 32(4): 28-31.

[6] SNOVER M G, DORR B J, SCHWARTZ R M, et al. A Study of Translation Edit Rate with Targeted Human Annotation [C]// Proceedings of the 7th Conference of the Association for Machine Translation in the Americas. 2006: 223-231.

[7] DO CARMO F, SHTERIONOV D, MOORKENS J, et al. A review of the state-of-the-art in automatic post-editing [J]. Machine Translation, 2020(2): 1-43.

[8] SHISH V, NOAM S, NIKI P, et al. Attention is all you need [C]// Advances in Neural Information Processing Systems. 2017: 5998-6008.

[9] SUTSKEVER I, VINYALS O, LE Q V. Sequence to Sequence Learning with Neural Networks [J]. Advances in Neural Information Processing Systems, 2014, 20: 3104-3112.

[10] CORREIA G M, MARTINS A. A Simple and Effective Approach to Automatic Post-Editing with Transfer Learning [C]// Proceedings of the 57th Conference of the Association for Computational Linguistics. 2019: 3050-3056.

[11] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding [C]// Proceedings of the 2019 Conference of the North Ameri-

can Chapter of the Association for Computational Linguistics. 2018.

[12] LEE W K, SHIN J, JUNG B, et al. Noising scheme for data augmentation in automatic post-editing[C]// Proceedings of the Fifth Conference on Machine Translation. 2020: 783-788.

[13] CAI Z L, YANG M M, XIONG D Y. Data Augmentation for Neural Machine Translation [J]. Journal of Chinese Information Processing, 2018, 32(7): 30-36.

[14] MATTEO N, MARCO T, RAJEN C, et al. eSCAPE: a Large-scale Synthetic Corpus for Automatic Post-Editing[C]// Proceedings of the Eleventh International Conference on Language Resources and Evaluation in Proceedings of LREC, 2018.

[15] HANSEN L K, SALAMON P. Neural network ensembles [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2002, 12(10): 993-1001.

[16] LI B, WANG Q, XIAO T, et al. On Ensemble Learning of Neural Machine Translation [J]. Journal of Chinese Information Processing, 2019, 33(3): 42-51.

[17] QIU Y M, YANG N S, LIU Z, et al. Cleaning, Filling and Plugging Device for Glass Bottle or Analogous Container: CN200943034Y[P]. 2007.

[18] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient Estimation of Word Representations in Vector Space[J]. Computer Science, 2013.

[19] BALCÁZAR J, DAI Y, WATANABE O. A random sampling Technique for training support vector machines[C]// International Conference on Algorithmic Learning Theory. Berlin, Heidelberg: Springer, 2001.

[20] SONG Y, SHI S, LI J, et al. Directional Skip-Gram: Explicitly Distinguishing Left and Right Context for Word Embeddings [C]// Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers). 2018.

[21] KISHORE P, SALIM R, TODD W, et al. Bleu: a Method for Automatic Evaluation of Machine Translation[C]// Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. 2002: 311-318.

[22] KUANGS H, XIONGD Y. The Influence of Different Use of Training Corpus on Neural Machine Translation Model [J]. Journal of Chinese Information Processing, 2018, 32(8): 53-59, 67.

ZHAO Sanyuan, born in 1997, postgraduate. His main research interests include NLP and machine translation.

WANG Peiyan, born in 1983, Ph.D, senior engineer, is a member of China Computer Federation. His main research interests include NLP, machine learning and knowledge engineering.