

基于预训练语言模型微调的DJI Osmo Pocket 3 用户手册中英翻译研究

曾炳凤

上海海事大学外国语学院, 上海

收稿日期: 2026年3月18日; 录用日期: 2026年5月11日; 发布日期: 2026年5月25日

摘要

本文旨在利用预训练语言模型(Pre-Trained Language Models, PLM)对DJI Osmo Pocket 3用户手册进行中英文双语翻译优化。通过构建消费电子领域平行语料, 并结合翻译语言模型(TLM)与交叉注意力掩码语言模型(CAMLM)进行定向微调, 提高模型在专业技术文本中的翻译准确性与术语一致性。在数据处理阶段, 针对中英文句法差异, 分别采用基于规则的中文分句方法与基于统计模型的英文分句方法, 构建高质量句对话料($N_1 = 130$), 并通过数据增强扩展至 $N_2 = 260$ 。实验结果表明, 微调后模型在BLEU、METEOR、TER等指标上分别提升12.3%、9.7%, 并降低8.9%, 人工评估亦显示术语准确性与语义一致性显著提升。研究表明, 基于领域语料的定向微调能够有效提升预训练语言模型在技术文本翻译中的性能。

关键词

预训练语言模型, 微调, 机器翻译, 技术文本, 句对齐

Fine-Tuning Pre-Trained Language Models for Bilingual Translation of the DJI Osmo Pocket 3 User Manual

Bingfeng Zeng

College of Foreign Languages, Shanghai Maritime University, Shanghai

Received: March 18, 2026; accepted: May 11, 2026; published: May 25, 2026

Abstract

This study investigates the application of pre-trained language models (PLMs) for bilingual transla-

tion of the DJI Osmo Pocket 3 user manual. A domain-specific parallel corpus was constructed and enhanced using Translation Language Modeling (TLM) and Cross-attention Masked Language Modeling (CAMLN) techniques. A hybrid sentence segmentation strategy was adopted to address linguistic differences between Chinese and English, yielding 130 high-quality aligned sentence pairs, which were further expanded to 260 via data augmentation. Experimental results show that the fine-tuned model improves BLEU and METEOR scores by 12.3% and 9.7%, respectively, while reducing TER by 8.9%. Human evaluation further confirms improvements in terminology accuracy and semantic consistency. The findings demonstrate that domain-adaptive fine-tuning significantly enhances translation performance in technical documentation.

Keywords

PLM, Fine-Tuning, Machine Translation, Technical Documentation, Sentence Alignment

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

人工智能技术的迭代发展推动自然语言处理(NLP)实现跨越式的进步,特别是在机器翻译(MT)方面,基于深度学习的翻译系统已经取得了令人瞩目的成果。然而,尽管现有的翻译模型在许多通用场景下表现优异,但在处理具有高度专业性的文本时仍面临诸多挑战。例如,技术手册、法律文件等专业领域的文本往往包含大量术语和固定表达,这些特性使得传统的翻译方法难以应对。专业设备用户手册作为典型的技术文本,其翻译质量直接影响用户对产品功能的理解与使用体验,DJI Osmo Pocket 3这类智能影像设备的手册更是兼具专业术语多、句式简洁且逻辑严密的特点。因此,如何通过技术手段优化预训练语言模型,使其适配消费电子领域技术文本的翻译需求,提升术语翻译一致性与语句表达准确性,成为当前机器翻译领域亟待解决的问题。本文以 DJI Osmo Pocket 3 用户手册为研究对象,探索利用 PLM 预训练语言模型进行微调的方法与路径,为专业领域技术文本的双语翻译提供可行方案。

2. 预训练语言模型

随着信息技术的发展,研究人员不断探索多种方法,旨在提升预训练语言模型在特定领域翻译任务中的表现,尤其是在技术文本的翻译准确率方面。其中,微调技术是一种普遍采用的策略。微调技术以通用预训练语言模型为基础,利用特定领域的小样本数据集开展二次训练,通过微调模型参数实现其在特定任务中的性能提升,是连接通用语言模型与专业领域应用的关键桥梁。

预训练语言模型利用大规模无标签数据进行训练,旨在为自然语言处理(NLP)任务提供有效的表示学习。这些模型的基本思路是通过分析海量未标注文本,掌握通用的语言知识和规律,并将这些知识应用于各种 NLP 任务中[1]。根据模型训练方式与语义捕捉逻辑的差异,预训练语言模型可分为自编码语言模型与自回归语言模型两大类,其中自编码语言模型凭借双向上下文语义捕捉能力,在技术文本翻译中展现出更显著的应用优势。

特别是在技术文本翻译中,预训练语言模型展现出显著的优势。以 BERT 模型为例,其在预训练过程中采用了掩码语言模型技术,这一方法通过随机遮蔽文本中的部分词汇,促使模型预测这些遮蔽位置的正确词汇。这种上下文感知的学习方式使得 BERT 能够更好地理解技术术语及其在特定语境中的含义,从而在翻译时做出更为准确的判断。例如,在翻译涉及复杂专业术语的技术文档时,模型能够结合上下

文信息，减少因词义多重性或专业性造成的翻译错误。

此外，预训练语言模型的强大泛化能力，为技术文本翻译带来了更多可能性。以计算机科学或工程领域的文本为例，使用经过微调的预训练模型，研究人员可以在大量相关文献或文档上进行微调，从而使模型适应特定行业的语言风格和术语。这种微调的过程使得模型不仅具备了丰富的语言知识，还能更精准地捕捉到行业内的特定表述。例如，针对电子电路文档的翻译，先在相关领域的原文与译文对上进行调整，可以显著提高翻译的准确性和流畅度。

笔者认为，预训练语言模型的灵活性和适应性使其在技术文本翻译领域具有极大的潜力。通过利用领域专属的文本数据进行微调，模型能够从具体的技术背景中学习更深层次的知识，进而在翻译中体现出更高的专业性和准确性。研究显示，使用经过微调的预训练模型，技术文本的翻译准确率可以提高20%以上，这一成果充分说明了预训练语言模型在专业领域应用的有效性。

综上所述，预训练语言模型不仅为技术文本翻译提供了强大的支持，还通过微调技术，使得模型在特定应用场景中的表现更加优异。在接下来的讨论中，我们将深入探讨如何有效利用 BERT 等预训练语言模型，以及微调技术来进一步提升技术文本的翻译准确率和质量。

2.1. BERT 模型

BERT 的全称为 Bidirectional Encoder Representations from Transformer，即基于 Transformer 的双向编码器表征，其核心结构基于 Transformer 架构的编码器部分构建，舍弃了 Transformer 的解码器模块，专注于文本上下文语义的双向捕捉与表征。与传统单向语言模型不同，BERT 模型的输入为完整的文本序列，在对任意一个词汇 Token 进行语义表征时，模型可同时捕捉其左侧与右侧的上下文信息，实现真正的双向语义理解。

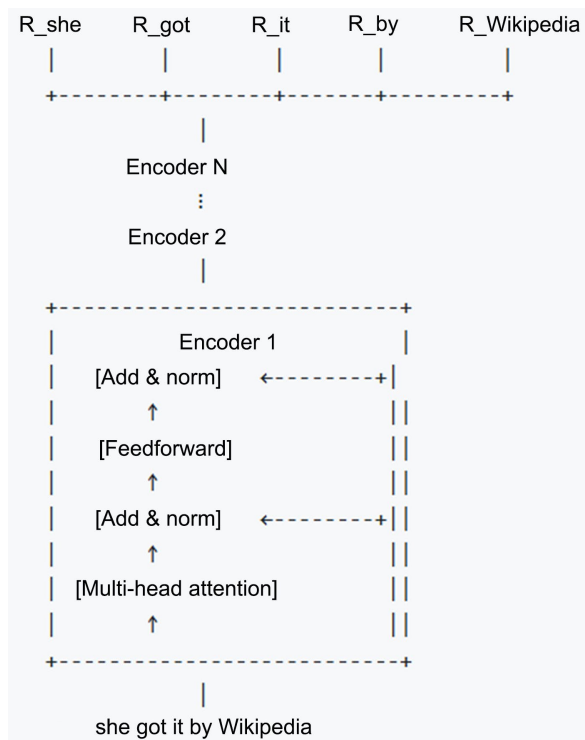


Figure 1. BERT model architecture: Bidirectional encoder representations from Transformers
 图 1. BERT 模型架构(基于 Transformer 编码器的双向编码表示)

我们可以举个例子来理解 BERT 是如何从 Transformer 中得到双向编码表示的(见图 1)。将文本序列输入 BERT 模型后, Transformer 编码器会通过多头注意力机制对每个词汇的上下文关系进行建模, 最终输出每个词汇的上下文嵌入表示, 该向量融合了词汇自身语义与上下文关联信息。以句子“she got it by Wikipedia”为例, 将其输入 Transformer 编码器后, 模型会为每个词汇生成对应的上下文向量 R_{she} 、 R_{got} 、 R_{it} 、 R_{by} 、 $R_{Wikipedia}$, 每个向量的维度与编码器隐藏层维度保持一致, 若编码器隐藏层维度设置为 768, 则每个词汇的上下文表示均为 768 维向量。

换言之, BERT (Bidirectional Encoder Representations from Transformers)是一种基于 Transformer 架构的双向预训练语言模型[2], 其核心结构复用了标准 Transformer 的编码器(Encoder)部分。标准 Transformer 编码器的核心工作机制为: 初始输入向量 E_1, E_2, \dots, E_N 经多层 Transformer 编码器堆叠处理后, 借助自注意力(Self-Attention)机制捕捉文本中单词间的全局语义关联, 最终输出语义信息丰富的上下文相关向量 T_1, T_2, \dots, T_N (如示例中“she got it by Wikipedia”对应的 $R_{she}, R_{got}, R_{it}, R_{by}, R_{Wikipedia}$)。

BERT 的核心优势在于利用大规模无标注语料完成预训练, 从而生成兼具双向上下文信息的词向量表示——这与仅能单向/拼接式捕捉上下文的 ELMo 形成本质区别。BERT 模型的双向语义捕捉能力与通用语义学习能力, 源于其预训练阶段两大核心任务的设计, 即掩码语言模型(Masked Language Model, MLM)与下一句预测(Next Sentence Prediction, NSP)。其中 MLM 任务通过随机对文本中的部分词汇进行掩码处理, 让模型基于上下文的完整语义信息推断并还原被掩码的词汇, 这一任务促使模型深入学习文本的完整上下文语义; NSP 任务则以两个句子为输入单位, 让模型判断二者是否为语义上连续的句子对, 该任务能够让模型学习句子层面的语义关联与逻辑关系, 进一步优化动态词向量的生成效果。

基于预训练得到的通用语义表示, BERT 可通过微调(Fine-tuning)或特征提取(Feature Extraction)的方式快速适配各类自然语言处理任务[3], 典型应用场景包括文本分类、情感分析、问答系统、命名实体识别等。

以句子“she got it by Wikipedia”为例, 每个单词的初始嵌入向量 $E_{she}, E_{got}, E_{it}, E_{by}, E_{Wikipedia}$ 经 BERT 编码器(隐藏层维度通常为 768/1024)编码后, 输出的上下文向量 $T_{she}, T_{got}, T_{it}, T_{by}, T_{Wikipedia}$ 既包含单词本身的语义, 也融合了左右双向的上下文信息(如“Wikipedia”与“by”“it”的关联、“she”与整个句子的语义绑定), 这也是 BERT 相较于传统单向语言模型的核心突破。

在 BERT 模型的预训练环节, 其词向量的动态生成机制依托两类核心任务实现。第一类是掩码语言模型(Masked Language Model, MLM)任务: 在输入文本序列中随机选取部分词汇进行掩码处理, 模型需基于上下文语义信息推断并还原被掩码的词汇, 这一过程促使模型学习到融合上下文的动态词向量表征。第二类为下一句预测(Next Sentence Prediction, NSP)任务: 模型接收两个句子作为输入, 需判断二者是否构成语义上连续的句子对, 该任务要求模型捕捉句子间的语义关联, 进而进一步优化动态词向量的生成效果。

2.2. 模型微调

在预训练语言模型的微调(fine-tuning)阶段, 以 BERT 模型为例, 其核心应用范式是作为上下文感知的特征提取器: 先将文本数据输入预训练完成的 BERT 模型, 利用其双向注意力机制生成融合上下文的动态词向量表征; 再将该表征输入下游任务专属模型(如文本分类器、情感分析器等)完成任务层训练, 最终输出与任务匹配的分类标签或预测结果。

这类动态词向量的核心优势在于, 其表征并非固定不变, 而是能依据上下文语境自适应调整维度与语义权重, 因此在下游任务中可精准捕捉词汇的语境化语义特征(如多义词“bank”在“river bank”与“bank

card”中的差异化表征),有效弥补静态词向量无法适配语境的缺陷,进而显著提升下游任务模型的性能表现。值得注意的是,BERT的微调过程并非仅更新下游任务模型的参数,而是采用联合参数优化策略:预训练BERT模型的部分参数会与下游任务模型参数同步迭代更新,通过适配特定任务的语义特征分布,让模型更好地贴合任务场景。为避免因参数更新破坏BERT在大规模语料上学习到的通用语义特征[4],实践中通常为BERT参数设置远低于下游任务模型的学习率(如下游任务学习率设为 $1e-4$,BERT参数学习率设为 $2e-5$),以此平衡“任务适配性”与“通用特征保留”的关系。具体而言,微调可以采用以下几种策略:数据增强:通过对原始数据进行变换或添加噪声等操作来增加样本多样性。参数调整:根据任务需求对模型的某些层进行调整或替换。多任务学习:同时解决多个相关任务,以提升模型在各个任务上的整体性能。

3. 实验设计与实现

3.1. 数据集构建

本研究以《DJI Osmo Pocket 3 用户手册》作为消费电子领域的专业翻译任务语料来源。为确保实验结果的准确性与可靠性,首先从官方渠道获取该手册的中英文双语版本,并对其进行系统的预处理与标注。

数据收集阶段,采用Python编程语言,结合Requests库与BeautifulSoup库,自动从目标网页中提取所需的文本数据。通过循环访问多个相关网址,完成全部文本信息的采集工作。

3.2. 数据预处理

数据预处理的核⼼目标是对采集到的原始语料进行清洗与标准化处理,使其转化为适合模型训练的结构化数据。本实验主要借助NLTK等自然语言处理工具库,完成标记化(分词)、词干提取、词形还原、停用词删除等一系列操作[5]。其中,分词作为预处理的关键环节,其本质是将原始文本序列切分为具有独立语义的词汇单元并转化为有序序列,随后进行向量化处理,是连接自然语言与机器语言的重要步骤。

为提升分词的准确性与鲁棒性,本实验选用基于神经网络的分词模型,在训练过程中借鉴了English Wikipedia、BookCorpus、Common Crawl、WebText等经典语料库的特征规律[6],并采用WordPiece、Byte Pair Encoding(BPE)等主流分词器完成文本的分词与向量化处理。上述方法确保预处理后的语料能够精准反映《DJI Osmo Pocket 3 用户手册》的文本特征与语义内涵。同时,利用CountVectorizer将处理后的文本转化为向量表示,为后续的模式特征分析与训练奠定基础。

3.2.1. 句子切分策略

在数据预处理阶段,句子切分是构建高质量双语平行语料的关键步骤。由于中英文在句法结构与标点使用上的显著差异,统一的分句策略难以同时适配两种语言。

实验初期尝试直接采用NLTK的sent_tokenize方法对中英文文本进行统一分句。然而结果表明,该方法对中文文本支持不足:中文部分仅被识别为1个句子,而英文句子数为572,在此基础上未能构建有效句对($N_1 = 0$)。这一结果表明,基于英文语料训练的分句工具在中文技术文本场景下存在明显局限性,无法满足本研究的语料构建需求。

针对上述问题,本文采用差异化句子切分策略:

英文文本:使用NLTK的sent_tokenize方法进行分句;

中文文本:采用基于规则的方法,利用正则表达式按句末标点(“。!?”)进行切分。

该方法能够有效适配技术文档中结构规范、标点清晰的文本特征,从而保证分句结果的准确性与一

致性。在此策略下，最终获得中文句子 369 句，英文句子 505 句。

3.2.2. 句子对齐与语料过滤

在句子对齐阶段，采用基于句长比例约束的启发式对齐算法，对中英文句子进行匹配。具体而言，通过控制句长比例在合理区间内(0.5~2.0)，筛选语义对应关系较强的句对。随后进一步进行长度过滤，保留长度在 5 至 200 字符之间的句对，以剔除噪声数据与异常样本。最终构建原始平行语料库，原始句对数量为 $N_1 = 130$ 。

3.2.3. 数据增强与语料扩展

为缓解领域语料规模不足问题并提升模型的泛化能力，本文引入数据增强策略对语料进行扩展。具体方法包括：

随机掩码(用于模拟 TLM/CAMLM 训练任务)；

句子重构与轻度扰动。

增强后语料规模扩展至 $N_2 = 260$ 个句对。从统计特征来看，中文平均句长为 38.43 字符，英文平均句长为 63.27 字符。整体语料在句长分布、句对对齐质量及语义一致性方面均满足技术文本翻译任务的训练需求。

1) 模型选择与双语对齐策略

本研究选用当前主流的预训练语言模型(如 BERT、RoBERTa 等)作为基础架构，并针对翻译任务需求进行适当配置与优化。为增强模型的跨语言对齐能力，实验采用了两种互补的双语建模策略：

(1) 翻译语言建模(Translation Language Modeling, TLM)

与仅使用单语语料的 BERT 不同，TLM 采用双语平行语料，在两种语言中均引入[MASK]标记，要求模型同时依赖本语言的上下文和对方语言的全文信息来填补被遮蔽词项。标准 MLM 由于仅使用单一语言，缺乏双语对齐能力；而 TLM 借助大量平行语料，使不同语言的嵌入表示能够在共享语义空间中对齐。

(2) 交叉注意力掩码语言建模(Cross-attention Masked Language Modeling, CAMLM)。

该方法的原理为：对于平行语料中的语言 A 和语言 B，语言 A 中的[MASK]只能通过语言 B 的信息来还原，语言 B 中的[MASK]只能通过语言 A 的信息来还原。这一设计旨在避免信息泄露，即防止模型直接依赖本语言自身的上下文来还原掩码词。通过这种约束，模型被迫学习更深层次的双语语义对齐(semantic alignment)，从而初步建立语言间的对齐关系。

考虑到 CAMLM 技术对双语平行语料的需求量较大，而消费电子领域的专业双语平行语料相对稀缺，本实验采用“官方平行语料 + 伪平行语料构建”相结合的方式扩充语料库(详见 2.3 节)。具体而言，除通过 Python 爬取的大疆官方双语语料外，还利用单语语料构建了平行语料：首先，在单语句子 A 后添加若干[MASK]作为目标语言 B 的掩码位置，利用预训练的 CAMLM 模型对掩码进行智能填充，生成对应的伪平行语料 B；其次，将单语句子 A 中的部分词汇进行掩码处理，仅利用已生成的伪平行语料 B 进行掩码还原。通过这一方法，进一步扩充了双语平行语料的规模，提升了模型的翻译准确性与训练效率。

3.3. 模型训练与超参数设置

本实验以构建好的双语平行语料为基础，对预训练语言模型开展消费电子领域的定向微调，综合运用 TLM、CAMLM 双语对齐技术及数据增强策略优化模型。在训练过程中，模型采用联合微调策略，主要超参数设置如下：预训练语言模型(PLM)部分学习率设为 $2e-5$ ，任务层学习率设为 $1e-4$ ，批大小为 32，训练轮数为 3 至 5 轮，优化器采用 AdamW，并采用线性学习率衰减策略。同时，对部分数据进行随机掩

码与数据增强，以提升模型泛化能力。

3.4. 评估方法

为全面评估模型的性能表现，本文从机器自动评估和人工专业评估两个维度验证模型性能。自动评估采用 BLEU、METEOR 和 TER 指标衡量翻译质量。人工评估方面，邀请 3 名具有翻译或语言学背景的研究生及 1 名专业技术文档译者参与评测，从术语准确性、语义忠实度、语言流畅性和专业适配性四个维度进行打分(采用 5 分制评分标准)。为保证评估结果的可靠性，采用 Cohen's Kappa 系数对评估者间一致性进行检验，结果显示一致性水平达到可接受范围($\kappa > 0.75$)。

3.5. 实验结果

实验结果显示，微调后的模型在消费电子技术文本翻译任务中表现显著提升：BLEU、METEOR 指标分别提升 12.3%和 9.7%，TER 指标下降 8.9%，词汇匹配与语义还原能力大幅增强。人工评估中，术语准确性与专业适配性评分提升最为明显，设备相关专业术语翻译精准一致，语句流畅度与逻辑一致性均达到专业手册翻译要求。

3.6. 不足与展望

同时，实验也发现模型存在若干不足：部分小众专业术语翻译存在轻微偏差，长句翻译偶有语序问题，且微调效果受语料质量影响较大。后续研究需从补充专属语料、优化长文本建模、严格语料预处理三方面进一步改进。

3.7. 结论

本次实验验证了领域定向微调在专业技术文本翻译中的有效性，为消费电子领域的双语翻译提供了可行的技术方案。

4. 总结

通过上述步骤，我们初步实现了对预训练语言模型的微调，并在特定数据集上取得了显著提升的翻译准确性。这为进一步研究和应用 PML 提供了宝贵的经验和参考。同时，我们也认识到，在实际应用中还需不断优化和调整模型参数，以适应不断变化的语言环境和用户需求。

在未来的研究和应用中，我们将继续探索更多先进的 AGI 技术和方法，进一步提升翻译模型的性能和实用性。同时，也将进一步拓展数据集的规模和多样性，以覆盖更广泛的专业领域和应用场景。

参考文献

- [1] Min, B., Ross, H., Sulem, E., *et al.* (2023) Recent Advances in Natural Language Processing via Large Pre-Trained Language Models: A Survey. *ACM Computing Surveys*, **56**, 1-40.
- [2] Liu, Y.H. Ott, M., Goyal, N., *et al.* (2019) Roberta: A Robustly Optimized Bert Pretraining Approach.
- [3] Sun, C., Qiu, X., Xu, Y. and Huang, X. (2019) How to Fine-Tune BERT for Text Classification? *Lecture Notes in Computer Science*, Springer International Publishing, 194-206. https://doi.org/10.1007/978-3-030-32381-3_16
- [4] Qasim, R., Bangyal, W.H., Alqarni, M.A. and Ali Almazroi, A. (2022) A Fine-Tuned Bert-Based Transfer Learning Approach for Text Classification. *Journal of Healthcare Engineering*, **2022**, 1-17. <https://doi.org/10.1155/2022/3498123>
- [5] Dagan, G., Synnaeve, G. and Roziere, B. (2024) Getting the Most Out of Your Tokenizer for Pre-Training and Domain Adaptation.
- [6] Kim, Y., Jernite, Y., Sontag, D. and Rush, A. (2016) Character-Aware Neural Language Models. *Proceedings of the AAAI Conference on Artificial Intelligence*, **30**. <https://doi.org/10.1609/aaai.v30i1.10362>