

基于LSTM模型与情感词典的翻译情感一致性分析

刘怡茹

上海海事大学外国语学院，上海

收稿日期：2025年11月27日；录用日期：2026年1月4日；发布日期：2026年1月13日

摘要

随着大语言模型广泛应用于翻译领域，机器译文质量评估模式受到更多关注。目前的译文质量评估框架多聚焦于词汇准确性和语法正确性，却往往忽略了原文与译文之间的情感一致性。情感表达的偏差不仅影响信息传递的完整性，还可能引发跨文化沟通中的误解。针对这一问题，本文提出在译后编辑过程中引入LSTM（长短期记忆网络）模型与情感词典的双路径融合情感分析方式进行情感分析，为翻译质量评估增加新的评估维度从而优化机器译文的情感表达效果。实验结果表明，引入该框架后，提高了对译文中的情感弱化、极性偏差等问题的有效识别，突破了原有质量评估框架的局限性。

关键词

LSTM模型，情感分类，翻译质量评估，深度学习

Sentiment Consistency Analysis of Translations Based on LSTM and Sentiment Lexicon

Yiru Liu

College of Foreign Languages, Shanghai Maritime University, Shanghai

Received: November 27, 2025; accepted: January 4, 2026; published: January 13, 2026

Abstract

With the widespread adoption of large language models (LLMs) in the field of translation, increasing attention has been drawn to the evaluation paradigms of machine translation (MT) quality. Existing

MT quality assessment frameworks primarily focus on lexical accuracy and grammatical correctness, while the consistency of sentiment between the source text and its translation is often overlooked. Deviations in sentiment expression not only compromise the integrity of information transfer but may also lead to misunderstandings in cross-cultural communication. To address this issue, this study proposes incorporating a dual-path sentiment analysis approach—combining a Long Short-Term Memory (LSTM) network with a sentiment lexicon—into the post-editing process. This method introduces a new evaluation dimension for translation quality assessment, thereby enhancing the emotional fidelity of machine-generated translations. Experimental results demonstrate that the proposed framework significantly improves the detection of sentiment weakening and polarity shifts in translations, overcoming limitations inherent in existing evaluation models.

Keywords

LSTM Model, Sentiment Classification, Translation Quality Assessment, Deep Learning

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

生成式人工智能(Generative Artificial Intelligence, GenAI)的崛起正在重塑语言服务业。以 ChatGPT、Gemini、Deepseek 等为代表的预训练大模型在跨语言翻译等领域取得突破性进展，显著提升了翻译效率与自动化水平。在新闻、科技、商务等领域的跨语言信息传递中，大语言模型已成为重要的辅助甚至主导翻译工具。由于该类模型在性能方面呈现指数级发展，翻译工作流程也面临着重构。使人机协同迈入人智协同的新阶段。在这场技术驱动的变革中，翻译模式正从机器翻译译后编辑(Machine Translation Post-Editing, MTPE)转向人工智能译后编辑(Artificial Intelligence Post-Editing, AIPE) [1]。

然而，大语言模型(LLMs)仍存在较多短板。首先，LLMs 在译文生成方面仍聚焦于词汇准确性、语法规范性、句法流畅性等表层语言特征，却无法分析跨文化交际行为中不可或缺的情感维度[2]。原文中所蕴含的情感倾向、情感强度及情感色彩，是作者意图与文本感染力的核心组成部分，情感传递的忠实度直接影响信息的完整接收与读者的心理共鸣。且 LLMs 的“黑箱”特性导致其生成的文本存在不可预测性，不可控程度仍偏高。因此，本研究提出基于 LSTM 模型与双语情感词典对机器译文进行情感一致性分析，优化目前质量评估框架忽略情感分析的短板。

2. 相关研究

2.1. 机器学习

早期的情感分类会通过手工设计的方法进行，比如通过计算词频、制作情感词汇表等，并从文本数据中提取特征并进行分类。后来随着计算机的发展，利用机器学习进行情感分类逐渐占据了学科的主导地位。传统的机器学习常运用支持向量机、朴素贝叶斯、决策树等进行情感分类。

机器学习的发展为情感分类带来了更多维度的分析。“然而，基于机器学习的传统方法有许多不足之处，在处理长文本和复杂语境时面临巨大挑战。传统方法高度依赖人工设计，需要人工进行特征工程、模型选择和参数调整等。这些步骤不仅需要人工付出大量的时间和精力，而且结果可能受到人为因素的影响。传统方法通常无法捕捉到文本的语义信息，它们主要依赖于浅层的特征表示” [3]。

2.2. 深度学习

深度学习即可以从大量的数据中提取文本特征并进行分类。这一方法为当今最普遍的方法，节省了许多时间。深度学习中，常用卷积神经网络(Convolutional Neural Network, CNN) 和循环神经网络(Recurrent Neural Network, RNN)来进行情感分析。

2.2.1. CNN

CNN (Convolutional Neural Network)即卷积神经网络，常用于图像识别任务。CNN 的结构组成依次为输入层(Input)、卷积层(CONV，包括激活操作)、池化层(POOL)、全连接层(FC)和输出层(OUTPUT)。CNN 处理使用卷积操作来提取图像中的局部特征。池化层则用于降低特征图的维度，保留关键特征并减少参数数量和计算量。常见的 CNN 算法有 LetNet5、VGG16、AlexNet、GoogleNet。

2.2.2. RNN

循环神经网络是将文本数据以序列的形式进行输入，网络中的相关节点也是以链式连接的递归神经网络进行排列[4]。RNN 主要用于图像处理、音频识别、文本相似度计算、机器翻译等。相较于 CNN，其结构比较简单，依次为：Input layer、Hidden Layer、Output layer。本文将用到的 LSTM 属于 RNN 的一种。

2.2.3. Transformer

该模型摒弃了前两种结构，提出自注意力机制(Self-Attention)处理序列关系，通过多头注意力(Multi-Head Attention)和位置编码(Positional Encoding)实现了对序列的高效并行计算和全局依赖建模。该模型也是目前生成式大模型的核心架构[5]。

虽然 Transformer 架构发展迅速并在多项 NLP 任务中取得了领先性能，但其固有的“黑箱”特性仍然是无法根本解决的问题。Transformer 模型内部复杂的多头注意力机制与海量参数的非线性交互，导致模型的决策过程高度不透明且难以解释[6]。且大语言模型在处理微妙情感表达、文化特定情感色彩以及复杂情感变化时，容易出现情感弱化、偏移甚至误判的现象。

相比之下，长短期记忆网络(LSTM)虽然在并行计算和全局建模能力上不及 Transformer，但其在情感分析任务上仍具有稳定的输出。LSTM 是基于 RNN 优化的模型，通过引入门控机制(包括输入门、遗忘门和输出门)和细胞状态(cell state)，能够有效地捕捉长距离依赖关系，避免梯度消失问题[7]。LSTM 在文本情感分类任务中能够达到较高的准确率。

情感分析(Sentiment Analysis)，亦称意见挖掘(Opinion Mining)，是自然语言处理(NLP)的核心子领域之一，旨在通过计算，识别、提取、量化和研究文本中所表达的主观性信息、情感状态、观点、评价、态度和情绪。情感分析试图让机器理解超越客观事实的、带有个人色彩的态度、情绪和评价[8]。情感分析对智能系统的输出质量具有重要作用，任何追求自然、友好、个性化的智能系统(如智能客服、虚拟伴侣、教育助手)都必须具备理解和回应人类情感的能力。情感分析是实现情感计算和人性化人机交互的基础模块[9]。

3. 实验过程

本研究采用双路径融合分析框架，使用基于 LSTM 的深度学习方法与基于情感词典的规则方法，对大语言模型译文与原文的情感一致性进行多维度评估。

3.1. 数据集介绍

本研究使用的公开数据集选自以下地址：<http://www.cs.cornell.edu/people/pabo/movie-review-data/>，其中包含了观众对电影的正面和负面评价数据集，情感规模数据以及对作品的主观评分数据集。本研究选

取了该数据集中的 polarity dataset 2.0。该数据集中的正面和负面数据均为 1000 条。

3.2. 文本预处理

数据集中的数据已有处理好的标签，但还应对文本进行(一) 分词(包括降噪，即删除没有语义的符号数据)；(二) 建立字典；(三) 将文本从词序列转化为索引序列，即编程语言可识别的数据集。

其中建立词典步骤较为复杂。建立字典即把词汇对应成数字，储存在字典中。词典构建中可以使用特殊字符填充短句并代替未来的新词汇。“数字编码：将文本单词映射为数字编码，使其能够被模型处理。为此，创建了一个词表(Vocabulary)，将每个单词映射到唯一的整数。在代码中，这一过程通过构建词汇表 vocab 实现”[10]。建立词典的具体步骤包括句子分词；词语储存(包括词语过滤和次数统计)；实现文本和数字序列的相互转换。

通过 Python 处理数据样本长度和出现频率间的关系，得到以下可视化关系图(见图 1)。

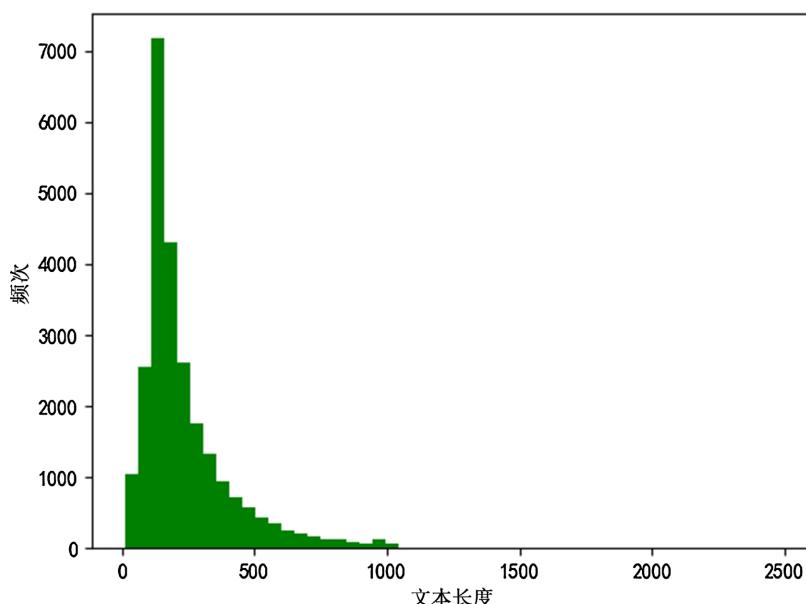


Figure 1. Sample length distribution diagram
图 1. 样本长度分布图

由上图可知，本数据集的绝大多数文本长度均在 1000 以下。

3.3. 基于 LSTM 情感分类

LSTM 单元包含记忆细胞、输入门、输出门和遗忘门，这些门机制帮助模型在训练过程中选择性地保存和丢弃信息。LSTM 模型通过一个或多个 LSTM 层来处理输入数据。在构建模型时，我们可以加入一个或多个 LSTM 层，并在其后添加全连接层，以输出最终的情感分类结果。

训练 LSTM 模型需要选择合适的超参数，如学习率、批次大小、训练轮数等。我们可以使用交叉验证来调整这些参数，以获得最佳的模型性能。在训练过程中，损失函数(如交叉熵损失)和优化器(如 Adam 或 RMSprop)被用来最小化训练误差。训练完模型后，我们需要使用测试集对模型进行评估，通过准确率、精确率、召回率和 F1 分数等指标来衡量模型的性能。

为了进一步提升模型的准确性，我们可以引入其他技术，如注意力机制、双向 LSTM 层，甚至结合卷积神经网络(CNN)以捕捉更复杂的特征。此外，通过增加数据集的多样性，或利用数据增强技术，可以

增强模型的泛化能力。

3.4. 模型准确率反思

3.4.1. 模型参数

```
update_w2v = True # 是否在训练中更新 w2v
vocab_size = 54848 # 词汇量, 与 word2id 中的词汇量一致
n_class = 2 # 分类数: 分别为 pos 和 neg
embedding_dim = 50 # 词向量维度
batch_size = 64 # 批处理尺寸
hidden_dim = 128 # 隐藏层节点数
n_epoch = 5 # 训练迭代周期, 即遍历整个训练样本的次数
lr = 0.0001 # 学习率; 若 opt='adadelta', 则不需要定义学习率
drop_keep_prob = 0.2 # dropout 层, 参数 keep 的比例
num_layers = 2 # LSTM 层数
bidirectional = True # 是否使用双向 LSTM
list1=[64.789,76.355,79.304,77.545,80.760]#lr=0.0001
list2=[70.812,79.179,80.263,81.595,83.123]#lr=0.0002
```

3.4.2. 实验结果(见图 2)

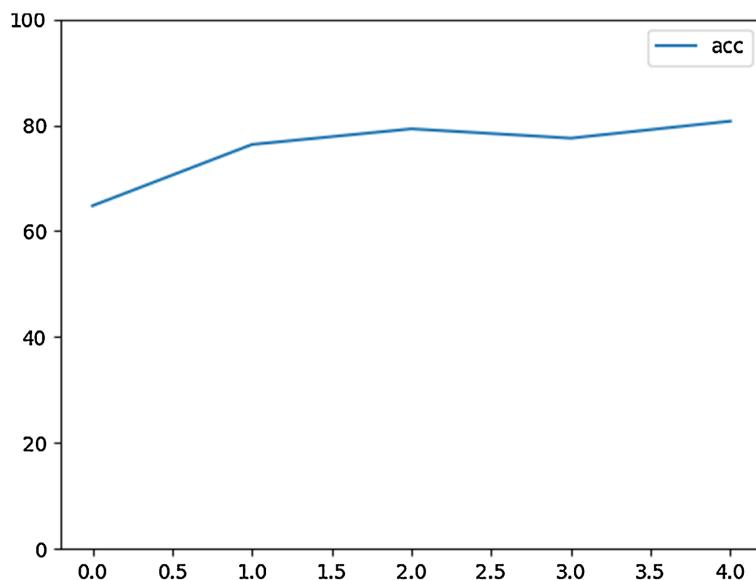


Figure 2. Model accuracy

图 2. 模型准确率

该图呈现模型准确率随训练轮数的增加而产生的变化, 其中横轴表示训练轮数, 纵轴表示准确率。每轮所得准确率为: [64.789, 76.355, 79.304, 77.545, 80.760]。当完成 4 轮训练后, 模型准确率已处于较高水平。

3.5. 情感词典的构建

为全面评估译文情感一致性, 本研究构建了三层级情感词典体系, 包括基础通用情感词典、领域专

用情感词典以及双语对齐情感词典。基础通用情感词典中的英文部分是基于 NLTK 的 VADER 词典，包含 7500+情感词，支持强度分级，中文部分融合知网 Hownet 词典和清华大学中文情感词典，覆盖 8000+情感词。领域专用情感词典包括从 IMDb、豆瓣等平台收集影评特色情感表达建立电影评价专用词库。双语对齐情感词典为人工创建的中英情感词对照表，确保跨语言情感分析的准确性。

完成对双路径分析框架的搭建后，本研究选取了一篇来自《北京周报》关于中国电影出海的评论型报道，并使用 ChatGPT-5 获取大模型译文。随后将原文与译文输入代码块并得出以下结果(见表 1)。

Table 1. Sentiment consistency analysis
表 1. 情感一致性分析表

情感分类		综合得分	情感极性一致性	情感强度一致性	情感一致性得分
原文	正面	1.457	1	0.076	0.626
译文	正面	0.533			

通过结果得出，虽然译文与原文保持了情感极性上的一致(都为正面)，但译文在情感强度上略低于原文。

4. 案例分析

4.1. 副词程度的弱化

程度副词的主要作用是表示某种事物的性状或动作行为的程度，以加深表达的语气[11]。汉语程度副词具有强烈的主观增强功能，尤其在新闻评论语体中，程度副词承担着情感强化和立场表达的双重功能。原文为表达两部影片取得的成功，多次使用如“极高(的水平)”、“令人瞩目”、“尤为”等程度副词。而译文对于这些程度副词的翻译仅保持准确性但忽略了情感强度，如“极高(的水平)”译为 especially high standards，“令人瞩目”译为 remarkable，“尤为(亮眼)”译为 eye-catching。这些词语虽然含有同样的赞美之意，却难以传达与原文相同的强度。因此，建议修改为(shine) with remarkable excellence, stunning, eye-popping。

4.2. 中英文句子结构差异

中文多流水句，且句与句之间无需逻辑词或结构连接，而英文则有固定的句子结构，需要显化逻辑关系。

原文“‘坚定文化自信、服务大众’的创作理念”被译为 the creative principle of “strengthening cultural confidence while serving the broader public”，用“while”承担逻辑词汇，使句子更偏事实陈述，弱化了原句情感强度。因此，建议修改为“Firmly uphold cultural confidence and serve the people.”。

4.3. 特定文化概念翻译

不同文化虽有相似的表达，但描述方式却不同。原文“酒香也怕巷子深”，译文采用直译的方式译为“Even the best wine fears hidden alleys”。这种对于原文文化的直译虽然保留了中文语境，但在传达含义与情感上会造成阻碍，因此建议采用保留情感强度的解释性翻译“No matter how great it is, it won’t sell itself if no one knows about it.”。

4.4. 隐喻性表达

大语言模型将原文“新蓝海”译为“a new blue ocean”。虽然“blue ocean”已在国际上作为一个战

略术语广泛使用，但并没有表达出原文所包含的“具有高潜力”的含义，因此建议修改为“ripe for growth”。

5. 结语

LSTM 模型因其在处理序列数据方面的优势被广泛应用于情感分析。本文在详细的实验过程中，首先通过数据预处理、特征提取等步骤为模型的训练做好准备。通过对数据集进行分词、建立词典以及将文本转化为索引序列等处理，确保数据能够被 LSTM 模型有效利用。并在此基础上引入基于情感词典的计算方式，构建针对情感一致性的双路径融合框架，对原文与译文的情感一致性进行分析，为翻译质量评估增加了新的维度，使译文能更准确传达原文情感，提高跨语言交流的质量。

参考文献

- [1] 王华树, 刘世界. 从 MTPE 到 AIPE: GenAI 时代翻译模式演变及其对翻译教育的启示[J]. 山东外语教学, 2025, 46(3): 111-121.
- [2] Bawden, R., Sennrich, R., Birch, A. and Haddow, B. (2018) Evaluating Discourse Phenomena in Neural Machine Translation. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1 (Long Papers), New Orleans, 1-6 June 2018, 1304-1313. <https://doi.org/10.18653/v1/n18-1118>
- [3] 薛嘉豪, 黄海, 孙宜琴. 基于 BiGRU-Attention 的中文微博评论情感分析[J]. 软件工程, 2024, 27(7): 12-16.
- [4] 蔡汶兴. 基于深度学习对中文影评的情感分析研究[D]: [硕士学位论文]. 兰州: 兰州交通大学, 2022.
- [5] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017) Attention Is All You Need. *Advances in Neural Information Processing Systems*, **30**, 3-6.
- [6] Bau, A., Belinkov, Y., Sajjad, H., et al. (2018) Identifying and Controlling Important Neurons in Neural Machine Translation. arXiv:1811.01157.
- [7] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [8] Pang, B. and Lee, L. (2008) Opinion Mining and Sentiment Analysis. *Foundations and Trends® in Information Retrieval*, **2**, 1-135. <https://doi.org/10.1561/1500000011>
- [9] Cambria, E., Das, D., Bandyopadhyay, S. and Feraco, A. (2017) Affective Computing and Sentiment Analysis. In: Cambria, E., Das, D., Bandyopadhyay, S. and Feraco, A., Eds., *Socio-Affective Computing*, Springer International Publishing, 1-10. https://doi.org/10.1007/978-3-319-55394-8_1
- [10] 郑志建, 林土水, 杨蕊平, 阙文华, 姚毅. 基于 LSTM 模型的用户情感分析方法研究[J]. 计算机与网络, 2024, 50(4): 366-370.
- [11] 王珊. 现代汉语程度副词修饰名词特征分析及理据解释[J]. 汉字文化, 2024(15): 13-15.