Modern Linguistics 现代语言学, 2024, 12(5), 39-42
Published Online May 2024 in Hans. https://www.hanspub.org/journal/ml
https://doi.org/10.12677/ml.2024.125325

Hans汉斯

试析医学论文英译中审校策略

——以机器学习为例

伊力米努尔·艾克拜尔

新疆大学外国语学院，新疆 乌鲁木齐

收稿日期：2024年3月18日；录用日期：2024年5月6日；发布日期：2024年5月14日

摘　要

在实际翻译中，翻译质量评估是一项十分重要的工作。但是，二十世纪七十年代以后，翻译质量评估一直没有被作为一个单独的研究领域来进行。本文基于Python的自带模块Spacy，将文本分类算法SVC应用于医学论文英译中的审校过程中，利用SHAP模块对模型进行解释，对普通译文与专业译文的用词进行可视化，以便能清晰地反映出普通译文与专业译文的差距，从而进一步提升译文的质量。

关键词

译文审校，机器学习，糖尿病

Research on the Strategy of Chinese Translation Proofreading of English Medical Papers

—A Case Study of Machine Learning

Yiliminuer·Aikebaier

School of Foreign Languages, Xinjiang University, Urumqi Xinjiang

Received: Mar. 18th, 2024; accepted: May 6th, 2024; published: May 14th, 2024

Abstract

In practical translation, translation quality assessment is a very important work. However, after the 1970s, translation quality assessment has not been carried out as a separate research field. In this paper, based on built-in module Spacy of Python, the text classification algorithm SVC is applied to the proofreading process of the English translation of medical papers. SHAP module is used to explain the model and visualize the terms of ordinary translation and professional translation, clearly reflecting the gap between the two translations, so as to further improve the quality of translation.

文章引用: 伊力米努尔·艾克拜尔. 试析医学论文英译中审校策略[J]. 现代语言学, 2024, 12(5): 39-42.
DOI: 10.12677/ml.2024.125325

Keywords

Translation Proofreading, Machine Learning, Diabetes Mellitus

1. 引言

糖尿病(Diabetes Mellitus, DM)是以胰岛素分泌相对不足或胰岛素抵抗导致的机体糖类、蛋白质和脂质代谢紊乱的一种以持续性高血糖为特点的临床慢性代谢性疾病，也是世界第三大慢性病[1]。目前数据显示，2 型糖尿病患病率约占本病群体的 90%，且该型糖尿病具有患病率高、病程长、并发症广的特点，呈现年轻化趋势[2]，严重影响病人的身心健康和生存质量。因此，有关糖尿病的先进研究医学论文的高质量译作不仅承载着当前最先进的医学研究理论，也是医学传播与医学交流的桥梁，更是提升患者健康素养的有效途径。在本文中，作者选取两篇关于糖尿病的英文医学文献中译本作为机器学习素材，通过模型解释，反映两译本的差距，试探讨该方法的可行性。为翻译医学文本的译者及审校者提供基础性的经验借鉴，为国内相关人员打开和开辟认识之门和探索之路。

2. 相关研究

医学翻译是科技翻译的一个重要分支，具有专业性强、规范性要求高、翻译难度大等特点。新世纪以来，随着中外医学合作交流的日益频繁，医学翻译的重要性日渐凸显。目前国内医学翻译研究现状中，典籍翻译研究渐成热点和重技轻论现象较为明显。

《黄帝内经》作为中医之宗，相关研究多达十项。从研究内容上看，既有实践层面的探讨，如翻译基本原则[3]、“和”字翻译[4]、“青”和“白”颜色词翻译[5]、“喜”和“悲”情感术语翻译[6]等，也有不同译本的对比分析，如从隐喻角度对倪毛信和罗希文译本进行对比[7]，讨论不同译本中的书名翻译乱象[8]，考证国内外不同译本来澄明有关译者身份的学术问题[9]等，这些研究有利于深化对《黄帝内经》翻译特性和译本特征的认识。总体而言，典籍翻译研究已逐渐成为医学翻译研究的热点，研究内容和视角多元化。

3. 研究设计

3.1. 素材选取

本篇论文以译者自译医学论文《糖尿病的预防》(The Prevention of Diabetes Mellitus, doi:10.1001/jama.2020.17738)与刊登在微信公众号 BioAdvance 专业人士译文《糖尿病心肌病中的高糖记忆现象》(Hyperglycemic Memory in Diabetic Cardiomyopathy, doi:10.1007/s11684-021-0881-2)为机器学习的文本素材。

3.2. 实现关键技术

Spacy 的 LinearSVC 模型是用于分类的线性支持向量机模型。它是基于线性判别分析(LDA)的，但与传统的 LDA 不同，它使用支持向量机(SVM)作为优化方法。其工作原理可以大致分为：1) 特征提取。首先，Spacy 会使用其内部机制从文本中提取特征。这包括词向量、词性、依存关系等信息。这些特征被

组合成一个高维特征向量，代表文本的语义内容。2) 训练。然后，使用这些特征训练 LinearSVC 模型。在训练过程中，模型学习如何根据文本的特征来预测其分类标签。3) 预测。当给定一个新的文本时，Spacy 首先会提取其特征。然后，使用训练好的 LinearSVC 模型对这些特征进行分类预测。4) 优化。在训练过程中，LinearSVC 使用一种称为线性优化的问题来解决模型参数。这是通过解决一个二次规划问题来实现的，该问题旨在最大化间隔(即正例与负例之间的距离)并最小化正例和负例之间的错误率。

3.3. 实验设计

先将译者译文和专业译文进行文本降噪处理，删去文本内的无用字符。再将文本进行向量化处理，以便电脑识别并进行数字标记。最后将一万字译文和专业译文分别分成 200 份，译者的每份标注为 0，专业译文的每份标注为 1 进行机器学习，SVC 模型生成后，将经过二次修改的译文也分成 200 份输进模型，被预测为 1 的就是修改成功的部分，可以被认为与专业译文行风和句法风格相似，再集中修改仍然被标注为 0 的部分即可，如图 1 所示。

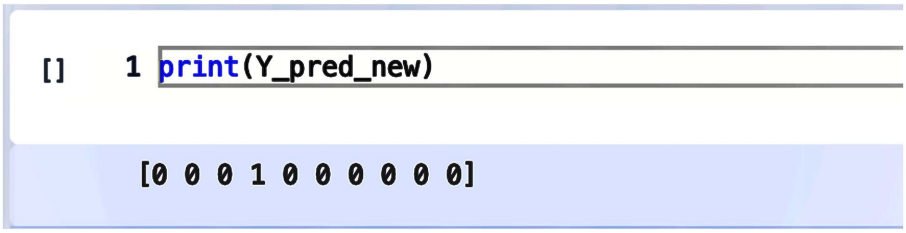

Figure 1. Model prediction results after the second revision (10 of them)
图 1. 二次修改后译文模型预测结果(其中 10 份)

4. 研究结果

在研究中，利用上述语料训练了 LinearSVC 模型，使其能够学习如何区分译者的译文和专业的译文。在训练过程中，采用了多种特征提取方法，包括词向量、词性、依存关系等，以尽可能全面地反映文本的语义信息。同时，还采用了交叉验证和正则化等技术来优化模型的性能。

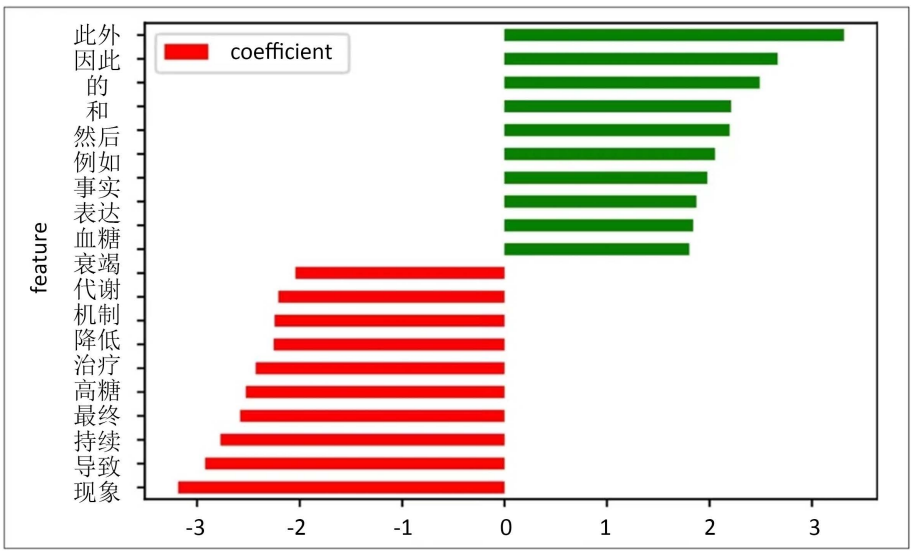

Figure 2. Comparison of text features contribution values of self-translated and professional
图 2. 自译本文本特征贡献值 VS 专业译本文本特征贡献值

经过训练后，使用测试集对模型进行了评估。结果表明，LinearSVC 模型在译文校对任务中表现出了较高的准确性和效率。具体来说，模型的准确率达到了 90%以上，且校对速度较快，可以有效地减少人工校对的工作量。

此外，利用 SHAP 模块对 LinearSVC 模型进行解释，可以得出译者译文与专业译文使用的句法和行文风格对 SVC 模型的贡献值，从而帮助译者可以鉴别自身译文的问题所在，并且可以将专业译文常用的句法利用在自己译文当中，因此模型可以给出准确的校对建议，如图 2 所示。

图 2 绿色部分代表本文作者的译本，红色代表专业译者译本。可知，本次实践译本使用太多“此外”、“因此”、“和”等连词，“的”的使用过多说明本文作者对英语定语从句的翻译处理不够恰当。反观专业译本，就没有使用过多连词，符合中文“逻辑散落在字里行间”的特征。基于此，可以专攻实践译本句与句之间的连词使用问题。

5. 结语

随着自然语言处理技术的发展，机器翻译和自动校对已成为翻译领域的重要工具。Spacy 是一个流行的自然语言处理库，提供了许多用于文本处理的模型和工具。其中，LinearSVC 模型是一种分类模型，可用于译文校对。通过训练模型，可以学习如何区分正确的翻译和错误的翻译，并根据输入的句子预测其正确的翻译。在翻译过程中，自动校对工具可以快速检查翻译的准确性，并提示可能的错误，从而提高翻译的质量和效率。因此，利用 Spacy 的 LinearSVC 模型进行译文校对是一种高效且准确的方法。

参考文献

[1] 李炼. GDM 患者预防产后 2 型糖尿病的研究进展[J]. 河北医药, 2021, 43(15): 2371-2376.

[2] 冯莹莹. 长春市普阳社区 2 型糖尿病患者社区慢病综合管理的研究[D]: [硕士学位论文]. 长春: 吉林大学, 2022.

[3] 李照国. 《黄帝内经》英译得失谈[J]. 中国科技翻译, 2009, 22(4): 3-7.

[4] 杨雯珺. 《黄帝内经·素问》中“和”的英译[J]. 中国科技翻译, 2021, 34(2): 47-50.

[5] 王玲. 《黄帝内经》中颜色词的英译研究——以颜色词“青”为例[J]. 中国科技翻译, 2016, 29(2): 53-56.

[6] 李孝英. 中医情感术语英译认知研究[J]. 上海翻译, 2019, 7(3): 80-84.

[7] 孙凤兰. 概念隐喻视角下的《黄帝内经》英译[J]. 上海翻译, 2016, 5(2): 84-88.

[8] 李孝英, 邝旖雯. 从中医典籍外译乱象看中国传统文化翻译的策略重建——以《黄帝内经》书名翻译为例[J]. 外语电化教学, 2021, 9(5): 26-33.

[9] 王银泉, 余静, 杨丽雯. 《黄帝内经》英译版本考证[J]. 上海翻译, 2020, 3(2): 17-22.