IEEE 版权声明 版权所有 (c)2023 IEEE。允许个人使用本材料。如需以任何其他方式使用本材料，无论是在当前还是未来的任何媒体中，包括为广告或宣传目的重印或再版本材料、创作新的汇编作品、转售或重新分发至服务器或列表，或在其他作品中重复使用本作品的任何受版权保护部分，均须事先获得 IEEE 的许可。

已获准发表于：2023 年 IEEE 第 1 届智能城市先进创新国际会议（ICAISC 2023），沙特阿拉伯吉达，2023年1月23日至25日。

arXiv:2303.01197v1 [cs.CL] 2 Mar 2023

通过作者分类实现文档来源追溯和身份验证

Muhammad Tayyab Zamir[1]，Muhammad Asif Ayub[2],Jebran Khan、Muhammad Jawad Ikram、Nasir Ahmad、Kashif Ahmad[l] Abasyn大学伊斯兰堡校区

[2] 工程与技术大学计算机系统工程系，巴基斯坦白沙瓦，
asifayub836@gmail.com，n.ahmad@uetpeshawar.edu.pk

[3] 吉达国际学院计算机科学与信息技术系，沙特阿拉伯，m.jawad@jicollege.edu.sa

[4] 韩国亚洲大学人工智能系，
jebran@ajou.ac.kr

[5] 爱尔兰科克芒斯特理工大学计算机科学系，kashif.ahmad@mtu.ie

风格分析是一个相对较少被探索的主题，却能够支持多种有趣的应用。例如，它使作者能够调整其写作风格，以便在协作中生成更连贯的文档。同样，风格分析也可作为文档来源追溯和身份验证的初始步骤。本文提出了一种基于集成的文本处理框架，用于单作者与多作者文档的分类，而这正是风格分析中的关键任务之一。所提出的框架结合了多种最先进的文本分类算法，包括经典机器学习（机器学习）算法、Transformer模型和深度学习算法，并分别单独使用及采用基于优选的后期融合。对于基于优选的后期融合，我们采用了多种权重优化和选择方法，为各个文本分类算法分配基于优选的权重。我们还通过在干净数据和未清洗数据上开展实验，分析了通常在自然语言处理应用的预处理过程中被排除的字符对该任务的影响。该框架在一个大规模基准数据集上进行了评估，显著提升了相较于现有解决方案的性能。

索引词—**文档身份验证；来源；BERT；RoBERTa；文本分类；后期融合；**

一、引言

近年来，在线社交网络上广泛传播的信息的可靠性日益受到业界关注。在此方面，人们尤为重视社交网络所分享信息的来源与真实性。对于书面文档而言，此类来源审查的一个重要方面与作者身份有关，即通过识别文档的原始作者来评估信息的真实性[1]。因此，可以认为，开发计算型作者身份识别系统意义重大；这类系统能够在新闻业、执法、内容审核等不同应用领域协助人们完成各项任务。风格分析是开发此类计算型作者身份识别系统的一项主要任务。风格分析主要通过利用合作创作文档中的内在特征，识别写作风格中的不同模式。文体变化检测（SCD）是风格分析的关键应用之一，与内在抄袭检测（IPD）相关。IPD 利用作者在文档篇幅范围内写作风格的一致性来检测抄袭，而不依赖任何其他参考语料库 [2], [3]。

关于多作者文档中的风格变化检测，文献中已有多项研究。多年来，风格变化检测也一直是 PAN[1],的一部分；PAN 是一项关于数字文本取证和文体计量学的基准共享任务。PAN 2021 共享任务[4] 中的风格变化检测包含三个子任务，即 (i) 单一作者与多作者、(ii) 基础风格变化和 (iii) 真实世界中的风格变化。在本研究中，我们针对第一个任务（即单一作者与多作者），该任务涉及对单一作者文档和多作者文档进行分类。文献中报道了多种有趣的解决方案，但该任务的若干方面仍未得到探索。例如，大多数解决方案依赖单独的机器学习和自然语言处理算法。然而，根据我们以往在类似自然语言处理应用方面的经验 [5],，我们认为，如果通过基于优点的融合将这些算法联合使用，它们可以相互补充。同样，预处理步骤对该任务中算法性能的影响尚未得到探索。预处理通常包括通过移除不必要的字符（例如停用词、问号、缩写形式等）来清理数据。我们认为，预处理步骤中通常被移除的文本可能会有用

[l]https://pan.webis.de/index.html

在风格变化检测中。例如，与其他作者相比，某位作者可能更频繁地使用停用词、问号、缩写形式等，这些特征可能有助于风格变化检测。

为克服这些局限性，我们提出一种基于优选的后期融合框架，单独及联合使用多种最先进的机器学习和自然语言处理算法，并采用多种权重优化和选择方法，对各个算法进行基于优选的融合。此外，我们还通过在干净数据集和未清洗数据集上进行实验，分析自然语言处理应用预处理过程中通常会被移除的字符对风格变化检测的影响。本文的贡献总结如下。

- 我们提出一种新颖的文档来源追溯与身份验证框架，采用多种基于优选的后期融合方法，能够区分由单个作者和多个作者撰写的文档。
- 我们评估了多种机器学习和自然语言处理算法，包括经典算法和最先进的 Transformer 模型，既单独评估，也在后期融合中联合评估。
- 我们开展了多项实验，以探索该任务的不同方面。实验在清洗后的数据和未清洗数据上进行，以分析自然语言处理应用在预处理过程中通常排除的字符对文体检测的影响。

II. 相关研究

文体计量学涉及对作者风格的分析，它将写作风格视为一种可量化的指标，以便量化给定文本中的独特特征。文体计量学包含多项任务，旨在分析文本的不同方面，例如作者归属判定、验证、作者画像、文体年代学和对抗性文体计量学。作者归属判定是文体计量学中一项关键且广受研究的任务 [6], [7]。作者归属判定旨在识别文档的作者，通常依赖于 SCD 技术。

文献中针对 SCD 报告了多种有趣的解决方案，从不同角度探索文本分析 [8]。例如，在较早的一项研究中，Glover等人 [9]运用文体计量特征，识别由多位作者撰写的文档中的作者边界。另一方面，Akiva等人 [10],提出了一种聚类框架，用于将多作者文档划分为由各个作者分别撰写的文本线索。连续三年，SCD 也作为一项共享任务纳入名为 PAN风格检测的基准项目中 [1], [4]。近期大多数相关研究都基于这项共享任务。该共享任务包含多个子任务，其中包括：(i) 区分单作者与多作者文档，并识别多作者文档中风格发生变化的段落序列；对于风格变化，考虑 (i) 基本变化和 (ii) 真实场景风格变化。本文针对第一项任务提出解决方案，并仅概述该任务的相关文献。

在相关文献中，已有研究提出了几种关于单作者与多作者文档分类的有趣解决方案。这些解决方案从不同角度探讨了该任务。例如，Zhang等人 [11] 提出了一种基于BERT和全连接神经网络的框架，用于区分单作者与多作者文档。Strom等人[12]提出了一种集成框架，将基于两种不同类型特征训练的不同分类器所获得的分类得分结合起来，这两类特征包括BERT嵌入和词汇特征。词汇特征包括字符、词语、句子、缩写词和功能词特征。Singh等人 [13]同样依赖词汇特征，使用了多样化的特征列表，包括n-gram的TF-IDF值、词性标签三元组和特殊字符、功能词频率，以及每个词的平均字符数。随后，在提取的特征上训练逻辑回归分类器，以对单作者与多作者文档进行分类。Deibel等人 [14] 使用了基于多层感知机（MLP）算法的模型，该模型在一组多样化的文本特征上进行训练，包括校正后的类型-标记比、平均句长和词长、功能词频率以及线性写作公式。另一方面，Nath等人 [15],针对该任务提出了一种结合双向长短期记忆网络的深度学习孪生神经网络架构。

在本研究中，与现有解决方案不同，我们在基于优选的后期融合框架中采用了多个 Transformer 模型。我们还分析了数据清洗和平衡技术对该框架性能的影响。

III. 方法论

图 1 展示了所提出方法论的不同组成部分。第一个组成部分采用多种预处理技术来清洗和平衡数据集。预处理阶段之后是分类阶段，在此阶段使用多种算法。最后，我们采用多种权重优化和选择方法，融合各个单个模型的得分。以下将详细介绍这些组成部分。

A. 预处理

在预处理阶段，我们执行了两项不同的任务：（i）数据清洗；（ii）数据平衡。在数据清洗任务中，我们删除了不必要的信息，例如用户名、网址、表情符号和停用词。此外，我们使用一个名为*Contractions*[2]的 Python 库扩展（即还原）了缩写形式。

在预处理阶段的第二部分，我们采用转置方法和合成少数类过采样技术（SMOTE）来平衡数据集。在转置方法中，通过交换两个连续/相邻字符来生成新样本。另一方面，SMOTE 方法利用数据增强技术生成与少数类样本更为相似的新样本。

B. 特征提取与分类

在这一阶段，我们采用了传统机器学习和深度学习方法。在传统机器学习方法中，我们选择了几种最先进的机器学习算法，包括随机森林（RF）、朴素

[2]https://pypi.org/project/contractions/

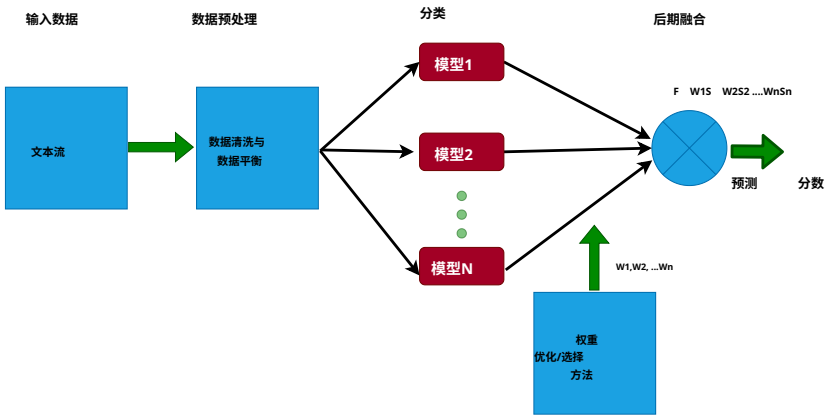

图 1. 所提出方法论的可视化示意图。

贝叶斯（NB）、XGBoost、K近邻（KNN）、逻辑回归、决策树（DT）和支持向量机（SVM）。我们指出，所有这些算法均基于 TF-IDF（词频和逆文档频率）特征进行训练。

另一方面，深度学习方法中使用了若干最先进的模型。这些模型包括 BERT（基于变换器的双向编码器表示）、RoBERTa（鲁棒优化的 BERT）、DistilBERT、XML-RoBERTa、XL-Net 和 ALBERT。这些模型是根据其在相关任务中的已验证性能选定的 [16], [5], [17]。下面简要介绍这些模型。

- **BERT**：BERT [18], 是一种基于 Transformer 的算法，也是最先进的自然语言处理算法之一。它最初在由大量未标注文本构成的语料库（维基百科）上进行预训练，并可针对任何自然语言处理任务进行微调。BERT 具有若干关键特征，使其适用于多种自然语言处理应用。不过，它的优势主要在于双向训练能力，因此可以同时从两个方向读取文本。这种双向训练有助于提取上下文信息。BERT 有两个不同版本，层数和参数数量各不相同。在本研究中，我们采用了具有 12 层和注意力头、参数总数为 1.1 亿的模型。

- **DistilBERT**：DistilBERT [19] 是 BERT 模型的一个更小、更快、更便宜、更轻量的版本。该模型通过对 BERT 基础模型进行知识蒸馏来训练。该模型在 GLUE基准数据集上的表现可达到原始 BERT 模型的 95%，同时参数数量显著减少（约 40%） [19]。参数数量的减少使该模型比原始 BERT 基础模型快 60%。

- **RoBERTa**：RoBERTa [20]是 BERT 的改进版本，对 BERT 的关键超参数进行了调整。例如，它采用更高的学习率和更大的小批量大小进行训练。更重要的是，它移除了 BERT 的下一句预训练目标，从而改进了掩码语言建模目标，相较于原始 BERT 而言，RoBERTa 还使用规模更大的数据集进行训练。

- **XML-RoBERTa**：XLM-RoBERTa 是 RoBERTa 的多语言版本。该模型使用来自 100 种不同语言、总量达 2.5 TB 的文本文件进行训练。选择该模型用于此任务，是基于其在类似任务中已验证的性能 [5]。
- **ALBERT**：ALBERT [21], 被认为是 BERT 的轻量版本，它以 BERT 基础模型为基础，针对三个方面进行了改进，包括：(i) 参数共享、(ii) 嵌入分解，以及 (iii) 句子顺序预测。在参数共享方面，ALBERT 不再为全部 12 层分别使用不同的参数，而是让所有层共用相同的权重/参数值。因此，与原始 BERT 模型相比，ALBERT 的参数数量显著减少。ALBERT 还采用了较小尺寸的嵌入。此外，与原始模型不同，ALBERT 使用基于句子连贯性的句子顺序预测（SOP）损失，而不是仅预测主题。

C. 融合

在融合阶段，使用公式 1 按照后期融合方案组合各个模型的得分。在公式中，S 表示组合后的分类得分，而 s_n 表示第 n 个模型的分类得分。w_1，w_2，w_3，$...w_n$ 表示分配给各个模型的权重，这些权重通过多种权重优化和选择方法计算得出。

$$S = w_1 s_1 + w_2 s_2 + w_3 s_3 + ... + w_n s_n \tag{1}$$

对于权重选择，主要采用三种基于优点的融合方法，即粒子群优化、Nelder-Mead方法和Powell方法。在这些融合方法中，模型权重在验证集上进行优化/选择。在所有融合方法中，适应度函数基于累积分类误差，如公式 2 所示。在该公式中，e 和 A_{acc}分别表示分类错误和累积准确率。

$$e = 1 - A_{acc} \tag{2}$$

累积准确率根据公式 3 计算，其中 x_1， x_2， x_3， $...x_n$ 是待优化的变量（权重值），而 p_1， p_2， p_3， $...p_n$ p_n 表示在验证集上通过各个单个模型获得的概率。

$$A_{acc} = x(1) * p_1 + x(2) * p_2 + ... + x(n) * p_n \quad (3)$$

之所以选择这些优化/选择方法，是因为它们已在类似任务中展现出经过验证的性能 [5]。下面将介绍这些方法的详细信息。

1) 基于*PSO*的融合

我们的第一种基于优点的融合方法基于PSO [22]。PSO本质上是一种启发式方法，这意味着使用PSO获得的解（在我们的应用中为权重组合）并非全局最优。然而，文献表明，使用PSO获得的解通常接近全局最优。PSO的工作机制源于鱼群或鸟群协同行动的方式：群体中的个体通过与其他成员分享各自的发现来相互帮助，从而为整个群体带来最佳的捕猎结果。在PSO中，采用了类似的方法，具体分为三个步骤。首先，根据适应度标准对每个候选解进行评估。其次，更新个体最优值和全局最优值。最后，相应地更新粒子群中每个粒子的位置和速度。

2）基于*Nelder-Mead*方法的融合

我们的第二种基于优点的融合方法基于Nelder-Mead算法，该算法可用于n维优化问题，其中“n”可以为1或大于1。它是一种数值方法，可以求出基于累积分类误差的目标函数的最大值或最小值。我们的实验基于Python开源库SciPy提供的该方法实现[3]。

3）基于*Powell*方法的融合

第三种方法基于Powell方法，这是一种截断最大梯度技术。该方法从初始猜测（权重组合）开始，通过寻找并计算一个良好方向上的距离，逐步向函数的最小值移动。该方法的主要限制之一是需要假设该方向上的步长。该方法简单易用，并且已被证明在类似任务中非常有效 [5]。在本研究中，我们的实现基于Python开源库SciPy[4]。与其他方法类似，我们针对Powell方法定义的适应度函数见公式2。

IV. 实验与结果

A. 数据集

在本研究中，我们使用了 PAN21 作者身份分析数据集[23]。该数据集专为风格变化检测任务创建，收集自热门 StackExchange 问答网站网络中随机选取的讨论串。数据集包含 16,000 份文本文件，来自主题聚焦于技术的相关网站。文本的主题范围受到限制，以确保数据集具有主题连贯性。

通过删除图形内容、网址、代码片段、引文、列表和极短帖子来清理数据。此外，为确保作者的自然性，也从数据集中删除了提交后经过编辑的帖子。文本清理后，将其拆分为段落，同时删除少于 100 个字符的段落。用于预测单作者和多作者文档的原始数据集存在不平衡，其中单作者文档占 25%，多作者文档占 75%。每篇文档的长度为 1,000 至 10,000 个词。此外，每篇文档至少包含 2 个、至多包含 100 个段落。数据集按照 70%（11,200）、15%（2400）和 15%（2400）的比例分别划分为训练数据、验证数据和测试数据。训练数据包含 8,400 篇多作者文档和 2800 篇单作者文档。同样，测试数据集和验证数据集各包含 18,00 篇多作者文档和 600 篇单作者文档。

B. 实验设置

本研究的目标涉及多个方面，我们希望探索该任务的不同层面。一方面，我们希望分析预处理步骤中通常会被删除的字符和内容所起的作用，例如停用词、网址、单独的字母数字字符以及缩写形式。我们认为，其中一些词语，尤其是缩写形式的使用，有助于检测文档中的风格／作者变更。同样，我们还希望分析并评估多种自然语言处理算法在该任务中的性能，包括传统机器学习算法，如随机森林、决策树、SVM，以及 Transformer 模型，如 BERT、RoBERTa、XLM-RoBERTa 和 ALBERT。我们还旨在分析并评估不同融合方法在整合单个模型结果时的性能。

为实现这些目标，我们进行了以下实验。

- 我们在清洗后的数据和原始数据上都进行了实验。
- 我们还在不平衡和经过平衡的数据集上进行了实验，其中采用了三种不同的方法来平衡数据集。
- 我们评估了若干单个模型的性能，包括经典算法和深度学习算法。
- 我们评估了若干融合技术的性能，包括一种朴素融合方法和三种基于评价指标的后期融合方案。

C. 结果

表 I 展示了我们的第一项实验结果：我们根据 F1 分数评估了各个单个模型在干净数据和未清洗数据上的性能。此外，还在平衡数据集和不平衡数据集上评估了这些模型。下面将介绍各个单个模型在不同实验设置下的性能。

- **经典模型与基于 Transformer 的模型**：总共使用了 8 种不同的模型，包括 7 种基于 TF-IDF 特征训练的经典机器学习算法，以及 5 种 Transformer 模型。

[3]https://scipy.org/ [4] https://scipy.org/

正如预期，在所有干净、不干净、平衡和不平衡数据集上，Transformer 模型的结果都优于经典方法。总体而言，XML-RoBERTa 在不干净的不平衡数据上取得了最高分，F1 分数为 0.84。不过，并没有哪一种模型能在所有类型的数据集上都拔得头筹。令人意外的是，与 Transformer 模型相比，一些经典机器学习算法（如随机森林）在干净数据集上的结果更好。这表明，与传统的自然语言处理和机器学习算法相比，Transformer 模型更适合从复杂数据中提取有意义的信息。

- **干净数据集与未清洗数据集**：我们观察到，模型在干净数据集和未清洗数据集上的表现存在有趣的差异。Transformer模型在干净数据集和未清洗数据集上的性能差异显著，而经典模型在两组数据上取得的结果则较为接近。在干净数据集上取得的最高 F1 分数为 0.79，而在未清洗数据集上则为 0.84。干净数据性能的下降表明，预处理步骤移除了对作者检测任务中的Transformer模型有用的词语或字符。值得注意的是，在预处理阶段，我们移除了各种特殊字符，例如用户名、网址、表情符号和停用词。我们还展开了缩写形式。未来，我们计划采用可解释自然语言处理技术，探究究竟是哪些预处理技术导致了有用特征的移除。

- **平衡数据集与不平衡数据集**：模型在平衡数据集和不平衡数据集上的表现也有所不同。总体而言，在不平衡数据集上观察到了更好的结果。此外，我们评估了两种不同的方法，即（i）SMOTE 和（ii）Transpose，用于平衡数据集。在数据平衡技术方面，没有一种方法明显胜出。在某些情况下，SMOTE 得分更高，而在其他情况下，Transpose 方法取得了更好的结果。

表 II 给出了融合实验的结果，我们在其中评估了多种融合方法的性能。需要指出的是，在这些融合方法中，我们仅融合 Transformer 模型的分类得分。这些方法包括三种基于优点的融合方法和一种简单平均融合方法。在平均方法中，所有模型的权重相同；而在其他方法中，则根据单个模型的表现对其权重进行优化或选择。

从表中可以看出，与单个模型相比，在干净数据集和未清洗数据集上，融合总体上都能取得更好的结果。在干净数据集上，融合方法相较于表现最佳的单个模型有显著提升。然而，在未清洗数据集上，融合方法带来的提升较小。

此外，在融合方法中，与朴素融合方法相比，基于优点的融合方法取得了更好的结果。这表明基于优点的融合方法有效，也说明融合时应考虑单个模型的性能。就基于优点的融合方法的评估而言，未观察到显著差异。总体而言，与其他权重优化/选择方法相比，基于 Powell 方法的融合取得了略好的结果。

我们还在表III中将所提出的方法与现有解决方案进行了比较。从表中可以看出，所有基于优点的融合方法相比现有研究成果都有显著提升。我们表现最佳的融合方法（即基于 Powell 方法的融合）比文献中得分最高的解决方案高出 5.7%。

V. 结论

本文提出了一种基于优点的融合框架，通过采用多种权重选择和优化方法，联合使用多种自然语言处理算法，对单作者与多作者文档进行分类。此外，我们评估了多种单个算法的性能，包括传统机器学习和深度学习模型。我们还使用干净数据和未清洗数据进行了实验，以分析字符对文体检测的影响；在自然语言处理应用中，这些字符通常会在预处理过程中被排除。我们还评估了多种数据平衡技术的性能。总体而言，Transformer模型取得了更好的结果。与单个模型相比，基于优点的融合技术显著提升了性能。我们还发现，在自然语言处理应用的预处理步骤中通常被排除的字符，在文体检测中发挥着关键作用。未来，我们计划通过探索该任务中其他尚未解决的问题来拓展这项工作，例如结合基于词汇特征训练的分类器与深度学习模型所获得的分数。

参考文献

[1] M. Kestemont、M. Tschuggnall、E. Stamatatos、W. Daelemans、G. Specht、B. Stein 和 M. Potthast，“PAN-2018 作者识别任务概述：跨领域作者归属判定和文体变化检测”，载于 *CLEF 2018*评测实验室工作笔记论文集。法国阿维尼翁，*2018*年*9*月*10*日至*14*日*/Cappellato, Linda [*编。*]*; 等人。，2018年，第1–25页。[2] F. Alvi、H. Algafri 和 N. Alqahtani，“使用话语标记进行文体变化检测。”CLEF，2022年。[3] J. Bevendorff、B. Chulvi、E. Fersini、A. Heini、M. Kestemont、K. Kredens、M. Mayerl、R. Ortega-Bueno、P. Pkezik、M. Potthast 等人，“PAN 2022 概述：作者身份验证、反讽画像分析和刻板印象传播者、文体变化检测以及触发词检测”，载于 欧洲信息检索会议。Springer，2022年，第331–338页。[4] E. Zangerle、M. Mayerl、M. Potthast 和 B. Stein，“PAN 文体变化检测任务概述”，2021年。[5] K. Ahmad、M. Ayub、J. Khan、N. Ahmad 和 A. Al-Fuqaha，“社交媒体作为获取水质即时反馈的来源”，*IEEE*《技术与社会汇刊》，2022年。[6] T. Neal、K. Sundararajan、A. Fatima、Y. Yan、Y. Xiang 和 D. Woodard，“文体计量技术及其应用综述”，*ACM*计算调查（*CSuR*），第50卷，第6期，第1–36页，2017年。[7] E. Stamatatos，“现代作者归属判定方法综述”，《美国信息科学与技术学会会刊》，第60卷，第3期，第538–556页，2009年。[8] K. Lagutina、N. Lagutina、E. Boychuk、I. Vorontsova、E. Shliakhtina、O. Belyaeva、I. Paramonov 和 P. Demidov，“文体计量文本特征综述”，载于 *2019*年第*25*届开放创新协会会议（*FRUCT*）。IEEE，2019年，第184–195页。[9] A. Glover 和 G. Hirst，“检测协作写作中的文体不一致”，载于 《新写作环境》。Springer，1996年，第147–168页。

表 I 单个模型在干净、原始、平衡及不平衡数据集上的评估。

模型	未清洗数据			干净数据		
	不平衡	平衡（SMOTE）	平衡（转置）	不平衡	平衡（SMOTE）	平衡（转置）
随机森林	0.62	0.6	0.61	0.63	0.49	0.56
朴素贝叶斯	0.43	0.45	0.43	0.43	0.51	0.48
XGBoost	0.62	0.62	0.62	0.62	0.48	0.36
KNN	0.49	0.48	0.49	0.49	0.51	0.48
SVM	0.57	0.46	0.21	0.57	0.45	0.47
决策树	0.51	0.53	0.51	0.52	0.49	0.46
逻辑回归	0.5	0.48	0.21	0.5	0.51	0.46
BERT	0.79	0.75	0.79	0.79	0.59	0.79
RoBERTa-Base	0.8	0.7	0.43	0.65	0.59	0.62
Albert	0.79	0.79	0.53	0.62	0.62	0.53
DistilBERT	0.75	0.72	0.2	0.76	0.6	0.2
XLM-RoBERTa	0.8481	0.75	0.79	0.66	0.59	0.79

表 II 干净、原始、平衡和不平衡数据集上的融合方法评估。

融合方法	未清洗数据			干净数据		
	不平衡	平衡（SMOTE）	平衡（转置）	不平衡	平衡（SMOTE）	平衡（转置）
简单平均	0.8272	0.7673	0.4546	0.6623	0.6167	0.2074
PSO方法	0.8486	0.8031	0.5301	0.6511	0.6007	0.54
Nelder Nead方法	0.8491	0.8031	0.52381	0.6439	0.56575	0.5397
Powell方法	0.8523	0.8031	0.52381	0.6412	0.5657	0.54

表III 与现有解决方案的比较。

方法	F1 分数
Zhang等人 [11]	0.753
Strom等人 [12]	0.795
Singh等人 [13]	0.634
Deibel等人 [14]	0.621
Nath等人 [15]	0.704
基于 Nelder–Mead 方法的融合（本工作）	0.849
基于PSO的融合（本研究）	0.848
基于 Powell 方法的融合（本研究）	0.852

[10] N. Akiva 和 M. Koppel，“一种用于分解多作者文档的通用无监督方法”，《美国信息科学与技术学会会刊》，第64卷，2264，2013年11月，2256期，第–页。[11] Z. Zhang，“基于写作风格相似性的文体变化检测。”[12] E. Strøm，“通过解决一个二元分类问题实现多标签风格变化检测。”载于 *CLEF*（工作笔记），2021年，第2146–2157页。[13] R. Singh、J. Weerasinghe 和 R. Greenstadt，“多作者文档的写作风格变化检测。”载于 *CLEF*（工作笔记），2021年，第2137–2145页。[14] R. Deibel 和 D. L¨offlad，“使用基于 LSTM 的归因算法对真实世界数据进行文体变化检测。”载于 *CLEF*（工作笔记），2021年，第1899–1909页。[15] S. Nath，“使用孪生神经网络进行文体变化检测。”载于 *CLEF*（工作笔记），2021年，第2073–2082页。[16] N. Said、K. Ahmad、M. Riegler、K. Pogorelov、L. Hassan、N. Ahmad 和 N. Conci，“社交媒体和卫星影像中的自然灾害检测：综述”，《多媒体工具与应用》，第78卷，第22期，第31 267–31 302页，2019年。[17] K. Ahmad、F. Alam、J. Qadir、B. Qolomany、I. Khan、T. Khan、M. Suleman、N. Said、S. Z. Hassan、A. Gul 等人，“全球用户层面对新冠肺炎接触者追踪应用的认知：使用自然语言处理的数据驱动方法”，《*JMIR Formative Research*》，第6卷，第5期，第e36238页，2022年。[18] J. Devlin、M.-W. Chang、K. Lee 和 K. Toutanova，“BERT：用于语言理解的深度双向 Transformer 预训练”，*arXiv* 预印本 *arXiv:1810.04805*，2018年。[19] V. Sanh、L. Debut、J. Chaumond 和 T. Wolf，“DistilBERT，BERT 的蒸馏版本：更小、更快、更便宜且更轻量”，*arXiv* 预印本 *arXiv:1910.01108*，2019年。

[20] Y. Liu、M. Ott、N. Goyal、J. Du、M. Joshi、D. Chen、O. Levy、M. Lewis、L. Zettlemoyer 和 V. Stoyanov，“RoBERTa：一种鲁棒优化的BERT预训练方法”，*arXiv* 预印本 *arXiv:1907.11692*，2019年。[21] Z. Lan、M. Chen、S. Goodman、K. Gimpel、P. Sharma 和 R. Soricut，“ALBERT：一种用于语言自监督学习的轻量级BERT”，*arXiv* 预印本 *arXiv:1909.11942*，2019年。[22] J. Kennedy 和 R. Eberhart，“粒子群优化”，载于*ICNN'95*—国际神经网络会议论文集，第4卷。IEEE，1995年，第1942–1948页。[23] E. Zangerle、M. Mayerl、M. Tschuggnall、M. Potthast 和 B. Stein，“PAN21作者身份分析：文体变化检测”，2021年3月。[在线]。链接：https://doi.org/10.5281/zenodo.4589145