面向小样本分类的随机配置网络研究

袁赛仙[1] 姚雪梅[1*] 唐 艳[1] 邹红梅[1] 蔡丽琼[1]
YUAN Saixian YAO Xuemei TANG Yan ZOU Hongmei CAI Liqiong

摘 要 针对小样本数据分类任务中存在的样本稀缺、标注成本高、模型泛化能力弱及类间区分性不足等问题，文章提出一种基于随机配置网络（SCN）的集成分类模型 E-SCN。该模型通过正态分布初始化网络节点，并引入了多数投票机制集成多个差异化的 SCN 基分类器，以增强模型在小样本条件下的判别能力与泛化性能。实验结果表明：E-SCN 相较于 SCN 及传统方法在分类精度与鲁棒性方面具有显著的优势，为小样本分类问题提供了新的有效解决方案，拓展了 SCN 在低资源学习场景中的应用潜力。

关键词 随机配置网络（SCN）；小样本分类；多数投票；带多数投票的正态 SCN（E-SCN）

doi：10.3969/j.issn.1672-9528.2026.01.028

0 引言

传统机器学习模型通常依赖大量数据来进行训练，以获得优异的成果。然而，在医学影像分析、用户行为评价等高度依赖专家标注的场景中，获取这一类数据成本高昂。因此，数据稀缺问题严重制约了传统机器学习模型。小样本分类任务要求模型基于极少量标注的样本准确识别新类别。以往的研究中通常将样本量与特征数之比小于 20 的数据集定义为小样本数据集 [1]。小样本分类任务要求模型就是仅基于极少量标注样本准确识别出新类别 [2]。

目前常见的小样本学习方法包括基于相似度度量的模型，如原型网络 [3] 和匹配网络 [4]，但其在高维数据上的泛化能力有限；元学习算法如 MAML[5] 虽具有较强适应性，但计算成本较高；自监督方法如 SimCLR[6] 则在类别区分任务中敏感性不足。随机配置网络（SCN）[7] 通过随机生成隐含层参数并结合增量学习机制，在抑制过拟合的同时提升了小样本下的泛化性能。后续研究提出了深度扩展 [8]、抗噪性能等版本 [9]，进一步拓宽了 SCN 的应用范围。然而，面对高维小样本数据时，SCN 的随机节点生成机制难以捕捉最优特征组合，对噪声较为敏感，过拟合风险依然存在。

因此，本文提出了一种带多数投票的正态 SCN 集成模型（E-SCN）提升小样本场景下的分类性能。E-SCN 通过正态分布初始化节点参数，并结合多数投票集成策略，构建多个具有多样性的基础 SCN 模型，进而通过多数投票显著提高了模型的泛化能力与分类准确率。

1 相关工作

1.1 随机配置网络（SCN）

RVFL 训练在逼近目标函数方面时，随机参数的设置没有被最佳配置，导致函数逼近失败的概率明显较高。为克服这个问题，Wang 等 [7] 开发了一种用于隐藏节点配置的监督机制，提出一种增量网络构建的方法，用来解决与 RVFL 相关的局限性，这种方法被称为 SCN。

该模型从一个小网络开始逐个增量生成隐藏节点，直到获得可接受的容错性。假设建立了一个具有 L−1 个隐藏节点的神经网络模型：$f_{L-1}=\sum_{j=1}^{L-1}\beta_j g_j$，$\boldsymbol{e}_{L-1}=f-f_{L-1}$表示残差。SCN 提供了一种快速的解决方案，以递增地添加第 L 个隐藏节点，并重复此过程，直到残差达到可接受的水平。Wang 等 [7] 在理论中详细描述了 SCN 的理论。当$L\to\infty$时，误差的范数$\left\|f-f_L^*\right\|\to 0$，网络能够逼近目标函数$f$。虽然 SCN 在处理一般分类任务时有其独特的优势，但对于小样本分类来说，存在过拟合、性能不稳定、模型复杂度与数据集规模不匹配等问题。这些问题使得 SCN 在小样本环境下的表现有限。在 Wang 等的理论中，生成的隐藏节点 h_L 必须满足不等式条件$\left\langle \boldsymbol{e}_{L-1,k},h_L\right\rangle^2\geq b_h^2\delta_{L,k}$，以确保能够与误差向量 $\boldsymbol{e}_{L-1,k}$ 有足够的相关性。

在小样本分类情况下，数量极少的训练样本会让生成的隐藏节点空间过于稀疏，进而影响模型的拟合能力。特别是在复杂任务下，小样本数不足以支撑网络生成有效的节点，从而难以捕捉数据中的复杂模式和特征。同时，当$L\to\infty$时，

1. 贵州民族大学数据科学与信息工程学院 贵州贵阳 550025
[基金项目] 贵州省高等学校机器智能产品研发创新团队（黔教技〔2022〕15 号）；贵州民族大学本科教学研究项目（GZMUJG202302）；教育部产学合作协同育人项目（2309173212）；贵州民族大学自然科学基金项目（GZMUZK〔2022〕YB07）；贵州大学省部共建公共大数据国家重点实验室开放课题（PBD2022-22）

理论上 SCN 的误差范数$\left\|f-f_L^*\right\|\to 0$，但在实际情况中，样本过少会使得这个逼近过程产生较大的方差，表现为过拟合。模型会趋向于记住训练数据的特定模式，而不是捕捉到数据的真实分布。于是本文提出带多数投票的正态 SCN 集成模型（E-SCN）解决 SCN 的泛化能力问题。

1.2 带多数投票的正态 SCN 集合模型（E-SCN）

E-SCN 是将 SCN 与多数投票法和正态分布节点初始化结合的模型。通过集成多个 SCN 模型，每个模型在初始化阶段使用正态分布进行随机生成，并通过执行多数投票决策进行分类任务。图 1 为集成模型的算法流程图。

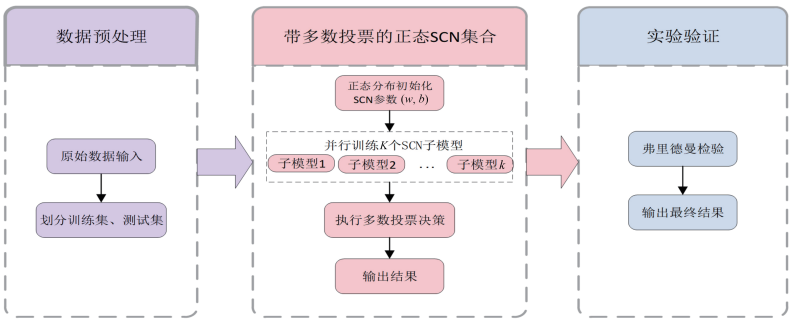

图 1 E-SCN 算法流程图

假设对于每个 SCN 模型 m，隐藏节点 $h_L^{(m)}$ 通过正态分布$\mathcal{N}\left(\mu_h,\sigma_h^2\right)$初始化，用公式表示为：

$$h_L^{(m)}(X)\sim\mathcal{N}\left(\mu_h,\sigma_h^2\right),\quad L=1,2,\ldots,L_m \tag{1}$$

式中：X 为输入数据；L_m 为第 m 个 SCN 模型中的隐藏节点数量；μ_h 和 σ_h^2 分别为正态分布的均值和方差，可以通过调参或数据统计确定。

对于第 m 个 SCN 模型，其输出是通过每个隐藏节点的加权和计算得到的，用公式表示为：

$$f^{(m)}(X)=\sum_{j=1}^{L_m}\beta_j^{(m)}h_j^{(m)}(X) \tag{2}$$

式中：$f^{(m)}(X)$ 是第 m 个 SCN 模型的预测结果；权重 $\beta_j^{(m)}$ 的优化方式依然采用误差最小化的形式。在集成框架下，有 M 个正态分布初始化的 SCN 模型，每个模型给出一个分类预测。

对于分类任务，集成的决策采用多数投票法。对于每个输入 X，第 m 个模型的预测类别为$\hat{y}^{(m)}$，集成后的最终输出结果为：

$$\hat{y}_{\text{ensemble}}=\text{Mode}\left(\left\{\hat{y}^{(1)},\hat{y}^{(2)},\ldots,\hat{y}^{(M)}\right\}\right) \tag{3}$$

式中：Mode 表示多数投票的结果，即出现次数最多的预测类别。如果出现平局，可以通过随机选择或预设规则来打破平局。假设有 M 个基于正态分布初始化的 SCN 模型，那模型集成的最终预测也可以表达为：

$$\hat{y}_{\text{ensemble}}=\text{Mode}\left(\left\{\hat{y}^{(m)}\right\}_{m=1}^{M}\right) \tag{4}$$

每个 SCN 模型的预测$\hat{y}^{(m)}$由其输出$f^{(m)}(X)$转换为类别标签得到式（5），其中$\mathcal{C}$是类别集合，$P\left(y=c|f^{(m)}(X)\right)$是模型 m 对于类别 c 的预测概率。该概率可以通过模型的输出 $f^{(m)}(X)$ 经过 softmax 函数得到。

$$\hat{y}^{(m)}=\arg\max_{c\in\mathcal{C}}P(y=c\mid f^{(m)}(X)) \tag{5}$$

综上，E-SCN 集成模型的输出公式可以归纳为：

$$\hat{y}_{\text{ensemble}}=\text{Mode}\left(\left\{\arg\max_{c\in\mathcal{C}}P\left(y=c|f^{(m)}(X)\right)\right\}_{m=1}^{M}\right) \tag{6}$$

E-SCN 结合了正态分布节点初始化的优点和多数投票集成策略的简单易实现，通过多个随机配置网络的投票决策，模型在分类任务中的表现能够得到显著提升。该方法不仅有效提高了分类准确性，还增强了模型的稳定性、泛化能力和抗过拟合性能，同时保持了较低的计算复杂度和易于实现的优势，使其在实际应用中具有重要价值。

2 实验

2.1 实验设置

所有实验均在 MATLAB R2023a、Intel Core CPU 12450H 2.0 GHz、32 GB RAM 和 Windows 11 操作系统上进行。实验将使用 4 种模型：随机向量函数链接网络（IRVFLN）、具有单个隐含层的多层感知器（MLP）、SCN、E-SCN。

MLP 模型的学习率固定为 0.01，训练周期设定为 100 次且不设置提前停止条件，采用误差反向传播算法进行训练。在 SCN、E-SCN 中，通过不同的方式来提取样本，以提升模型在小样本数据中样本的多样性。SCN 利用随机配置和增量学习策略来构建能够适应小样本数据分布的神经网络结构。E-SCN 在 SCN 的基础上结合了正态分布节点初始化的优点和多数投票集成策略。为了评估 E-SCN 和 SCN 的性能，本文采用了来自 UCI 机器学习库的数据集，每个数据集的属性如表 1 小样本分类数据集属性所示。

表 1 小样本分类数据集属性

数据集	样本	特征	类别	比率 /%
Mfeat-fou	2 000	76	10	7.89
Mfeat-pix	2 000	240	10	2.5
Mfeat-kar	2 000	64	10	9.38
iris	150	4	3	11.25
vowel	528	10	11	15.8
spect	267	22	2	3.68

为了模拟小样本数据的情况，实验中所有数据集仅使用总数据的 30% 作为训练数据。表中列出了所有 10 个数据集的训练样本与特征的比率。由于当样本与特征的比率小于 20% 时，数据集被认为是小数据集，因此本实验中的所有数据集都保持了小样本数据属性。

2.2 实验结果及分析

表 2 展示了四种模型（MLP、IRVFLN、SCN、E-SCN）在六个不同分类数据集（iris、Mfeat-fou、Mfeat-kar、Mfeat-pix、vowel、spect）上的分类准确率表现。其中，粗体显示的是每个数据集上表现最佳的分类准确率。

表 2 MLP、IRVFLN、SCN 和 E-SCN 在小样本分类数据集上的平均分类准确率（ACC）

Datasets	Test Result（%）			
	MLP	IRVFLN	SCN	E-SCN
iris	77.24	88.76	81.04	**89.52**
Mfeat-fou	61.78	52.34	**91.49**	90.18
Mfeat-kar	73.30	43.85	92.14	**93.00**
Mfeat-pix	81.82	76.58	90.95	**96.57**
vowel	25.52	45.41	90.66	**90.88**
spect	79.57	77.76	75.79	**79.25**

从表 1 中可以看出，E-SCN 模型在所有数据集上表现最优，其准确率普遍高于其他模型，尤其是在 vowel 和 Mfeat-pix 数据集上分别达到了 90.88% 和 96.57%，展现出较高的泛化能力和稳定性。这表明 E-SCN 的多数投票策略增强了原始 SCN 的分类能力和泛化性能。

另外，还可以观察到 MLP 和 IRVFLN 在大多数数据集上的准确率相对较低，表现出较高的不稳定性。所以在面对不同特征分布的数据集时，MLP 和 IRVFLN 的适应能力不如 SCN 及其变体模型。

综上所述，表 2 的结果验证了 E-SCN 的设计有效性。该模型增加的多数投票机制在部分数据集上提升了 SCN 的分类性能，为小样本数据集的分类任务提供了更优的解决方案。

为了更准确地评估各模型的整体性能，本文对四种模型进行了弗里德曼检验（Friedman Test），弗里德曼检验适用于比较多个相关样本的性能差异，且是一种非参数检验方法，在本实验中，显著性水平设定为 0.05。弗里德曼检验的结果如表 3 所示。

表 3 Friedman Test 表

值	平方和	自由	均方	卡方统计量	显著性水平
模型	14	3	4.67	8.4	0.038 4
误差	16	15	1.07	—	—

检验结果显示，各模型之间的卡方统计量为 8.4，自由度为 3，对应的 p 值为 0.038 4。由于 p 值小于 0.05，说明模型之间的性能差异在统计上具有显著性。因此，可以拒绝零假设，即认为至少有一个模型的性能显著不同于其他模型，实验合理。

3 结论

针对小样本数据分类的问题，本文引入了随机配置网络（SCN）作为基础方法，为了进一步提升 SCN 在小样本分类任务中的性能，本文提出了一种基于多数投票策略的正态随机配置网络的集成模型（E-SCN）。在小样本分类任务中，传统模型常常因为过拟合和泛化能力的不足而受到限制。于是，E-SCN 通过构建多个正态分布初始化的 SCN 子模型并集成它的输出，显著提升了 SCN 的分类精度与鲁棒性。实验研究结果表明，E-SCN 在多个小样本数据集上，能充分利用有限样本，有效缓解过拟合，从而提高小样本的分类性能。在未来面向大规模或高维数据的任务时，E-SCN 还可以引入并行计算、模型剪枝等这些技术进一步提升效率，增强在医学影像、金融风控等这些实际场景中的应用。

参考文献：

[1]OOI B P, RAHIM N A, MASNAN M J, et al. A study of extreme learning machine on small sample-sized classification problems[J].Journal of physics: conference series. IOP publishing, 2021, 2107:012013.

[2]SONG Y S, WANG T, CAI P Y, et al. A comprehensive survey of few-shot learning: evolution, applications, challenges, and opportunities[J]. ACM computing surveys, 2023, 55(13s): 1-40.

[3] 邱云飞 , 牛佳璐 . 融合小样本元学习和原型对齐的点云分割算法 [J]. 中国图象图形学报 , 2023, 28(12):3884-3896.

[4] 胡正伟 , 夏思懿 , 王文彬 , 等 . 基于深度强化学习的 Π 型阻抗匹配网络多参数最优求解方法 [J]. 电力系统保护与控制 , 2024,52(6):152-163.

[5]HE Y P, ZANG C Z, ZENG P, et al. Convolutional shrinkage neural networks based model-agnostic meta-learning for few-shot learning[J]. Neural processing letters, 2023,55(1): 505-518.

[6]CHEN T, KORNBLITH S, NOROUZI M ,et al.A simple framework for contrastive learning of visual representations[EB/OL].(2020-07-01)[2025-12-31].https://doi.org/10.48550/arXiv.2002.05709.

[7]WANG D H, LI M. Stochastic configuration networks: fundamentals and algorithms[J].IEEE transactions on cybernetics, 2017, 47(10): 3466-3479.

基于大语言模型的多路径一致性蒸馏

蔡丽琼[1] 霍雨佳[1*] 袁赛仙[1]
CAI Liqiong HUO Yujia YUAN Saixian

摘 要 提升小型语言模型在复杂推理任务中的表现，是当前自然语言处理领域的一个关键挑战。基于此，文章提出了一种基于知识蒸馏的推理增强方法（CGRD），旨在提升学生模型在复杂推理任务中的表现。具体而言，采用 GPT-3.5-turbo-instruct 作为教师模型，通过 API 调用生成多样化的推理路径，从中筛选出一致性最高的路径用于指导学生模型的训练。实验结果表明，CGRD 方法在 AddSub 和 Date Understanding 两个数据集上均显著优于传统的方法。还使教师模型在 AddSub 和 Date Understanding 数据集上的准确率分别提升了 15.11 和 10.05 个百分点，更重要的是，在 GPT-2 和 T5 等不同架构的一系列小规模学生模型上均取得了一致性的性能提升，充分证明了 CGRD 在增强小模型推理能力方面的有效性，通过提升教师模型生成推理路径的质量，有效增强了学生模型的推理能力和任务表现。

关键词 蒸馏；大型语言模型；思维链；一致性推理

doi：10.3969/j.issn.1672-9528.2026.01.029

0 引言

近年来，随着大规模语料库上预训练的 Transformer 模型的广泛应用，预训练语言模型（PLM）在自然语言处理（NLP）领域展现了显著的优势 [1-2]。研究表明，当模型的参数规模超过一定阈值时，这些大规模语言模型在性能上实现了显著提升，并展现出小规模模型（如 BERT[3]）无法实现的能力，例如上下文学习和推理能力。为区分不同规模的语言模型，学术界和工业界引入了“大语言模型”（LLMs）这一术语，这类模型通常拥有数百亿至数千亿参数。在人工智能的发展中，专有的大语言模型（如 GPT-3.5、GPT-4、Gemini 和 Claude 作为开创性技术，重塑了对自然语言处理（NLP）的理解。其核心优势在于突显能力 [4-5]，即模型展示出超出明确培训目标的能力，使其能够高效处理各种复杂任务。然而，尽管像 GPT-4、GPT-4o 和 Gemini 等专有 LLM 展示了卓越的性能，也存在明显缺点，尤其是在考虑开源模型的优势时。使用这些模型往往伴随着高昂的费用和访问限制，且数据隐私和安全问题 [6] 也成为一个重要考量。

与此相比，开源模型，如 LLaMA 在可访问性和适应性方面具有显著优势。未许可使用或使用限制，开源模型使更多的用户能够受益，从个人研究人员到小型组织都可以轻松使用。然而，开源 LLM 也存在一些不足，主要体现在模型规模和资源上，相较于专有模型，通常具备较少的参数，这导致其在处理复杂任务时性能较弱 [7]。

因此，知识蒸馏作为一种有效的模型压缩技术，逐渐成为提升大语言模型应用效率的关键方法之一。通过蒸馏，小型学生模型能够借助教师模型的知识，接近甚至超越大模型的性能，从而在减少计算成本的同时提高推理效率。Bucilă

1. 贵州民族大学数据科学与信息工程学院 贵州贵阳 550025
[基金项目] 贵州省教育厅自然科学研究项目（黔教技〔2022〕047 号）（贵州省高等学校智慧教育工程研究中心支持完成）；贵州民族大学自然科学研究基金（GZMUZK〔2021〕YB22）

[8]LI J Q, WANG D H. 2D convolutional stochastic configuration networks[J/OL].Knowledge-based systems, 2024[2025-12-31].https://dl.acm.org/doi/10.1016/j.knosys.2024.112249.

[9]DENG X G, ZHAO Y , ZHANG J, et al. A holistic global-local stochastic configuration network modeling framework with antinoise awareness for efficient semi-supervised regression[J].Information science, 2024, 661:120132.

【作者简介】

袁赛仙（1999—），女，贵州毕节人，硕士，研究方向：多源数据融合、小样本学习。

姚雪梅（1985—），通信作者（email:yaomei0119@126.com），女，云南大理人，博士，副教授、硕士生导师，研究方向：多源数据融合及大数据挖掘与分析。

（收稿日期：2025-09-18 修回日期：2026-01-12）