

一种空间几何角度最大化的随机增量学习模型及应用

南静¹ 代伟¹ 袁冠² 周平³

摘要 针对随机神经网络 (Random weight neural networks, RWNNs) 隐含层节点随机生成过程可解释性不足和节点随机生成而导致的网络结构不紧致等问题, 提出了一种空间几何角度最大化随机增量学习模型 (Stochastic incremental learning model with maximizing spatial geometry angle, SGA-SIM). 首先, 以空间几何视角深入分析随机增量学习过程, 建立了具有可解释性的空间几何角度最大化约束, 以改善隐含层节点质量, 并证明该学习模型具有无限逼近特性; 同时, 引入格雷维尔迭代法优化学习模型输出权值计算方法, 提高模型学习效率. 在真实的分类和回归数据集以及数值模拟实例上的实验结果表明, 所提增量学习模型在建模速度、模型精度和模型网络结构等多个方面具有明显优势.

关键词 随机神经网络, 增量学习, 空间几何角度最大化约束, 无限逼近性

引用格式 南静, 代伟, 袁冠, 周平. 一种空间几何角度最大化的随机增量学习模型及应用. 自动化学报, 2023, 49(6): 1283–1294

DOI 10.16383/j.aas.c211041

A Stochastic Incremental Learning Model With Maximizing Spatial Geometry Angle and Its Application

NAN Jing¹ DAI Wei¹ YUAN Guan² ZHOU Ping³

Abstract Aiming at the problems of insufficient interpretability and incompact network structure caused by random generation of hidden nodes in random weight neural networks (RWNNs), this paper proposes a stochastic incremental learning model with maximizing spatial geometry angle (SGA-SIM). Firstly, the random incremental learning process is deeply analyzed from the perspective of spatial geometry, then an interpretable spatial geometric angle maximization constraint is established to improve the quality of the hidden nodes, and the universal approximation property of this model is proved. Besides, the Grenville iteration method is introduced to optimize the output weight calculation, which improves the learning efficiency of the learning model. Experimental results on the real datasets and numerical simulation examples show that the proposed model has obvious advantages in modeling speed, model accuracy and model network structure.

Key words Random weight neural networks (RWNNs), incremental learning, spatial geometric angle maximization constraint, universal approximation property

Citation Nan Jing, Dai Wei, Yuan Guan, Zhou Ping. A stochastic incremental learning model with maximizing spatial geometry angle and its application. *Acta Automatica Sinica*, 2023, 49(6): 1283–1294

收稿日期 2021-11-03 录用日期 2022-09-26

Manuscript received November 3, 2021; accepted September 26, 2022

国家自然科学基金 (61973306), 江苏省自然科学基金优秀青年项目 (BK20200086), 江苏省研究生科研与实践创新计划 (KYCX22_2552), 中国矿业大学未来杰出人才助力计划 (2022WLKXJ077) 资助
Supported by National Natural Science Foundation of China (61973306), Outstanding Youth Project of Jiangsu Provincial Natural Science Foundation (BK20200086), Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX22_2552), and Assistance Program for Future Outstanding Talents of China University of Mining and Technology (2022WLKXJ077)

本文责任编辑 刘艳军

Recommended by Associate Editor LIU Yan-Jun

1. 中国矿业大学信息与控制工程学院 徐州 221116 2. 中国矿业大学计算机科学与技术学院 徐州 221116 3. 东北大学流程工业综合自动化国家重点实验室 沈阳 110819

1. School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116 2. School of Computer Science & Technology, China University of Mining and Technology, Xuzhou 221116 3. State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819

随着计算机技术和处理能力的逐渐提升, 神经网络模型及其学习技术得到了长足发展. 特别是深度神经网络, 其凭借着强大的数据分析和函数表达能力被广泛应用于图像处理^[1]和语音识别^[2]等领域. 但庞大的网络超参数和复杂的网络结构使得大多数深度神经网络训练过程非常耗时且对硬件设备要求较高. 实际上, 扁平结构的神经网络对非线性映射具有无限逼近能力, 其强大的表达能力足已胜任许多数据分析任务, 典型的模型为单隐含层前馈神经网络^[3-4]. 但是, 传统单隐含层前馈神经网络在训练过程中使用梯度下降法获取网络参数, 故存在收敛速度慢和易陷入局部极小值等问题^[5], 这些问题也一直是困扰神经计算发展的难题.

随机学习方法的快速建模和低计算消耗等特性, 为单隐含层前馈神经网络提供了一种不同的学

习思路^[6]。目前,已涌现出多种单隐含层随机权前馈神经网络^[7-9],它们统称为随机权神经网络(Random weight neural networks, RWNNs),这种网络的特点是:从固定区间内随机产生隐含层节点参数(输入权值和偏置)保持不变,然后使用解析法求解网络输出权值。文献[10]对RWNNs的基础理论进行了重新讨论,并证明若目标函数满足Lipschitz条件,则在对称区间内随机生成隐含层节点参数也能保证RWNNs具有无限逼近性。但是,传统RWNNs难以确定合适的网络拓扑结构,其往往通过实验方式试凑确定。在实际建模任务中,网络拓扑结构过大会导致模型过度泛化,过小则会引起模型表达能力不足^[11],因此,寻找合适的网络拓扑结构通常是一项繁重的工作。构造法^[11],剪枝法^[12]和正则化法^[13]是三种解决网络拓扑结构问题的主要技术。剪枝法作为一种模型压缩方法,其首先训练一个远超出建模任务需求的大型模型,然后根据模型误差变化、输出相关度和隐含层节点灵敏度等规则对冗余节点进行修剪,同时保持模型性能不变^[14]。但剪枝法先建大模型再压缩的建模方式效率低下,且需要消耗大量的计算和存储资源。构造法从一个小网络(通常是一个隐含层节点)开始,然后逐步添加隐含层节点,使网络残差下降,直到满足预期目标。相比于剪枝法,构造法往往会构造具有相同表达能力但网络拓扑结构更紧致的模型,因此,相比于剪枝法效率更高。正则化方法主要是通过设置惩罚系数来平衡目标函数值和约束违反值以提升模型性能。但惩罚系数需要通过交叉验证^[15]或人为调节^[16]等实验方法进行确定,且正则化方法并没有改变网络结构。鉴于构造法良好的建模效率,本文集中研究如何利用构造法建立RWNNs。

构造法本质上是一种增量式方法,文献[17]将构造法引入RWNNs,提出了增量式RWNNs(Incremental RWNNs, IRWNNs),其首先以点增量方式为网络添加隐含层节点,并从 $[-1, 1]$ 的固定区间内随机生成节点的输入权值和偏置且保持不变;然后,利用Moore-Penrose广义逆计算隐含层和输出层之间的输出权值;最后,计算整个网络的残差,重复上述步骤直到满足期望目标为止。当前,IRWNNs已发展了正交IRWNNs^[18]和集成IRWNNs^[19]等多个版本。然而,文献[20]和文献[21]的研究结果从理论上证明了从固定区间内随机产生隐含层节点参数难以保证IRWNNs的无限逼近能力。同时,文献[22]也证明了将随机学习方法应用到单隐层前馈神经网络时需要基于建模任务对参数随机生成过程进行约束,否则难以用很高的概率逼近目标函

数。在此基础上,文献[23]提出了一种具有监督机制约束的构造算法来构建神经网络,即随机配置网络(Stochastic configuration networks, SCNs)。SCNs与其他构造算法最大的不同在于其利用不等式约束从隐含层参数分配区间内直接选择优质参数而不是简单的随机生成。进一步地,鲁棒SCNs^[24]、并行SCNs^[25]和块增量SCNs^[26]等被提出以增强SCNs的鲁棒性和学习效率。此外,为了提高随机学习方法对数据特征的提取能力和模型更新效率,文献[27]提出了一种宽度学习(Broad learning system, BLS)构造算法,BLS在模型结构上与其他构造算法有着明显的区别,即:BLS首先将输入样本映射为特征节点,然后由这些特征节点生成增强节点,最后再将特征节点和增强节点一起输入模型。BLS主要关注如何快速构建模型,因此BLS仍然存在由参数随机分配而导致的网络结构不紧致问题。目前,BLS已经被广泛应用于工业噪声处理^[28]、模糊系统^[29]和类脑^[30]等研究领域。虽然构造算法进一步提高了RWNNs的潜力,但缺少关于隐含层参数如何实现它们目标的研究。即,现有构造算法很难将隐含层参数对模型残差的影响可视化。因此,如何增强参数随机生成过程可解释性和避免参数随机生成而导致网络结构不紧致等问题,仍然是一个巨大挑战。

本文旨在以空间几何角度视角深入分析扁平神经网络随机增量建模过程中隐含层节点参数与残差变化之间的内在几何关系,建立一种具有可解释性的高效学习模型,即,空间几何角度最大化的随机增量学习模型(Stochastic incremental learning model with maximizing spatial geometry angle, SGA-SIM)。其主要贡献是:

- 1) 以空间几何视角对随机增量学习模型建模过程进行分析,建立隐含层节点参数与残差变化之间的几何关系;
- 2) 建立具体可解释性的空间几何角度最大化约束以提升隐含层节点参数质量,并结合残差变化证明在该约束下随机增量学习模型的无限逼近性;
- 3) 提出了一种基于格雷维尔迭代的高效输出权值求解方法,以提高网络建模效率。

本文使用一个数值模拟实例、四个真实数据集以及一个自主研发的手势识别(Gesture recognition, GR)系统来评价所提学习模型。实验结果表明,与现有IRWNNs模型相比,SGA-SIM在建模速度、模型精度和模型网络拓扑结构等方面有明显优势。

本文的内容安排如下:第1节利用空间几何知识对扁平神经网络随机增量学习过程进行描述和分

析; 第 2 节对所提的空间几何角度最大化的随机增量学习模型进行描述; 第 3、4 节展示所提模型的性能评估与应用; 第 5 节对本文进行总结。

1 扁平神经网络随机增量学习

1.1 随机增量学习过程

给定含 n 个不同样本的数据集 $\{x, f\}$, 其中 $x = \{x_1, x_2, \dots, x_n\}$ 为输入, $x_i = [x_{i1}, x_{i2}, \dots, x_{id}]^T \in \mathbf{R}^d$, $f = \{f_1, f_2, \dots, f_n\}$ 为输出, $f_i = [f_{i1}, f_{i2}, \dots, f_{im}]^T \in \mathbf{R}^m$, $i = 1, 2, \dots, n$. d 表示输入的维数, m 表示输出维数. 假设已经构建了带有 $L-1$ 个隐含层节点的扁平神经网络随机学习模型, 即

$$f_{L-1}(x) = \sum_{j=1}^{L-1} \beta_j g_j(w_j^T x + b_j) \quad (1)$$

其中, w_j , b_j 和 β_j 分别表示第 j 个隐含层节点的输入权值、偏置和输出权值. T 表示矩阵转置, $j = 1, 2, \dots, L-1$. $g_j(\cdot)$ 表示第 j 个隐含层节点的激活函数, 其通过最大化 $\langle e_{j-1}, g_j \rangle^2 / \|g_j\|^2$ 获得. 输出权值 β_j 利用下式获得

$$\beta_j = \frac{\langle e_{j-1}, g_j \rangle}{\|g_j\|^2} \quad (2)$$

利用下式计算当前网络残差

$$e_{L-1} = f - f_{L-1} \quad (3)$$

如果残差 e_{L-1} 不满足期望容差要求, 那么将继续向网络中添加新节点, 并计算该节点所对应的网络输出权值. 重复上述步骤直到残差满足期望容差要求或隐含层节点数达到预设最大值为止.

注 1. 随机增量学习过程中, 期望容差和隐含层节点数分别表征模型精度和模型紧致性 (模型大小), 它们的预设值基于实际建模任务需求而设置. 通常情况下以期望容差为目标, 通过不断扩充隐含层节点数使所建模型尽可能满足期望容差的要求, 整个过程无需人工干涉. 然而, 硬件资源有限性以及建模任务复杂性可能会使得模型隐含层节点数达到最大时仍不满足期望容差, 此时需要更换映射能力更强的激活函数或增加隐含层来提高模型的表达能力.

1.2 随机增量学习特性分析

扁平神经网络随机增量学习模型理论结果可以被描述为:

定理 1^[22]. 假设 $\text{span}(\Gamma)$ 在 L_2 空间中是稠密的, $\forall g_L \in \Gamma$, 对于 $b_g \in \mathbf{R}^+$ 有 $0 < \|g_L\| < b_g$. g_L 通过最

大化 $\langle e_{L-1}, g_L \rangle^2 / \|g_L\|^2$ 获得, 输出权值 β_L 通过 $\beta_L = \langle e_{L-1}, g_L \rangle / \|g_L\|^2$ 计算得到.

由第 1.1 节描述可知, 若 f_{L-1} 没有满足期望容差, 那么需要继续添加节点, 构建模型 $f_L = f_{L-1} + \beta_L g_L$ 逼近期望容差. 但从空间几何视角分析该过程可知, 当 $\beta_L = \langle e_{L-1}, g_L \rangle / \|g_L\|^2$ 时, $e_L = f - (f_{L-1} + \beta_L g_L)$. 因此, e_L 与 g_L 满足如下空间几何特性关系

$$\begin{aligned} \langle e_L, g_L \rangle &= \langle e_{L-1} - \beta_L g_L, g_L \rangle = \\ &\langle e_{L-1}, g_L \rangle - \beta_L \langle g_L, g_L \rangle = \\ &\langle e_{L-1}, g_L \rangle - \frac{\langle e_{L-1}, g_L \rangle}{\|g_L\|^2} \langle g_L, g_L \rangle = 0 \quad (4) \end{aligned}$$

故, $e_L \perp g_L$.

即, e_{L-1} 与 g_L 之间的夹角 $\angle(e_{L-1}, g_L)$ 满足 $0 < \angle(e_{L-1}, g_L) < \pi/2$. 由于 β_L 是一个数值, 因此, e_{L-1} , e_L 和 g_L 三者的关系可利用图 1 近似表示. 通过观察图 1 可知, 图 1 是以 e_{L-1} 为半径构成的 $1/4$ 圆, 因此 e_L 和 g_L 的解存在多样性. 而利用最大化 $\langle e_{L-1}, g_L \rangle^2 / \|g_L\|^2$ 的方法来获取 g_L , 在一定程度上避免了对节点构造的盲目性, 但其随机的本质依然缺少合理的解释性, 容易导致随机增量学习模型建模过程耗时和网络结构不紧致等问题.

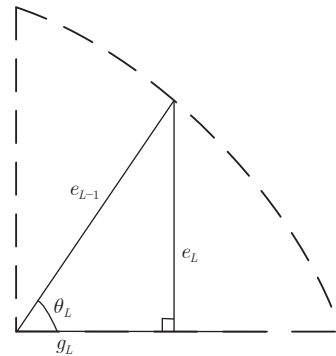


图 1 e_{L-1} , e_L 和 g_L 关系示意图

Fig. 1 Relationship diagram of e_{L-1} , e_L and g_L

根据空间几何知识, 最大化 $\langle e_{L-1}, g_L \rangle^2 / \|g_L\|^2$ 可以转为如下形式

$$\begin{aligned} \Delta_{\max} &= \frac{\langle e_{L-1}, g_L \rangle^2}{\|g_L\|^2} = \\ &\frac{\|e_{L-1}\|^2 \|g_L\|^2 \cos^2 \angle(e_{L-1}, g_L)}{\|g_L\|^2} = \\ &\|e_{L-1}\|^2 \cos^2 \angle(e_{L-1}, g_L) \quad (5) \end{aligned}$$

当 $L \rightarrow \infty$ 时, 若要式 (5) 满足最大值, 即从备选池中选择使 $\cos \angle(e_{L-1}, g_L)$ 最大的 g_L . 此外有

$$\begin{aligned} \|e_L\|^2 - \|e_{L-1}\|^2 &= \\ \|e_{L-1}\|^2 \sin^2 \theta_L - \|e_{L-1}\|^2 &= \\ \|e_{L-1}\|^2 (\sin^2 \theta_L - 1) &< 0 \end{aligned} \quad (6)$$

其中, $\theta_L = \angle(e_{L-1}, g_L)$, 故 $\{\|e_L\|^2\}$ 单调递减.

当 $L \rightarrow \infty$ 时, 结合式 (6) 有

$$\frac{\|e_{L-1}\|^2 - \|e_L\|^2}{\|e_{L-1}\|^2} = 1 - \sin^2 \theta_L < 1 \quad (7)$$

令 $1 - \sin^2 \theta_L = \varepsilon_L$, 易得

$$\lim_{L \rightarrow \infty} \prod_{k=1}^L (1 - \varepsilon_k) = \varepsilon > 0 \quad (8)$$

由式 (7) 和式 (8) 可知

$$\begin{aligned} \|e_L\|^2 &= (1 - \varepsilon_L) \|e_{L-1}\|^2 = \\ \prod_{k=1}^L (1 - \varepsilon_k) \|f\|^2 &\end{aligned} \quad (9)$$

因此, 有

$$\lim_{L \rightarrow \infty} \|e_L\|^2 = \lim_{L \rightarrow \infty} \prod_{k=1}^L (1 - \varepsilon_k) \|f\|^2 = \varepsilon \|f\|^2 \quad (10)$$

由上述分析可知

$$\lim_{L \rightarrow \infty} \|f - f_L\| = \sqrt{\varepsilon} \|f\| \quad (11)$$

由此可见, 当 RWNNs 使用式 (5) 选择隐含层参数时难以保证网络无限逼近性.

2 空间角度最大化约束的随机增量学习模型

本文从空间几何视角, 针对扁平神经网络随机增量学习模型建模过程的固有问题, 提出一种空间几何角度最大化随机增量学习模型. 其通过建立空间几何角度最大化约束, 保证了 SGA-SIM 的无限逼近特性; 同时根据格雷维尔迭代理论, 建立了高效的网络输出权值计算方法, 提高了模型学习效率.

2.1 空间几何角度最大化约束

给定一个目标函数 $f: \mathbf{R}^d \rightarrow \mathbf{R}^m$, 假设已经构建好带有 $L-1$ 个隐含层节点的 SGA-SIM. 即, $f_{L-1}(x) = \sum_{j=1}^{L-1} \beta_j g_j(w_j^T x + b_j)$, $L = 1, 2, \dots$, $f_0 = \mathbf{0}$. $e_{L-1} = f - f_{L-1} = [e_{L-1,1}, e_{L-1,2}, \dots, e_{L-1,m}]$. 若 e_{L-1} 不满足期望容差, 将利用空间几何角度最大化约束产生新的隐含层节点 $g_L(w_L$ 和 $b_L)$ 计算输出权值 β_L , 进而让 $f_L = f_{L-1} + \beta_L g_L$ 改善容差, 重复上述过程直到满足目标要求为止.

SGA-SIM 理论描述如下:

定理 2. 假设 $\text{span}(\Gamma)$ 在 L_2 空间中是稠密的, $\forall g_L \in \Gamma$, 对于 $b_g \in \mathbf{R}^+$ 有 $0 < \|g_L\| < b_g$. 给定一个任意小的 σ . 若激活函数 g_L 和残差 e_{L-1} 满足如下空间几何角度最大化约束

$$\begin{cases} 1 - \sin \theta_L > \sigma \\ (\theta_L)_{\max} \end{cases} \quad (12)$$

其中, $(\theta_L)_{\max} = (\angle(e_{L-1}, g_L))_{\max}$, 输出权值利用如下方法计算

$$\beta_L = \frac{\langle e_{L-1}, g_L \rangle}{\|g_L\|^2} \quad (13)$$

那么, $\lim_{L \rightarrow \infty} \|f - f_L\| = 0$.

证明. 基于上述分析易知, 当 $L \rightarrow \infty$ 时, $\|e_L\|$ 单调递减且收敛. 故 $\exists \gamma \geq 0$, $\lim_{L \rightarrow \infty} \|e_L\| = \gamma$. 即, $\forall \delta > 0$, $\exists N \in \mathbf{N}^+$, $\forall L > N$ 时, 有 $\gamma \leq \|e_L\| \leq \gamma + \delta$.

因此, 当 $\forall L > N$ 时, 一定存在子序列 $\{e_{L+t}\}$, $t = 1, 2, \dots$, 其极限为 \tilde{e} , 且 $\|\tilde{e}\| = \lim_{t \rightarrow \infty} \|e_{L+t}\| = \gamma$. 结合空间几何特性可知

$$\|e_{L+t}\| < \frac{\|\tilde{e}\|}{\sin(\theta_{L+t+1})} = \frac{\gamma}{\sin(\theta_{L+t+1})} \quad (14)$$

结合极限定义: \exists 正整数 N_1 , 当 $t \geq N_1$ 时

$$\|e_{L+t}\| < \frac{\gamma}{1 - \sigma} \quad (15)$$

结合空间几何特性、式 (12) 和式 (15) 可知

$$\begin{aligned} \|e_{L+t}\| &= \|e_{L+(t-1)}\| \sin(\theta_{L+t}) < \\ \frac{\gamma}{1 - \sigma} \sin(\theta_{L+t}) &< \gamma \end{aligned} \quad (16)$$

由于序列 $\{e_{L+t}\}$ 是序列 $\{e_L\}$, $\gamma \leq \|e_L\| \leq \gamma + \delta$ 的一个子序列且单调递减, 因此有 $\lim_{t \rightarrow \infty} \|e_{L+t}\| = \|\tilde{e}\| = \gamma$. 但通过式 (14) ~ 式 (16) 的证明可知, $\|e_{L+t}\| < \gamma$. 因此, 出现 $\|e_{L+t}\| < \gamma$ 和 $\gamma \leq \|e_L\| \leq \gamma + \delta$ 相矛盾的结论.

假设 e^* 是序列 $\{e_L\}$ 的极限, 如果 $\|e^*\| = \gamma \neq 0$, 那么一定存在一个 g^* 不垂直于 e^* . 进而根据图 1 可知, 一定存在 $\|e_{L+t}\| < \lim_{L \rightarrow \infty} \|e_L\| = \gamma$. 然而, 这与 $\gamma \leq \|e_L\| \leq \gamma + \delta$ 相矛盾, 故 $\|e^*\| = \gamma$ 必须等于零.

综上, 结合式 (16) 与 $\gamma \leq \|e_L\| \leq \gamma + \delta$ 相互矛盾可知 $\gamma = 0$. 即 $\lim_{L \rightarrow \infty} \|f - f_L\| = \lim_{L \rightarrow \infty} \|e_L\| = 0$. \square

在实际应用中, 定理 2 首先使用节点池策略生成 T_{\max} 个候选隐含层节点; 进而 $1 - \sin \theta_L > \sigma$ 被用来选择满足收敛条件的节点, 根据 $1 - \sin^2 \theta_L = \varepsilon_L$ 和空间几何特性可知, θ_L 对模型无限逼近性有着至关重要的影响, 其越大越有助于收敛. 因此, $(\theta_L)_{\max}$ 能够从满足条件的节点中选择最有利于收敛的节

点. 此外, 若在节点池中找不到满足条件的节点, 那么节点参数随机区间 $[-\lambda, \lambda]^d$ 和 $[-\lambda, \lambda]$ 以步长 $\Delta\lambda$ 从 λ_{\min} 开始增加, 每次增加节点参数都会重新生成, 直到找到合适参数或到达 λ_{\max} 为止. 同时, SGA-SIM 构建模型时, 本文利用 early stop 机制通过验证集对模型泛化性进行监控以防止模型过拟合.

注 2. 定理 2 提供了一种与最大化 $\langle e_{L-1}, g_L \rangle^2 / \|g_L\|^2$ 不同的构造方法, 该方法利用空间几何知识, 从备选池 (T_{\max} 个节点) 中直接选取加快收敛的隐含层节点, 并保证了学习模型的无限逼近性. 此外, 对比式 (5) 与式 (12) 可知, 两种构造方法在复杂度上的区别主要体现在计算 $\|e_{L-1}\|^2$ 和 $\arccos()$ 函数. 通过数学分析可知, 相比于 $\|e_{L-1}\|^2$, $\arccos()$ 函数的计算复杂度非常小. 同时, 由于 $0 < (\theta_L)_{\max} < \pi/2$, 故 $1 - \sin(\theta_L)_{\max} > \sigma$ 这一条件极容易满足. 因此, 本文所提 SGA-SIM 具有建模复杂度更低的优势.

注 3. 本文从空间几何视角阐明了模型隐含层节点的构造对学习模型的重要性, 将神经网络“黑箱”构造过程进行可视化解析, 从而提出可解释性构造约束, 提高隐含层节点质量, 实现学习模型的快速收敛, 提高模型的紧致性.

2.2 权值计算方法

利用定理 2 在求解输出权值时, 只求解与新增节点相关输出权值而保持之前的输出权值不变. 由文献 [23] 结果可知, 这种局部权值更新方法, 从整个模型角度看, 输出权值并非最优值, 因此势必会导致模型构建过程收敛速度缓慢. 为此, 本文采用输出权值全局更新算法

$$\beta = \arg \min_{\beta} \left\| f - \sum_{j=1}^L \beta_j g_j \right\| \quad (17)$$

矩阵形式表示如下

$$H\beta = f \quad (18)$$

其中, $H = [g_1, g_2, \dots, g_L]$ 表示当前隐含层输出矩阵, L 表示当前隐含层节点数, g_j 表示第 j 个节点对应的隐含层向量, $j = 1, 2, \dots, L$.

进一步可得

$$\beta = H^\dagger f \quad (19)$$

其中, \dagger 表示矩阵的 Moore-Penrose 广义逆^[31]. 由于 H 很难满足存在逆矩阵的条件, 因此本文使用 H 的 Moore-Penrose 广义逆来近似表示其逆矩阵.

显然根据式 (19) 计算的输出权值全局最优解优于式 (13) 输出权值局部最优解, 但每次添加节点都需要求解 Moore-Penrose 广义逆, 而广义逆通常

需要使用高复杂度的奇异值分解求解^[32], 这将会极大地增加 SGA-SIM 的建模复杂度. 因此, SGA-SIM 在进行需要大量隐含层节点或输入数据规模较大的建模任务时存在效率低下的问题.

为解决这一问题, 本文利用格雷维尔迭代方法提出了一种高效 (低复杂度) 输出权值计算方法^[33], 具体描述如下.

假设 $H_{L-1} = [g_1, g_2, \dots, g_{L-1}]$ 是未添加新节点之前的隐含层输出矩阵, 则新添加节点之后的隐含层输出矩阵可表示为 $H_L = [H_{L-1}|g_L]$, 根据格雷维尔迭代理论, H_L^\dagger 可以利用下式求得

$$H_L^\dagger = \begin{bmatrix} H_{L-1}^\dagger - d_L b_L^T \\ b_L^T \end{bmatrix} \quad (20)$$

其中, d_L, b_L^T 为辅助变量并满足如下关系

$$b_L^T = \begin{cases} (c_L)^\dagger, & c_L \neq \mathbf{0} \\ (1 + d_L^T d_L)^{-1} d_L^T H_{L-1}^\dagger, & c_L = \mathbf{0} \end{cases}$$

$$d_L = H_{L-1}^\dagger g_L$$

$$c_L = g_L - H_{L-1} d_L$$

结合式 (19), 输出权值可以利用如下方法计算

$$\beta = \begin{bmatrix} \beta_{\text{previous}} - d_L b_L^T f \\ b_L^T f \end{bmatrix} \quad (21)$$

其中, $\beta_{\text{previous}} = [\beta_1, \beta_2, \dots, \beta_{L-1}]$ 表示未添加新节点之前网络的最优输出权值; $\beta = [\beta_1, \beta_2, \dots, \beta_L]$ 表示当前网络最优输出权值. 此方法能够在原有输出权值的基础上直接计算新添加节点后模型的输出权值, 同时这种方法只涉及简单运算, 因此具有计算简单、复杂度低等优势.

由于本文基于式 (19) 和式 (21) 构建了两种不同输出权值计算方法, 因此 SGA-SIM 包括两种不同的算法实现, 即 SGA-SIM-I 和 SGA-SIM-II, 具体如算法 1 所示.

算法 1. SGA-SIM 算法

初始化:

1) 给定训练集 $\{x, f\}$, $x = \{x_1, x_2, \dots, x_n\}$, $x_i \in \mathbf{R}^d$, $f = \{f_1, f_2, \dots, f_n\}$, $f_i \in \mathbf{R}^m$, T_{\max} 为最大的随机分配次数, L_{\max} 为最大的隐含层节点数, ℓ 为期望容差, $\Upsilon = \{\lambda_{\min} : \Delta\lambda : \lambda_{\max}\}$ 为一组正的标量, $e_0 = f$, Ω 和 W 为空集.

网络训练:

2) **while** $L \leq L_{\max}$ 和 $\|e_0\| > \ell$ 时, **do**

3) **for** $\lambda \in \{\lambda_{\min} : \Delta\lambda : \lambda_{\max}\}$ **do**

4) **for** $s = 1, 2, \dots, T_{\max}$ **do**

5) 根据式 (12) 从区间 $[-\lambda, \lambda]^d$ 和 $[-\lambda, \lambda]$ 中产生隐含层节点 w_L, b_L 以构建节点池, 并将参数存储于 W 内;

- 6) 计算 $\tau_L = \sum_{q=1}^m (\theta_L)_{\max_{L,q}}$ 存储于 Ω 内;
 7) **end for**
 8) 若 Ω 不为空, 在 Ω 中找到 τ_L 最大时对应的隐含层参数 w_L^* 和 b_L^* . $H_L = [g_1, g_2, \dots, g_L]$ 若为空, 则返回 3);
 9) **break** (返回 2));
 10) **end for**
 11) 计算输出权值 β : 通过式 (19) (对应 SGA-SIM-I) 或式 (21) (对应 SGA-SIM-II) 计算;
 12) 更新 $e_0 = e_L$, $L = L + 1$;
 13) **end while**
 14) β_1, \dots, β_L , $w^* = \{w_1^*, \dots, w_L^*\}$, $b^* = \{b_1^*, \dots, b_L^*\}$.

3 性能评估与应用

本节将展示 SGA-SIM 与 IRWNNs 和 SCNs 在 5 种数据集和一个自主研发的手势识别系统上的对比结果. 其中, 数据集主要包括 1 个数值模拟例子、2 个实际回归应用例子和 2 个实际分类应用例子. 手势识别系统来源于课题组和计算机学院共同研究的手势识别课题. 该系统包括硬件和软件两部分, 其中硬件部分主要作用是手势数据采集, 软件部分主要是数据处理、模型建立与识别.

3.1 实验数据与设置

实验中数值模拟例子数据集是利用具有高非线性单输入单输出特性的式 (22) 随机生成. 4 种实际应用数据集分别来自 UCI 数据库^[34] 和 KEEL 数据库^[35]. 它们分别是 Abalone、Compactiv、Iris 和 HAR (Human activity recognition) 数据集. 公共数据集和自建手势识别数据集都被随机分成训练集和测试集. 具体的详细信息被展示在表 1.

$$f(x) = 0.2e^{-(10x-4)^2} + 0.5e^{-(80x-40)^2} + 0.3e^{-(80x-20)^2} \quad (22)$$

上式中, 输入 x 和输出 $f(x)$ 都被归一化到 $[0, 1]$.

本文的所有实验都是利用 Matlab 2019b, 在 CPU 为 E3-1225 3.3 GHz、内存为 32 GB RAM 的 Think Server 上进行. Sigmoid 函数被选为所有实验的激活函数, 各模型在不同数据集上的初始参

表 1 数据集信息

Table 1 Information of datasets

	数据集	训练样本数	测试样本数	特征	类别
	nonlinear function	600	400	1	—
回归问题	Abalone	2000	2177	7	—
	Compactiv	6144	2048	21	—
	Iris	120	30	4	3
分类问题	HAR	7352	2947	561	6
	Gesture recognition	3595	1241	54	24

数详见表 2. 此外, 本节中所展示的实验结果都是运行 30 次后的统计值.

3.2 性能分析

本文首先使用数值模拟例子对 SGA-SIM-I、SGA-SIM-II、IRWNNs 以及 SCNs 四者进行比较, 并以隐含层节点数、建模时间和均方根误差 (RMSE) 的平均值和标准差为评判标准对实验结果进行分析 (由于 BLS 与其他模型结构不同, 故这里不做比较). 具体实验结果如表 3 所示, 其中 L 表示隐含层节点数, AVE 和 DEV 分别表示模型测试 RMSE 的平均值和标准差.

从表 3 可以看出, 在保证相同的 stop RMSE 时, 本文所提 SGA-SIM-I、SGA-SIM-II 和 SCNs 三者需要较少的隐含层节点, 这表明本文所提模型具有结构紧凑的优势. 此外, 上述三者在训练时间上也具有明显的优势, 特别是 SGA-SIM-II. SGA-SIM-II 相比于 SGA-SIM-I、IRWNNs 和 SCNs 分别降低了 94.44%、95.39% 和 94.88% 的训练时间. 综上, 随机增量构建模型时对隐含层参数生成过程进行约束有助于提高模型紧致性和降低建模时间. 此外, 在 SGA-SIM-I、SGA-SIM-II 和 SCNs 所用隐含层节点数极为接近的情况下, SGA-SIM-II 在建模时间上具有如此巨大的优势, 这也直接说明本文高效输出权值计算方法的有效性. 而在 RMSE 的 AVE 和 DEV 方面, 本文所提 SGA-SIM-I 和 SGA-SIM-II 都远小于 IRWNNs. 因此, 说明本文所建模型的

表 2 各模型在不同数据集上的初始参数

Table 2 Initial parameters of each model on different datasets

数据集 (期望残差 ℓ)	IRWNNs (L_{\max} , λ , T_{\max})	SGA-SIM (L_{\max} , Υ , T_{\max})	SCNs (L_{\max} , Υ , T_{\max})
nonlinear function (0.05)	100, 150, 1	100, 150:10:200, 20	100, 150:10:200, 20
Abalone (0.16)	100, 0.5, 1	100, 0.5:0.1:10, 20	100, 0.5:0.1:10, 20
Compactiv (0.15)	200, 0.5, 1	200, 0.5:0.1:10, 20	200, 0.5:0.1:10, 20
Iris (0.01)	50, 1, 1	50, 1:1:10, 20	50, 1:1:10, 20
HAR (0.01)	500, 50, 1	500, 1:1:10, 20	500, 1:1:10, 20
Gesture recognition (0.05)	500, 0.5, 1	500, 0.5:0.5:10, 20	500, 0.5:0.5:10, 20

表 3 数值模拟例子的实验结果

Table 3 Experimental results of numerical simulation examples

模型	节点数 (L)	建模时间 (s)	AVE	DEV
IRWNNs	100.0	3.25	0.1060	0.0301
SGA-SIM-I	79.8	2.70	0.0014	0.0003
SGA-SIM-II	79.2	0.15	0.0010	0.0002
SCNs	79.3	2.93	0.0014	0.0003

表达能力更好, 且更加稳定.

特别地, SCNs 虽然在性能上与 SGA-SIM 相似, 但不等式约束很难解释其为什么会有良好的性能. 而 SGA-SIM 通过空间几何可视化模型性能与节点参数之间的关系增强了模型的可解释性, 揭示了 RWNNs 的部分黑盒子性质. 这是由于 SCNs 中用于生成隐含层参数的不等式约束 ($\langle e_{L-1, q}, g_L \rangle \geq b_g^2 \delta_{L, q}, q = 1, 2, \dots, m$)^[23] 本质上是一种以残差和激活函数等为条件的数据驱动方法. 而数据驱动方法像一个黑盒子, 很难对参数随机生成的机理进行解释. 即, SCNs 难以从机理方面回答为什么选择这些节点参数构建模型. 不同于 SCNs, SGA-SIM 使用的空间几何角度最大化约束 (式 (12)) 是一种数据驱动与机理相融合的方法, 其中, $1 - \sin \theta_L > \sigma$ 是一种用于选择合适节点参数的数据驱动条件, 而 $(\theta_L = \angle(e_{L-1}, g_L))_{\max}$ 则是通过刻画添加节点前的残差与激活函数之间的机理关系来选择能够快速收敛的节点参数进行模型构建.

通过上述实验对比结果可知, 本文提出的具有可解释性的空间几何角度最大化节点约束方法有助于减少建模任务对隐含层节点的需求 (模型结构更紧致) 及改善模型表达能力, 同时, 高效输出权值计算方法有助于降低模型建模时间, 特别适合处理复杂网络结构建模和大数据建模任务.

为了比较 IRWNNs、SCNs、SGA-SIM-I 和 SGA-SIM-II 四者的函数逼近能力, 本文将 stop RMSE 设置为 0.1 后绘制了详细的收敛曲线和拟合曲线, 如图 2、图 3 所示, 它们分别显示了 RMSE 随着节点增加的变化趋势和相应的模型逼近能力变化趋势. 从图 2 可以看出, 相比于 IRWNNs, 本文所提 SGA-SIM 和 SCNs 优势明显. SGA-SIM-I、SGA-SIM-II、SCNs 分别使用 18、19、18 个隐含层节点便达到了期望的 RMSE. 而 IRWNNs 在用尽所有节点后才接近期望的 RMSE. 同时, IRWNNs 在隐含层节点数达到 37 个后, 很难通过增加更多节点来提高 IRWNNs 收敛性. 这也说明本文所提空间几何角度最大化节点约束方法有助于选择更加优质的隐含层节点. 从图 3 可以看出, 相比于 IRWNNs,

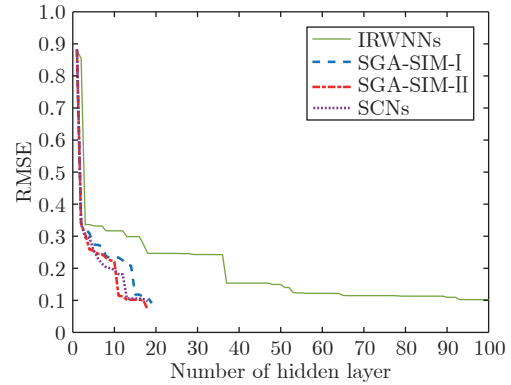


图 2 4 种模型的收敛曲线图

Fig. 2 Convergence curves of four models

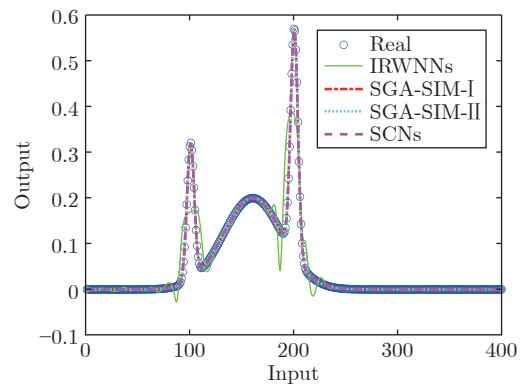


图 3 4 种模型的拟合曲线图

Fig. 3 Fitting curves of four models

SGA-SIM 和 SCNs 的函数拟合能力更强.

对于所有机器学习模型来说, 如果建模过程中过度学习含噪数据, 将容易导致模型过拟合, 使得泛化性能变差, 本文所提 SGA-SIM 也不例外. 为阐明这一问题, 本文给出在 stock 数据集上的实验结果, 如图 4 所示, 其中 L_{\max} 设置为 200, T_{\max} 设置为 20, $\Upsilon = \{0.5:0.1:10\}$, 期望容差设置为 0.03, 训练集人为加入均值为 0、方差为 0.5 的噪音数据. 由图 4 可知, 学习模型始终没有达到期望的容差, 训练在节点增加到最大 200 时结束. 通过观察可知, 学习模型在 155 之前训练误差和测试误差均随着节点增加而降低, 155 个节点之后, 测试误差开始增加, 说明出现了过拟合. 这说明期望容差的设置较低, 导致模型因过度学习出现过拟合. 为此, 在实际建模过程中往往需要引入 early stop 机制, 使 SGA-SIM 模型在学习过程中一旦过拟合即停止训练. 此外, 期望容差应基于实际建模任务需求而设置. 假设将期望容差设置为 0.05, 根据图 4 可知, SGA-SIM 在 116 个节点时即可达到期望容差, 此时模型没有出现过拟合. 进一步, 若期望容差设置小于 0.044619,

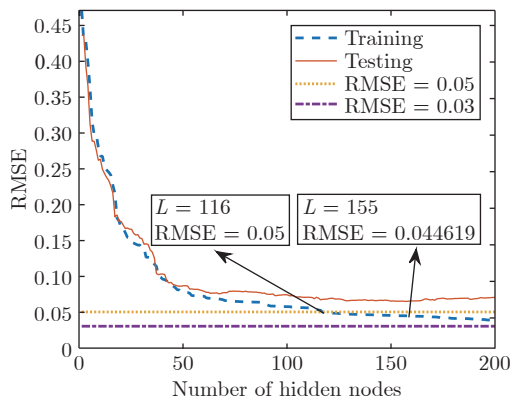


图 4 节点对拟合性能的影响

Fig.4 Effect of nodes on fitting performance

那么当所建模型的 RMSE 达到 0.044619 时, early stop 机制将会被触发并立刻停止训练.

进一步地, 本文在 Abalone、Compactiv、Iris 和 HAR 4 个公共的实际数据集上进行了实验探究. 实验结果如表 4 所示. 基于这 4 个数据集本文使用隐含层节点、建模时间、训练误差和测试误差对 SGA-SIM-I、SGA-SIM-II、IRWNNs 以及 SCNs 四者进行评估. 对于 Abalone 数据集, 在使用 100 个隐含层节点进行建模后, SGA-SIM 和 SCNs 的训练误差和测试误差都远小于 IRWNNs. 上述实验结果证明了节点参数约束存在的意义. 但约束也导致在使用相同的隐含层节点数时, SGA-SIM-I 和 SCNs 的建模时间高于 IRWNNs. 而高效输出权值计算方法的提出使得 SGA-SIM-II 相比于 SGA-SIM-I、IRWNNs 和 SCNs 分别减少了 75.27%、20.41% 和 73.05% 的建模时间. 因此, 这种方法能够有效避免由空间几何角度最大化节点约束方法和不等式约束带来的建模耗时问题. 该方法极大地提高了 SGA-SIM-II 的建模速度和实际应用能力. 特别地, SGA-SIM-I 的建模时间低于 SCNs, 这说明本文所提的可解释性的空间几何角度最大化节点约束方法比不等式约束方法更轻量.

对于 Compactiv 数据集, IRWNNs 实现了最小的建模时间, 这主要是因为 SGA-SIM 和 SCNs 需要进行节点参数的选择. 而通过表 1 可知, Compactiv 的样本数和特征数都远远大于 Abalone, 因此参数约束方法计算较为耗时. 故出现 SGA-SIM 和 SCNs 建模时间高于 IRWNNs 的情况. 但 SGA-SIM 和 SCNs 在训练误差和测试误差上具有明显优势. 对于 Iris 数据集, IRWNNs 实现了最小的建模时间. 这主要是因为当隐含层节点较少时, 高效输出权值计算方法的优势很难体现. 但是, SGA-SIM 具有相同的最低训练误差和测试误差. 为了验证本

表 4 公共数据集的实验结果

Table 4 Experimental results of public datasets

数据集	模型	节点数 (L)	建模时间 (s)	训练误差	测试误差
Abalone	IRWNNs		0.2543	0.2209	0.2178
	SGA-SIM-I	100	0.8190	0.1479	0.1763
	SGA-SIM-II		0.2024	0.1446	0.1727
	SCNs		0.8876	0.1477	0.1723
Compactiv	IRWNNs		1.3461	0.2720	0.2715
	SGA-SIM-I	200	5.3443	0.0573	0.0695
	SGA-SIM-II		1.8827	0.0571	0.0670
	SCNs		5.7920	0.0573	0.0695
Iris	IRWNNs		0.0169	0.0222	0.0556
	SGA-SIM-I	50	0.0794	0.0167	0.0333
	SGA-SIM-II		0.0616	0.0167	0.0333
	SCNs		0.0918	0.0173	0.0333
HAR	IRWNNs		50.0381	0.0739	0.1233
	SGA-SIM-I	500	110.7238	0.0147	0.0450
	SGA-SIM-II		27.4965	0.0140	0.0441
	SCNs		111.3737	0.0160	0.0550

文所提 SGA-SIM 在处理包含大样本和多特征的数据集上的性能, 本文选择 HAR 数据集进行实验探究. 结果表明, 相比于 SGA-SIM-I、IRWNNs 和 SCNs, SGA-SIM-II 在建模时间上分别减少了 75.15%、45.05% 和 75.31%. 同时 SGA-SIM-I 和 SGA-SIM-II 实现了最低的相似训练误差和测试误差. 综上, 本文所提 SGA-SIM 对于处理大数据任务和需要隐含层多的复杂建模任务具有极大优势, 特别是 SGA-SIM-II.

4 手势识别应用

手势识别已经被成功应用于虚拟现实、健康监测和智能家居等多个领域^[36-37]. 本文开发的基于智能手套的手势识别系统如图 5 所示. 在手势识别系统中共包括硬件和软件两部分. 硬件部分主要作用是手势数据采集. 在这部分设计了一款集成加速度传感器、陀螺仪传感器和弯曲传感器等众多传感器的新型智能手套, 并用其捕捉手部的细微动态轨迹变化以及上肢部分的全部姿态信息. 采集到的数据可以通过 WI-FI、蓝牙和 USB 等通信方式传输到软件部分. 软件部分由数据预处理模块和手势识别模块组成. 在数据预处理模块, 采集的手势数据需要经历平滑去噪、滑动分割、归一化处理 and 特征提取等步骤. 手势识别模块包括建立模型和手势识别. 在这一模块上, 本文使用所提的 SGA-SIM 建立 GR 模型并进行手势识别.

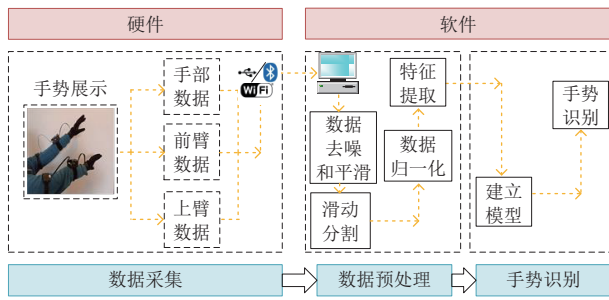


图 5 手势识别系统框架图

Fig.5 Frame diagram of gesture recognition system

4.1 硬件部分

本文设计的智能手套主要由电池组管理、组件访问、信号处理、存储模块和通信处理模块等 5 部分组成。具体结构如图 6 所示。电池组管理的作用是提供能够支持数据采集、处理和传输的能量。信号处理模块是对手势信号进行滤波和分析。存储模块主要是程序存储区和数据缓存区。通信处理模块的作用是与计算机之间进行数据传输, 主要包括 WI-FI、蓝牙和 USB 等方法。本系统所采集的数据为手指的三维弯曲数据, 手掌、前臂和上臂的姿态欧拉角和四元数数据。

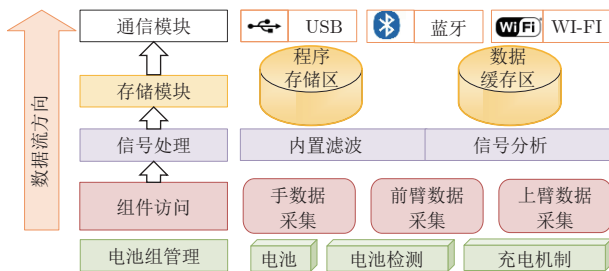


图 6 智能手套框架图

Fig.6 Frame diagram of smart gloves

在智能手套中, 本文将两个弯曲传感器进行集成, 进而设计一个三维手指传感器来测量每个手指的弯曲程度和相邻两个手指的张开角度。利用陀螺

仪传感器采集姿态欧拉角和手掌四元数数据。惯性传感器被用来采集前臂和上臂的姿态信息和运动变化数据。从而将整个上肢的运动和姿态数据进行采集。每个惯性传感器由三个陀螺仪传感器、三个加速度传感器和一个可扩展的数字运动处理器组成。运动传感器可以消除陀螺仪与加速度计的轴间差异, 并直接输出姿态欧拉角和四元数。智能手套中传感器的详细信息被展示在表 5。

4.2 软件部分

本部分设计了一款基于智能手套的手势识别软件。该软件的主要模块包括: 数据接收、特征提取和手势建模与识别。其中, 手势建模与识别最重要, 这里使用本文所提算法进行建模和识别。具体结构如图 7 所示。在特征提取步骤中, 本文使用 CNN 从手势数据中提取特征, 这主要是因为 CNN 的参数共享和稀疏连接等特点能够在特征提取时降低数据的冗余性, 进而利于构建识别精度更高的模型。

对于数据采集而言, 本文所使用数据集来自 6 名志愿者 (男女各 3 名), 实验中他们佩戴智能手套完成指定的手势, 智能手套以 20 Hz 的固定频率采集详细的手势数据。然后, 经由软件部分处理得到一个含有 64 个特征、24 个类别的数据集。最后, 将其随机分成训练集和测试集。具体结果如表 6 所示。通过观察表 6 可知, IRWNNs、SCNs、SGA-SIM-I 和 SGA-SIM-II 四者在节点数上都是 500。对于建模时间而言, SGA-SIM-I 和 SCNs 的建模时间最长, 这也说明节点约束以及耗时的输出权值计算方法会极大地增加模型的建模消耗。而 SGA-SIM-II 在建模时间上相比于 IRWNNs、SGA-SIM-I 和 SCNs 有着明显的优势, 分别节省了 77.51%、86.23% 和 86.50% 的时间。进而说明 SGA-SIM-II 是一种更加适合在资源有限设备上应用的高效轻量型模型。而从识别精度方面来说, 本文所提模型有着很大的优势。但需要注意到 SGA-SIM-II 的识别精度比 SGA-SIM-I 会高一些, 这主要是因为 Moore-Penrose 广

表 5 智能手套传感器描述

Table 5 Smart gloves sensor description

传感器	描述
加速度传感器	1) 加速度传感器是一种能够测量加速度的传感器, 其能感受加速度并转换成可用输出信号; 2) 数据主要来自 x, y, z 三个轴。
陀螺仪传感器	1) 陀螺仪通过测量物体运动时的角速度来计算物体旋转的角度和方向; 2) 数据主要来自 x, y, z 三个轴。
弯曲传感器	1) 弯曲传感器通过阻值将弯曲程度数字化; 2) 弯曲传感器能够测量的弯曲范围为 $[1^\circ, 180^\circ]$ 。

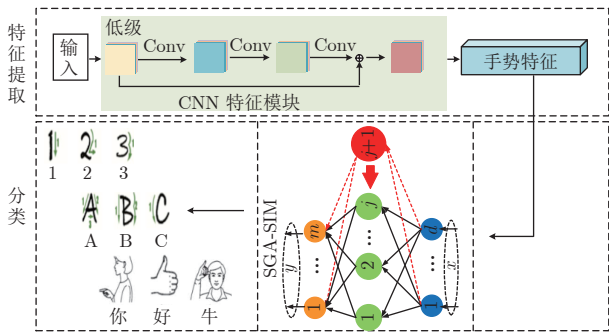


图 7 软件部分结构图

Fig.7 Software structure diagram

表 6 手势识别结果

Table 6 Gesture recognition result

算法	建模时间 (s)	测试精度	节点数
IRWNNs	67.84	81.92%	500
SGA-SIM-I	110.80	94.49%	500
SGA-SIM-II	15.26	95.19%	500
SCNs	113.08	94.49%	500

义逆是伪逆, 它只能反映逆的大部分特性, 多次使用会对最终的结果产生很大的影响. 而 SGA-SIM-II 中逐步动态更新, 仅通过上一步的逆就可以推导出现有模型的逆. 因此, 其识别精度比 SGA-SIM-I 会高一些.

除了上述指标外, 本文还使用了精确率 (Precision)、召回率 (Recall) 和 F1 值 (F1-score) 等指标评价 IRWNNs, SGA-SIM-I 和 SGA-SIM-II 在 GR 问题上的实验结果. 具体结果如图 8 ~ 图 11 所示. 通过对比 4 张图可知, IRWNNs 的数值幅度波动变化最大, 而 SGA-SIM-I, SGA-SIM-II 和 SCNs 的三种指标在数值上都比较接近 1. 说明这三种算法所建模型的表达能力更好. 从召回率上看, 四者对手

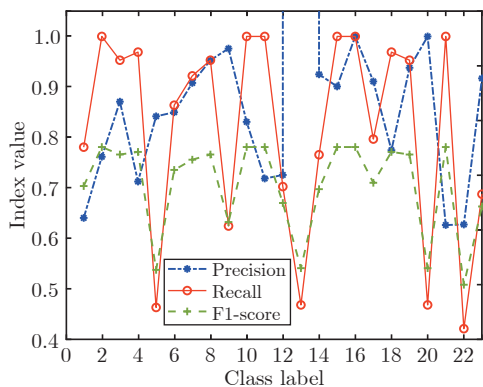


图 8 IRWNNs 的三种指标结果对比图

Fig.8 Comparison chart of three index results of IRWNNs

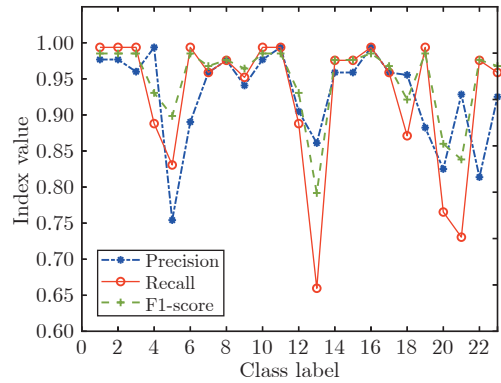


图 9 SGA-SIM-I 的三种指标结果对比图

Fig.9 Comparison chart of three index results of SGA-SIM-I

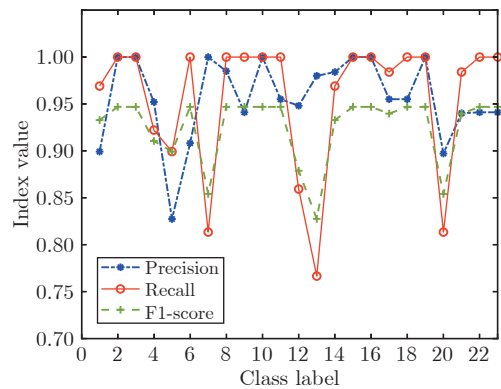


图 10 SGA-SIM-II 的三种指标结果对比图

Fig.10 Comparison chart of three index results of SGA-SIM-II

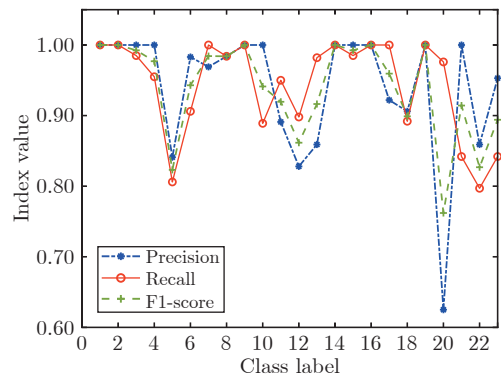


图 11 SCNs 的三种指标结果对比图

Fig.11 Comparison chart of three index results of SCNs

势 1 和手势 3 具有很低的召回率, 这是由于手势数据是一种时间序列数据, 因此数据之间的干扰性很强. 从精确率上看, SGA-SIM-II 几乎对 83.3% 的手势都能够实现大于等于 95% 的精确率. 这是 IR-

WNNs、SGA-SIM-I 和 SCNs 难以达到的。此外，SGA-SIM-II 模型的 F1 值也能够大致稳定在 0.95 附近，这表明该模型具有更好的泛化能力。

5 结论

本文针对扁平随机神经网络增量构建过程中出现的结构不紧致和参数生成可解释性不足等问题，基于空间几何和对模型增量构建过程的分析，提出了空间几何角度最大化约束随机增量学习模型。该模型利用具有可解释性的空间几何最大化方法对隐含层节点参数进行约束，以选择优质节点构建任务模型，进而减少建模时所需节点数，即模型结构紧致性更好。此外，本文还提出一种高效的输出权值计算方法，以降低建模消耗。实验探究表明，相比于 IRWNNs 和 SCNs，本文所提 SGA-SIM 在建模时间、模型精度和模型结构紧致性等方面都具有明显优势，特别是 SGA-SIM-II。然而，所提模型缺少对原始数据的多层次特征提取能力，并不是 end to end 模型，因此，在面对复杂数据分析任务时，需要与特征提取或特征选择方法联合使用，性能更优。这些不足将成为未来的研究方向。同时，如何将其应用到实际工业中也是一个很重要的研究课题。

References

- Wang X, Chen H, Gan C X, Lin H G, Dou Q, Tsougenis E, et al. Weakly supervised deep learning for whole slide lung cancer image analysis. *IEEE Transactions on Cybernetics*, 2020, **50**(9): 3950–3962
- Chai L, Du J, Liu Q F, Lee C H. A cross-entropy-guided measure (CEGM) for assessing speech recognition performance and optimizing DNN-based speech enhancement. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2021, **29**: 106–117
- Wang N, Er M J, Han M. Generalized single-hidden layer feedforward networks for regression problems. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, **26**(6): 1161–1176
- Ginanjar R R, Kim D S. Real-time SLFN-based node localization using UAV. In: Proceedings of the IEEE International Conference on Industrial Cyber Physical Systems (ICPS). Taiwan, China: IEEE, 2019. 101–106
- Cao W P, Wang X Z, Zhong M, Gao J Z. A review on neural networks with random weights. *Neurocomputing*, 2018, **275**: 278–287
- Broomhead D S, Lowe D. Multivariable functional interpolation and adaptive networks. *Complex System*, 1988, **2**: 321–355
- Igel'nik B, Pao Y H. Stochastic choice of basis functions in adaptive function approximation and the functional-link net. *IEEE Transactions on Neural Networks*, 1955, **6**(6): 1320–1329
- Pao Y H, Takefuji Y. Functional-link net computing: Theory, system architecture, and functionalities. *Computer*, 1992, **25**(5): 76–79
- Schmidt W F, Kraaijveld M A, Duijn R P W. Feedforward neural networks with random weights. In: Proceedings of the 1st IAPR International Conference on Pattern Recognition. Hague, Netherlands: IEEE, 1992. 1–4
- Husmeier D. *Neural Networks for Conditional Probability Estimation: Forecasting Beyond Point Predictions*. New York: Springer, 2012.
- Kwok T Y, Yeung D Y. Objective functions for training new hidden units in constructive neural networks. *IEEE Transactions on Neural Networks*, 1997, **8**(5): 1131–1148
- Lauret P, Fock E, Mara T A. A node pruning algorithm based on a fourier amplitude sensitivity test method. *IEEE Transactions on Neural Networks*, 2006, **17**(2): 273–293
- Ainsworth T L, Wang Y T, Lee J S. Model-based polarimetric SAR decomposition: An L_1 regularization approach. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, **60**: 1–13
- Lin M B, Ji R R, Li S J, Wang Y, Wu Y J, Huang F Y, et al. Network pruning using adaptive exemplar filters. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, **33**(12): 7357–7366
- Amari S, Murata N. Asymptotic statistical theory of overtraining and cross-validation. *IEEE Transactions on Neural Networks*, 1997, **8**(5): 985–996
- Hanson S J, Pratt L Y. Comparing biases for minimal network construction with back-propagation. In: Proceedings of the 1st International Conference on Neural Information Processing Systems. Colorado, USA: MIT Press, 1988. 177–185
- Liu D R, Chang T S, Zhang Y. A constructive algorithm for feedforward neural networks with incremental training. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 2002, **49**(12): 1876–1879
- Zhou P, Jiang Y, Wen C Y, Chai T Y. Data modeling for quality prediction using improved orthogonal incremental random vector functional-link networks. *Neurocomputing*, 2019, **365**: 1–9
- Qiu X H, Suganthan P N, Amaratunga A J G. Ensemble incremental random vector functional link network for short-term crude oil price forecasting. In: Proceedings of the IEEE Symposium Series on Computational Intelligence (SSCI). Bengaluru, India: IEEE, 2018. 1758–1763
- Li M, Wang D H. Insights into randomized algorithms for neural networks: Practical issues and common pitfalls. *Information Sciences*, 2017, **382**: 170–178
- Dudek G. A method of generating random weights and biases in feedforward neural networks with random hidden nodes. *Information Sciences*, 2019, **481**: 33–56
- Tyukin I Y, Prokhorov D V. Feasibility of random basis function approximators for modeling and control. In: Proceedings of the IEEE Control Applications, (CCA) & Intelligent Control, (ISIC). St. Petersburg, Russia: IEEE, 2009. 1391–1396
- Wang D H, Li M. Stochastic configuration networks: Fundamentals and algorithms. *IEEE Transactions on Cybernetics*, 2017, **47**(10): 3466–3479
- Li M, Huang C, Wang D H. Robust stochastic configuration networks with maximum correntropy criterion for uncertain data regression. *Information Sciences*, 2018, **473**: 73–86
- Dai Wei, Li De-Peng, Yang Chun-Yu, Ma Xiao-Ping. A model and data hybrid parallel learning method for stochastic configuration networks. *Acta Automatica Sinica*, 2021, **47**(10): 2427–2437
(代伟, 李德鹏, 杨春雨, 马小平. 一种随机配置网络的模型与数据混合并行学习方法. *自动化学报*, 2021, **47**(10): 2427–2437)
- Dai W, Li D P, Zhou P. Stochastic configuration networks with block increments for data modeling in process industries. *Information Sciences*, 2019, **484**: 367–386
- Chen C L P, Liu Z L. Broad learning system: An effective and efficient incremental learning system without the need for deep architecture. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, **29**(1): 10–24
- Chu F, Liang T, Chen C L P, Wang X, Ma X. Weighted broad learning system and its application in nonlinear industrial process modeling. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, **31**(8): 3017–3031

- 29 Feng S, Chen C L P. Fuzzy broad learning system: A novel neuro-fuzzy model for regression and classification. *IEEE Transactions on Cybernetics*, 2020, **50**(2): 414–424
- 30 Wang X H, Zhang T, Xu X M, Chen L, Xing X F, Chen C L P. EEG emotion recognition using dynamical graph convolutional neural networks and broad learning system. In: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM). Madrid, Spain: IEEE, 2018. 1240–1244
- 31 Albert A. *Regression and the Moore-Penrose Pseudoinverse*. New York: Academic Press, 1972.
- 32 Howland P, Park H. Generalizing discriminant analysis using the generalized singular value decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(8): 995–1006
- 33 Ben-Israel A, Greville T. *Generalized Inverses: Theory and Applications*. New York: John Wiley and Sons, 1974.
- 34 Anguita D, Ghio A, Oneto L, Parra X, Reyes-Ortiz J L. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In: Proceedings of the International Workshop on Ambient Assisted Living. Vitoria-Gasteiz, Spain: Springer, 2012. 216–223
- 35 Fdez J A, Fernandez A, Luengo J. KEEL data-mining software tool: Data set repository. *Integration of Algorithms and Experimental Analysis Framework, Journal of Multiple-Valued Logic & Soft Computing*, 2011, **17**(2–3): 255–287
- 36 Cheng L, Liu Y, Hou Z G. A rapid spiking neural network approach with an application on hand gesture recognition. *IEEE Transactions on Cognitive and Developmental Systems*, 2019, **99**: 151–161
- 37 Cheng H, Yang L, Liu Z C. Survey on 3D hand gesture recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 2016, **26**(9): 1659–1673



南 静 中国矿业大学信息与控制工程学院博士研究生。2021 年获得中国矿业大学硕士学位。主要研究方向为随机神经网络, 模式识别。

E-mail: jingn@cumt.edu.cn

(**NAN Jing** Ph.D. candidate at the School of Information and Control

Engineering, China University of Mining and Technology. He received his master degree from China University of Mining and Technology in 2021. His research interest covers random weight neural networks and pattern recognition.)



代 伟 中国矿业大学信息与控制工程学院教授。2015 年获得东北大学博士学位。主要研究方向为复杂工业过程建模、运行优化与控制。本文通信作者。E-mail: weidai@cumt.edu.cn

(**DAI Wei** Professor at the School

of Information and Control Engineering, China University of Mining and Technology. He received his Ph.D. degree from Northeastern University in 2015. His research interest covers modeling, operational optimization, and control for complex industrial process. Corresponding author of this paper.)



袁 冠 中国矿业大学计算机科学与技术学院教授。2012 年获得中国矿业大学博士学位。主要研究方向为数据挖掘, 软件工程。

E-mail: yuanguan@cumt.edu.cn

(**YUAN Guan** Professor at the School of Computer Science &

Technology, China University of Mining and Technology. He received his Ph.D. degree from China University of Mining and Technology in 2012. His research interest covers data mining and software engineering.)



周 平 东北大学教授。分别于 2003 年、2006 年、2013 年获得东北大学学士学位、硕士学位和博士学位。主要研究方向为工业过程运行反馈控制, 数据驱动建模与控制。

E-mail: zhouping@mail.neu.edu.cn

(**ZHOU Ping** Professor at North-

eastern University. He received his bachelor degree, master degree and Ph.D. degree from Northeastern University in 2003, 2006 and 2013, respectively. His research interest covers operation feedback control of industrial process and data-driven modeling and control.)