

面向二分类问题的直觉模糊深度随机配置网络^{*}

丁世飞^{1,2}, 朱姜兰¹, 张成龙¹, 郭丽丽^{1,2}, 张健^{1,2}

¹(中国矿业大学 计算机科学与技术学院, 江苏 徐州 221116)

²(矿山数字化教育部工程研究中心 (中国矿业大学), 江苏 徐州 221116)

通信作者: 朱姜兰, E-mail: TS21170128P31@cumt.edu.cn; 张健, E-mail: zhangjian10231209@cumt.edu.cn



摘 要: 深度随机配置网络 (deep stochastic configuration network, DSCN) 采取前馈学习方式, 基于特有的监督机制随机分配节点参数, 具有全局逼近性. 但是, 在实际场景下, 数据采集过程中潜在的离群值和噪声, 易对分类结果产生负面影响. 为提高 DSCN 解决二分类问题的性能, 基于 DSCN 引入直觉模糊数思想, 提出了一种直觉模糊深度随机配置网络 (intuitionistic fuzzy deep stochastic configuration network, IFDSCN). 与标准 DSCN 不同, IFDSCN 通过计算样本隶属度和非隶属度, 为每个样本分配一个直觉模糊数, 通过加权的方法来生成最优分类器, 以克服噪声和异常值对数据分类的负面影响. 在 8 个基准数据集上的实验结果表明, 所提出的模型与直觉模糊孪生支持向量机 (intuitionistic fuzzy twin support vector machine, IFTWSVM)、核岭回归 (kernel ridge regression, KRR)、直觉模糊核岭回归 (intuitionistic fuzzy kernel ridge regression, IFKRR)、随机函数向量链接神经网络 (random vector functional link neural network, RVFL) 和 SCN 等学习模型相比, IFDSCN 具有更好的二分类性能.

关键词: 直觉模糊数; 随机配置网络; 二分类; 数据噪声; 神经网络

中图法分类号: TP18

中文引用格式: 丁世飞, 朱姜兰, 张成龙, 郭丽丽, 张健. 面向二分类问题的直觉模糊深度随机配置网络. 软件学报, 2025, 36(10): 4660–4670. <http://www.jos.org.cn/1000-9825/7291.htm>

英文引用格式: Ding SF, Zhu JL, Zhang CL, Guo LL, Zhang J. Intuitionistic Fuzzy Deep Stochastic Configuration Network for Solving Binary Classification Problems. Ruan Jian Xue Bao/Journal of Software, 2025, 36(10): 4660–4670 (in Chinese). <http://www.jos.org.cn/1000-9825/7291.htm>

Intuitionistic Fuzzy Deep Stochastic Configuration Network for Solving Binary Classification Problems

DING Shi-Fei^{1,2}, ZHU Jiang-Lan¹, ZHANG Cheng-Long¹, GUO Li-Li^{1,2}, ZHANG Jian^{1,2}

¹(School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China)

²(Mine Digitization Engineering Research Center of Ministry of Education (China University of Mining and Technology), Xuzhou 221116, China)

Abstract: Deep stochastic configuration network (DSCN) adopts a feedforward learning approach and randomly assigns node parameters based on a unique supervisory mechanism, which has universal approximation. However, in actual scenarios, the potential outliers and noise during data collection can negatively affect the classification results. To improve the performance of DSCN in solving binary classification problems, this study introduces the idea of intuitionistic fuzzy numbers based on DSCN and proposes an intuitionistic fuzzy deep stochastic configuration network (IFDSCN). Different from the standard DSCN, IFDSCN assigns an intuitionistic fuzzy number to each sample by calculating the sample membership and non-membership, and generates the optimal classifier by a weighting method to overcome the negative effect of noise and outliers on data classification. The experimental results on eight benchmark datasets show that

* 基金项目: 国家自然科学基金 (62276265, 61976216)

收稿时间: 2023-05-02; 修改时间: 2024-04-07, 2024-06-09; 采用时间: 2024-09-05; jos 在线出版时间: 2025-01-24

CNKI 网络首发时间: 2025-01-26

compared to other learning models including the intuitionistic fuzzy twin support vector machine (IFTWSVM), kernel ridge regression (KRR), intuitionistic fuzzy kernel ridge regression (IFKRR), random vector functional link neural network (RVFL), and SCN, IFDSCN has better binary classification performance.

Key words: intuitionistic fuzzy number; stochastic configuration network; binary classification; data outlier; neural network

近年来, 随机函数向量链接神经网络 (random vector functional link neural network, RVFL) 等随机神经网络模型因其快速的建模效率, 得到快速发展^[1]. 但是, 随机神经网络全局逼近性易受到节点参数范围和节点数量的影响. 与其他随机神经网络不同, Wang 等人^[2]通过不等式约束分配隐含层节点权重, 并采用点增量式建模方式根据数据样本构建网络结构, 提出了随机配置网络 (stochastic configuration network, SCN). SCN 的特点在于不需要事先给定节点参数的初始值, 而是通过一种随机的机制来初始化节点的参数. 此后, SCN 会根据监督机制对节点参数进行选择, 以最大化网络的性能.

自提出以来, SCN 得到快速推广和发展, 被广泛应用于电力大数据处理^[3]、工业过程建模^[4]、在线测量问题^[5]、振动信号故障诊断^[6]等大规模数据分析领域.

为提高 SCN 图像分类性能, Li 等人^[7]将 SCN 扩展为具有矩阵输入的 2D-SCN (2-D stochastic configuration network). 为提高节点增量配置效率, Dai 等人^[8]提出了一种基于块增量学习的随机配置网络 (stochastic configuration network with block increments, BSCN), 对隐含层节点参数进行批量配置. 为对含有噪声和异常值的回归问题进行分析, Wang 等人^[9,10]分别采用核密度估计 (kernel density estimation, KDE) 和最大相关熵准则 (maximum correntropy criterion, MCC) 更新节点参数, 提高了 SCN 的鲁棒性. 为进行大规模数据分析, Wang 等人^[11]对数据特征进行分解, 分别训练不同的基模型, 采用负相关学习 (negative correlation learning, NCL) 方法进行模型集成. 为进行数据流分析, Pratama 等人^[12]将 stacking 方法和 SCN 进行结合, 提出了一种深度堆叠随机配置网络 (deep stacked stochastic configuration network, DSSCN).

尽管上述 SCN 变体从不同角度提高了 SCN 的性能, 但并未对 SCN 的深层扩展进行研究. Wang 等人^[13]对 SCN 进行了深度扩展, 提出了深度随机配置网络 (deep stochastic configuration network, DSCN). 当面临二分类问题时, DSCN 易受到数据中离群值或噪声的影响, 降低了模型分类性能. 在直觉模糊孪生支持向量机 (intuitionistic fuzzy twin support vector machine, IFTWSVM) 中, IFTWSVM 借助直觉模糊数 (intuitionistic fuzzy number, IFN) 为每个训练样本分配隶属度和非隶属度函数, 其中隶属度函数用于计算训练样本与类中心之间的距离, 非隶属度函数用于评价非调和样本个数与其邻域样本个数之间的关系^[14]. 鉴于直觉模糊数能够有效降低噪声的影响, 提高二分类精度, 随后被引入至核岭回归 (kernel ridge regression, KRR)^[15]中, Hazarika 等人^[16]提出了一种直觉模糊核岭回归 (intuitionistic fuzzy kernel ridge regression, IFKRR), 提高了分类精度.

因此, 为提高 DSCN 模型对于解决含有离群值和噪声的二分类问题性能, 本文将直觉模糊的相关概念与 DSCN 相结合, 构建了一种新的监督机制分配隐含层节点参数, 并对输出权重进行加权, 提出了一种直觉模糊深度随机配置网络 (intuitionistic fuzzy stochastic deep configuration network, IFDSCN).

本文的主要贡献如下.

- (1) 引入直觉模糊的概念, 定义了隶属度和非隶属度函数, 构造了直觉模糊得分函数.
- (2) 利用得分函数对数据样本进行加权, 设计了一种新的监督机制分配隐含层节点参数, 构建了基于直觉模糊的加权重.
- (3) 基于 8 个基准二分类问题数据集, 验证了 IFDSCN 的有效性.
- (4) 将噪声引入到数据集中, 验证了 IFDSCN 的鲁棒性.

1 相关工作

1.1 深度随机配置网络

DSCN (deep stochastic configuration network) 为 SCN 的深度扩展, 是一种多隐层前馈神经网络, 有两个隐藏层

的 DSCN 网络结构如图 1 所示, 网络由输入层、多个隐含层和输出层构成.

假定训练数据为: $X = \{(x_1, t_1), \dots, (x_N, t_N)\}$, 其中, $x_i = [x_{i1}, x_{i2}, \dots, x_{id}] \in \mathbb{R}^d$ 表示输入数据; $T = \{t_1, t_2, \dots, t_N\}$, $t_i = [t_{i1}, t_{i2}, \dots, t_{im}] \in \mathbb{R}^m$ 表示相应的输出数据; $i = 1, 2, \dots, N$ 表示样本序号. 具体算法过程可以描述如下.

令 $\Gamma := \{g_1, g_2, g_3, \dots\}$ 表示一组实值函数, $\text{span}(\Gamma)$ 表示由 Γ 张成的函数空间. $L_2(K)$ 表示所有勒贝格可测函数 $f = [f_1, f_2, \dots, f_m] : \mathbb{R}^d \rightarrow \mathbb{R}^m$ 的空间.

在第 n 层的第 L_n 个节点加入网络前, 网络的输出表示为:

$$f_p^{(n)}(x) = \sum_{k=1}^n \sum_{j=1}^p \beta_j^{(k)} g_{k,j}(x^{(k-1)}; w_j^{(k-1)}, b_j^{(k-1)}) \quad (1)$$

网络的残差表示为:

$$e_{L_{n-1}}^{(n)} = f - f_p^{(n)} = e_{L_{n-1}}^{(n)}(X) = [e_{L_{n-1},1}^{(n)}(X), \dots, e_{L_{n-1},m}^{(n)}(X)]^T \quad (2)$$

令 ϵ 为设定的容差, 网络在构造的初期并不满足精度需求, 即 $e_{L_n}^{(n)} > \epsilon$. 随机配置网络能够在每次新增节点时快速选择随机参数 $w_{k,j}$ 和 $b_{k,j}$, 使得网络残差逐渐满足设定的容差, 接下来将描述这种随机参数的配置方法.

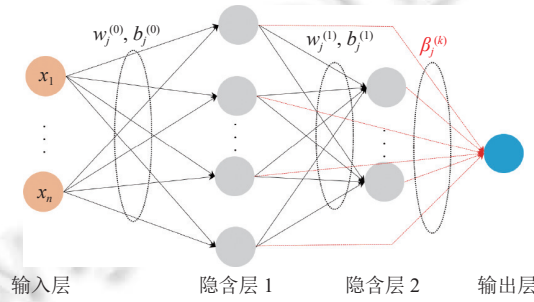


图 1 深度随机配置网络结构

假设 $\text{span}(\Gamma)$ 在 L_2 空间中是稠密的, $\forall g \in \Gamma$, $b_g \in \mathbb{R}^+$, 有 $0 < \|g\| < b_g$ 成立. 给定 $0 < r < 1$, 非负并满足 $\lim_{L \rightarrow +\infty} \mu_l = 0$ 及 $\mu_l \leq (1-r)$ 的序列 $\{\mu_l\}$, 令 $\delta_{j,q}^{(n)} = (1-r-\mu_j) \|e_{j-1,q}^{(n)}\|^2$, $q = 1, 2, \dots, m$. 若满足不等式约束 $\langle e_{j,q}^{(n)}, g_{n,j} \rangle^2 \geq b_g^2 \delta_{j,q}^{(n)}$, $q = 1, 2, \dots, m$, 则有 $\lim_{L \rightarrow +\infty} \|f - f_p^{(n)}\| = 0$ 成立, 即满足网络的全局逼近能.

当第 L_n 个节点加入第 n 个隐含层后, 则第 n 个隐含层的输出为:

$$h_{L_n}^{(n)} := h_{L_n}^{(n)}(X) = [g_{n,L_n}(x_1^{(n-1)}), \dots, g_{n,L_n}(x_N^{(n-1)})]^T \quad (3)$$

令

$$\xi_{L_n,q}^{(n)} = \frac{\langle g_{L_n-1,q}^{(n)}, h_{L_n}^{(n)} \rangle_{l_2}^2}{\langle g_{L_n}^{(n)}, h_{L_n}^{(n)} \rangle_{l_2}} - (1-r) \langle g_{L_n-1,q}^{(n)}, g_{L_n-1,q}^{(n)} \rangle_{l_2} \quad (4)$$

我们选取令公式 (4) 最大的一组隐含层参数, 这也就是 DSCN 配置隐含层参数的方法和理论.

将所有隐层的输出矩阵表示为:

$$H_{L_n}^{(n)} = [h_1^{(n)}, h_2^{(n)}, \dots, h_{L_n}^{(n)}] \quad (5)$$

则 DSCN 的输出权值方法计算如下:

$$\beta^* = \arg \min_{\beta} \left\| f - \sum_{k=1}^n \sum_{j=1}^{L_k} \beta_j^{(k)} g_{k,j} \right\|^2 \quad (6)$$

常用穆尔-彭罗斯广义逆解决公式 (6) 的最小二乘问题, 即 $\beta = H^+ T$.

综上所述, DSCN 的构造过程可以表述如下: 网络从只有单个节点开始逐步增加隐含层节点, 当前隐含层节点

数目满足最初设定的要求后便会增加隐含层, 最终完成网络的构建. 网络中的所有隐含层结点都与输出层直接连接, 通过解决最小二乘目标函数求得输出权重.

1.2 直觉模糊集

直觉模糊集 (intuitionistic fuzzy set, IFS) 用于处理不确定性和模糊性, 最早由 Atanassov^[17] 在 1983 年提出. 直觉模糊集是模糊集的推广, 因为它提供了更多的信息来描述模糊性. 与传统模糊集相比, 直觉模糊集可以更好地描述现实世界中的模糊概念. 本节对直觉模糊数的基本数学概念和性质进行介绍.

对于非空集合 X , 论域 X 中的模糊集合 A 可以定义为:

$$A = \{(x, \mu_A(x)) \mid x \in X\} \quad (7)$$

其中, $\mu_A: X \rightarrow [0, 1]$, $\mu_A(x)$ 是 $x \in X$ 的隶属度. 直觉模糊集定义为:

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x), \nu_{\tilde{A}}(x)) \mid x \in X\} \quad (8)$$

其中, $\mu_{\tilde{A}}(x)$ 和 $\nu_{\tilde{A}}(x)$ 分别定义了 x 的隶属度和非隶属度函数, 有 $\mu_{\tilde{A}}: X \rightarrow [0, 1]$, $\nu_{\tilde{A}}: X \rightarrow [0, 1]$ 且满足 $0 \leq \mu_{\tilde{A}}(x) + \nu_{\tilde{A}}(x) \leq 1$, 则 x 的犹豫度可以表示为:

$$\pi_{\tilde{A}}(x) = 1 - \mu_{\tilde{A}}(x) - \nu_{\tilde{A}}(x) \quad (9)$$

直觉模糊数定义为 $\alpha = (\mu_{\alpha}, \nu_{\alpha})$, 其中 $\mu_{\alpha} \in [0, 1]$, $\nu_{\alpha} \in [0, 1]$, 且满足 $0 \leq \mu_{\alpha} + \nu_{\alpha} \leq 1$. 最大的直觉模糊数为 $\alpha^+ = (1, 0)$, 最小的为 $\alpha^- = (1, 0)$. 给定 $\alpha = (\mu_{\alpha}, \nu_{\alpha})$, 则直觉模糊数可以由下式计算:

$$s(\alpha) = \mu_{\alpha} - \nu_{\alpha} \quad (10)$$

其中, $s(\alpha)$ 表示直觉模糊数 α 的得分值. 然而某些直觉模糊数的得分值无法确定. 为了解决这一问题, 可以用公式 (11) 的函数代替:

$$h(\alpha) = \mu_{\alpha} + \nu_{\alpha} \quad (11)$$

由公式 (10) 和公式 (11), 可以得到:

$$h(\alpha) + \pi(\alpha) = 1 \quad (12)$$

根据公式 (8), 我们可以定义一个如下的得分函数:

$$H(\alpha) = \frac{1 - \nu(\alpha)}{2 - \mu(\alpha) - \nu(\alpha)} \quad (13)$$

隶属度函数和非隶属度函数具有如下的性质:

- 1) 若 $s(\alpha_1) < s(\alpha_2)$, 则有 $H(\alpha_1) < H(\alpha_2)$;
- 2) 若 $s(\alpha_1) = s(\alpha_2)$, $h(\alpha_1) < h(\alpha_2)$, 则可以推出 $H(\alpha_1) < H(\alpha_2)$.

2 直觉模糊深度随机配置网络

本节将直觉模糊与 SCN 结合, 定义了隶属度和非隶属度函数, 构造了直觉模糊得分函数. 构建了一种新的监督机制分配隐含层节点参数, 并对训练数据样本进行加权, 提出了一种直觉模糊深度随机配置网络 IFDSCN.

2.1 直觉模糊隶属度

IFDSCN 模型采用一种基于直觉模糊数进行加权的方法^[18]. 直觉模糊数为隶属度与非隶属度构成的二元组. 其中, 隶属度表示一个元素属于某个集合的程度, 非隶属度则表示一个元素不属于某个集合的程度. 得分函数同时考虑了隶属度与非隶属度, 更全面地刻画了某一元素与集合的关系. 利用得分函数, 可以量化每个数据点相对于预期或常态数据分布的异常程度. 因此, 每个训练样本被赋予一个直觉模糊数, 即 (α, α^*) , α 通过隶属函数计算, 而 α^* 则通过非隶属函数进行计算. 最后, 根据 α 和 α^* 的值定义一个得分函数来评估数据集中的异常值. 其中, 隶属度、非隶属度函数、得分函数定义如下^[18].

1) 隶属度函数考虑了在高维特征空间中每个样本与其对应的类质心之间的距离. 对于每个训练样本 x_i , 隶属度函数定义为:

$$\alpha(x_i) = \begin{cases} 1 - \frac{\|\phi(x_i) - D_+\|}{r^+ + \xi}, & y_i = +1 \\ 1 - \frac{\|\phi(x_i) - D_-\|}{r^- + \xi}, & y_i = -1 \end{cases} \quad (14)$$

其中, D_+ (D_-) 和 r_+ (r_-) 分别为正 (负) 类的中心和半径, ξ 为可调参数. 两个样本之间的距离 P 定义为:

$$P(\phi(x_i), \phi(x_j)) = \|\phi(x_i) - \phi(x_j)\| \quad (15)$$

每个类的半径定义为:

$$r^\pm = \max_{y_i = \pm 1} \|\phi(x_i) - \phi(x_j)\| \quad (16)$$

2) 非隶属度函数: 除了隶属度外, 每个训练样本还被赋予一个非隶属度, 它给出了异质点的数目与其邻域内所有点的数目之间的比例. 因此, 非成员函数定义为:

$$\alpha^*(x_i) = (1 - \alpha(x_i)) \Phi(x_i), 0 \leq \alpha(x_i) + \alpha^*(x_i) \leq 1 \quad (17)$$

其中,

$$\Phi(x_i) = \frac{\left| \{x_j : \|\phi(x_i) - \phi(x_j)\| \leq \eta, y_j \neq y_i\} \right|}{\left| \{x_j : \|\phi(x_i) - \phi(x_j)\| \leq \eta\} \right|} \quad (18)$$

其中, η 为可调参数, $|\cdot|$ 为集合的基数. 可以看出来, 我们需要利用核技巧进行运算.

定理 1^[18]. 设 $K(x, y)$ 为核函数, 则内积距离定义为:

$$\|\phi(x) - \phi(y)\| = \sqrt{K(x, x) + K(y, y) - 2K(x, y)} \quad (19)$$

推论 1^[11]. 利用公式 (14) 和定理 1, 可以计算出 r^+ 和 r^- :

$$r^+ = \max_{y_i = +1} (K(x_i, x_i) + \frac{1}{n_+^2} \sum_{y_m = +1} \sum_{y_n = +1} K(x_m, x_n) - \frac{2}{n_+} \sum_{y_j = +1} K(x_i, x_j))^{1/2} \quad (20)$$

$$r^- = \max_{y_i = -1} (K(x_i, x_i) + \frac{1}{n_-^2} \sum_{y_m = -1} \sum_{y_n = -1} K(x_m, x_n) - \frac{2}{n_-} \sum_{y_j = -1} K(x_i, x_j))^{1/2} \quad (21)$$

3) 得分函数: 在计算出每个样本的隶属度和非隶属度后, 训练样本被赋予直觉模糊数. 因此, 训练集设为: $X = \{(x_1, y_1, \alpha_1, \alpha_1^*), \dots, (x_N, y_N, \alpha_N, \alpha_N^*)\}$. 其中 α_i 和 α_i^* 分别表示样本 x_i 的隶属度和非隶属度. 得分函数可以定义为:

$$\theta_i = \begin{cases} \alpha_i, & \alpha_i^* = 0 \\ 0, & \alpha_i \leq \alpha_i^* \\ \frac{1 - \alpha_i^*}{2 - \alpha_i - \alpha_i^*}, & \text{其他} \end{cases} \quad (22)$$

2.2 IFDSCN

RVFL、KRR 和 SCN 等模型为数据集中的所有样本赋予相等的权值不同, 我们提出的 IFDSCN 为每个样本赋予一个评分值来应对数据集中的离群值. 我们定义 IFDSCN 为:

$$\min_{\beta} \frac{1}{2} \|\beta\|_2^2 + \frac{C}{2} \sum_{i=1}^N s_i \left\| \sum_{k=1}^n \sum_{j=1}^{L_k} \beta_j^{(k)} g_{k,j} - y_i \right\|^2 \quad (23)$$

目标函数 (23) 中 $s_i = \theta_i$, 为样本 i 的惩罚权重, 代表样本对目标函数 (23) 的贡献度. $g_{k,j} = g_{k,j}(w_{k,j}, b_{k,j}, x)$, $w_{k,j}$, $b_{k,j}$ 为 DSCN 网络随机配置产生的隐层权重和偏置, $\beta_j^{(k)}$ 为第 k 层节点的输出权重.

通常情况下, 惩罚权重 s_i 是由样本 x_i 的可靠程度确定的, 加入惩罚权重可以更好地刻画样本的真实分布. 更高的可靠性指标意味着样本更加值得信任, 反之, 较低的可靠性意味着噪声或异常值样本置信度较低. 因此, 采用惩罚权重可以降低噪声或离群值的负面影响. 本文结合了直觉模糊集的概念构造了惩罚权重.

给定训练样本输入 $X = \{x_1, x_2, \dots, x_N\}$, $x_i = [x_{i,1}, \dots, x_{i,d}]^T$, 输出 $T = \{t_1, t_2, \dots, t_N\}$, $t_i = [t_{i,1}, t_{i,2}]^T$, 网络的残差为:

$$e_{L_n-1,q}^{(n)} = [e_{L_n-1,1}^{(n)}(x_1), \dots, e_{L_n-1,q}^{(n)}(x_N)], q = 1, 2 \quad (24)$$

当第 L_n 个节点加入到第 n 个隐含层后, 隐含层的输出矩阵可以表示为:

$$h_{L_n}^{(n)} := h_{L_n}^{(n)}(X) = [g^{n,1}(x^{(n-1)}), \dots, g^{n,L_n}(x^{(n-1)})]^T \quad (25)$$

依据公式 (23), 由于进行了加权处理, 因此我们重新定义残差 $e_{L_n-1}^{(n)}$ 及隐层输出 $h_{L_n}^{(n)}$ 如下:

$$\tilde{e}_{L_n-1}^{(n)} = \Theta e_{L_n-1}^{(n)}(X) \quad (26)$$

$$\tilde{h}_{L_n}^{(n)} = \Theta h_{L_n}^{(n)}(X) \quad (27)$$

公式 (26) 和公式 (27) 中, $\Theta = \text{diag}\{\sqrt{\theta_1}, \sqrt{\theta_2}, \dots, \sqrt{\theta_N}\}$, 为了便于表示与理解, 接下来依然采用 $e_{L_n-1}^{(n)}$ 和 $h_{L_n}^{(n)}$ 表示残差和隐层输出.

我们加入了 L_2 正则化项来避免输出权重 β 求解出现问题. 依据王前进等人^[19]的研究, 结合公式 (26) 及公式 (27), 重新定义 IFDSCN 的不等式约束如下, 并且我们可以很容易地证明模型的全局逼近性.

$$\xi_{L_n,q}^{(n)} = \frac{\langle e_{L_n-1}^{(n)}, h_{L_n}^{(n)} \rangle^2}{\gamma} - (1-r) \langle e_{L_n-1,q}^{(n)}(X), e_{L_n-1,q}^{(n)}(X) \rangle > \quad (28)$$

其中,

$$\gamma = \frac{\left(\langle h_{L_n}^{(n)}, h_{L_n}^{(n)} \rangle + \frac{1}{C} \right)^2}{\langle h_{L_n}^{(n)}, h_{L_n}^{(n)} \rangle + \frac{2}{C}} \quad (29)$$

公式 (23) 的优化问题是凸二次优化问题. 因此, 该问题存在唯一的全局最优解. 该问题的解如下:

$$\beta = \begin{cases} H^T \left(\frac{I}{C} + SHH^T \right)^{-1} SY, & N < L \\ \left(\frac{I}{C} + H^T SH \right)^{-1} H^T SY, & N \geq L \end{cases} \quad (30)$$

其中, $S = \text{diag}\{s_1, s_2, \dots, s_N\}$, s_i 可能是零值, 从而惩罚权重矩阵 S 可能为非奇异矩阵, 导致输出权重 β 求解出现病态解, IFDSCN 的目标函数加入了 L_2 正则化项解决了这一问题.

IFDSCN 的具体算法流程如算法 1 所示.

算法 1. IFDSCN 伪代码.

输入: 训练输入 $X = \{x_1, x_2, \dots, x_N\}, x_i \in \mathbb{R}^d$, 训练输出 $T = \{t_1, t_2, \dots, t_N\}, t_i \in \mathbb{R}^2$; 最大隐含层数 M , 第 n 层最大节点数 $L_{\max}^{(n)}$; 容差 ϵ , 最大配置次数 $T_{\max}$; 两组标量 $\Upsilon = \{\lambda_1, \dots, \lambda_{\text{end}}\}, \mathcal{R} = \{r_1, \dots, r_{\text{end}}\}$; 样本直觉模糊惩罚权重 S ;
输出: IFDSCN 模型.

1. **while** $n \leq M$ and $\|e_0^{(1)}\|_F > \epsilon$ **do**
 2. **while** $L_n \leq L_{\max}^{(n)}$ and $\|e_0^{(1)}\|_F > \epsilon$
 3. **for** $\lambda \in \Upsilon$ **do**
 4. **for** $r \in \mathcal{R}$ **do**
 5. **for** $k = 1, 2, \dots, T_{\max}$ **do**
 6. 在 $[-\lambda, \lambda]^d$ 及 $[-\lambda, \lambda]$ 中随机生成 $\omega_{L_n}^{(n-1)}$ 和 $b_{L_n}^{(n-1)}$;
 7. 通过公式 (26), (27) 计算 $h_{L_n}^{(n)}, \xi_{L_n,q}^{(n)}$;
 8. **if** $\min\{\xi_{L_n,1}^{(n)}, \xi_{L_n,2}^{(n)}\} \geq 0$ **then**
 9. 将 $w_{L_n}^{(n-1)}$ 和 $b_{L_n}^{(n-1)}$ 保存在 W 中,
 10. 将 $\xi_{L_n}^{(n)} = \sum_{q=1}^2 \xi_{L_n,q}^{(n)}$ 保存在 Ω 中;
-

```

11.      else
12.          回到步骤 5
13.      end
14.  end
15.  if  $W$  非空 then
16.      找到使得  $\xi$  最大的  $w_{L_n}^{(n-1)*}, b_{L_n}^{(n-1)*}$ , 令  $H_{L_n}^{(n)} = [h_1^{(n)*}, \dots, h_{L_n}^{(n)*}]$ ;
      Break (转到步骤 22)
17.  else
18.      Continue 转到步骤 4
19.  end
20. end
21. end
22. 令  $\mathcal{H} := [\mathcal{H}, H_{L_n}^{(n)}]$ , 通过公式 (30) 计算  $\beta^*$ ,
    令  $\mathcal{E}_{L_n}^{(n)} = \mathcal{H}\beta^* - T$ ,  $e_0^{(1)} := e_{L_n}^{(n)}, L_n := L_n + 1$ ;
23. end
24. 令  $e_0^{(n+1)} = e^{(n)}, \Omega, W = []$ 
25. end
26. Return  $\beta^*, \omega^*, b^*$ 

```

2.3 算法复杂度分析

设数据集的样本个数为 N , 由于隶属度和非隶属度值是根据每个数据样本计算的, 因此这一部分的时间复杂度为 $O(N)$. 算法模型另一部分的时间复杂度主要取决于矩阵求逆的计算量 N 阶矩阵的求逆时间复杂度为 $O(N^3)$. 因此, IFDSCN 算法的时间复杂度为 $O(N^3) + O(N) \approx O(N^3)$.

3 实验

3.1 实验设置

所有实验都是在 Matlab R2022b、Intel(R) Core(R) CPU 12900H 2.50 GHz、32 GB RAM、Windows 11 系统的实验平台上进行的. 在 IFTWSVM、IFKRR 和 IFSCN 模型中, 我们通过核函数将样本映射到高维空间中生成直觉模糊权值. 在本实验中使用高斯核:

$$K(x_1, x_2) = \exp\left(-\left(\frac{\|x_1 - x_2\|^2}{\mu^2}\right)\right) \quad (31)$$

其中, μ 为核参数. 将数据集按训练集:测试集=7:3 的比例进行随机划分. 我们利用网格搜索法对不同模型对应的超参数进行优化, 调整参数的过程中利用了五折交叉验证法. 在五折交叉验证中, 将数据集随机划分为 5 个不相交集, 其中一个集用于测试, 其余用于训练. 每个模型对应于最佳超参数的性能作为模型的最终精度.

用于评价算法的性能指标 ACC 计算方式如下:

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \quad (32)$$

其中, 假阳性、真阳性、假阴性、真阴性分别用 FP 、 TP 、 FN 、 TN 表示.

3.2 实验数据集介绍

为了验证本文提出的模型性能, 我们在 UCI 和 KEEL 部分二分类基准数据集上进行了实验, 数据集的详细信息如表 1 所示. 值得注意的是, 我们选用了部分不平衡数据集, 即正负样本比例悬殊的数据集进行实验. 不平衡数据集由于某类样本数量较少, 因而模型需要利用有限的该类样本特征信息进行学习, 这对模型的二分类性能形成

了巨大的挑战. 我们混合采用了普通二分类数据集和不平衡二分类数据集, 以期对模型的二分类精度有一个更加全面的评估.

表 1 数据集详细信息

数据集	样本数	负样本数	正样本数	特征数
ecoli2	335	283	52	7
ecoli-0-1-4-7_vs_2-3-5-6	336	308	28	7
ecoli3	335	300	35	7
wdbc	198	151	98	33
heart	270	150	120	14
pima	768	268	500	8
vehicle1	845	628	217	18
abalone918	730	688	42	8

3.3 实验结果分析

我们选取了 KRR^[15]、RVFL^[11]这两种经典的机器学习模型, 以及 IFTWSVM^[14]、IFKRR^[16]这两种结合直觉模糊理论改进的算法模型, 还有 SCN、DSCN 与本文提出的 IFDSCN (在本实验中, 隐含层数量设为 3) 进行了对比实验.

由于模型在某一个数据集的优异表现可以弥补其他数据集的差距, 因此采用单一的平均精度衡量标准并不能很好地体现模型的精度. 本实验对每个数据集的模型进行排序以分析性能. 我们对每一个模型在各个数据集上分别进行排序, 性能越低的模型秩越大, 而性能越高的模型秩越小, 故秩最小的模型精度最优. 设在 N 个数据集上评估模型, 并设 r_j^i 是第 j 个模型在第 i 个数据集上的秩. 则第 j 个模型的平均秩为 $R_j = (1/N) \sum_{i=1}^N r_j^i$.

从表 2 的实验结果中可以看到, 我们提出的 IFDSCN 模型的二分类精度在 8 个数据集上有 7 个领先于其他模型, 平均秩最高. 这表明 IFDSCN 在二分类任务上有着较大的精度优势. 除此之外, IFDSCN 在全部数据集上的二分类精度都优于 DSCN, 这证明了本文提出方法对 DSCN 改进的有效性.

表 2 基准数据集实验结果 (ACC)

数据集	IFTWSVM	KRR	IFKRR	RVFL	SCN	DSCN	IFSCN	IFDSCN
ecoli2	0.8571	0.6964	0.8087	0.8072	0.9431	0.9453	0.9455	0.9501
ecoli-0-1-4-7_vs_2-3-5-6	0.9174	0.9	0.8391	0.8	0.9592	0.9563	0.9687	0.9732
ecoli3	0.8571	0.7218	0.8087	0.8072	0.9257	0.9260	0.9322	0.9355
wdbc	0.564	0.4944	0.5195	0.6716	0.7569	0.7622	0.7796	0.7822
heart	0.8398	0.8717	0.7866	0.8574	0.8037	0.8053	0.8185	0.8213
pima	0.7582	0.7613	0.6886	0.7452	0.7582	0.7625	0.7709	0.7798
vehicle1	0.7859	0.7766	0.7366	0.7683	0.8029	0.8125	0.8225	0.8336
abalone918	0.7502	0.6964	0.7022	0.7035	0.9489	0.9547	0.9528	0.9644
平均准确率	0.7912	0.7398	0.7362	0.7700	0.8623	0.8656	0.8738	0.8861
平均秩	4.93	6.13	7.13	6.12	4.44	3.25	2.62	1.375

我们还采用统计检验 Friedman 对各学习模型的性能差异进行了详尽地分析. Friedman 检验将每个数据集所用的算法模型分别排序, 若出现平局的则赋予平均排名. 零假设表明所有的算法都有相同的性能. 在拒绝零假设的情况下, 为了判断两种学习模型在 k 个学习模型中的性能是否存在显著差异, 我们采用 Nemenyi 后续检验对所有学习模型进行比较. Nemenyi 检验结果如图 2 所示. 较好的模型排名靠后, 较差的模型排名靠前, 因此本文提出的 IFDSCN 的性能最优.

此外, 我们采用配对 win-tie-loss 符号检验来评估模型. win-tie-loss 符号检验评估了每个分类器是赢家、输家或平局的数据集的数量. 在 win-tie-loss 符号检验中, 原假设认为, 两个模型的表现无显著差异. 当两个模型的胜数至少为 $N/2 + 1.96 \sqrt{N}/2$ 时, 称两个模型有显著差异. 如果出现平局, 则在被比较的模型中平均分配数量. 通过简单

的计算, win-tie-loss 检验结果如表 3 所示. $N = 8$ 时, 当任意一个模型在大于等于 5.96 个数据集上获胜时, 两种模型的差异显著. 从表 3 可以看出, IFDSCN 与其他对比模型均存在显著差异, 这体现了 IFDSCN 分类精度的优越性.

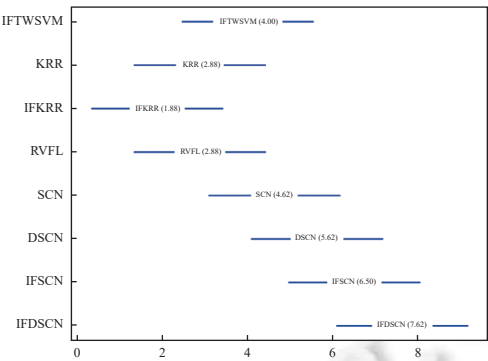


图 2 Nemenyi 检验结果

表 3 win-tie-loss 检验结果

模型	IFTWSM	KRR	IFKRR	RVFL	SCN	DSCN	IFSCN
KRR	[2, 0, 6]	—	—	—	—	—	—
IFKRR	[0, 0, 8]	[4, 0, 4]	—	—	—	—	—
RVFL	[2, 0, 6]	[4, 0, 4]	[5, 0, 3]	—	—	—	—
SCN	[6, 1, 1]	[6, 0, 2]	[8, 0, 0]	[7, 0, 1]	—	—	—
DSCN	[7, 0, 1]	[7, 0, 1]	[8, 0, 0]	[7, 0, 1]	[7, 0, 1]	—	—
IFSCN	[7, 0, 1]	[7, 0, 1]	[8, 0, 0]	[7, 0, 1]	[8, 0, 0]	[7, 0, 1]	—
IFDSCN	[7, 0, 1]	[7, 0, 1]	[8, 0, 0]	[7, 0, 1]	[8, 0, 0]	[8, 0, 0]	[8, 0, 0]

综上所述, 我们采用不同的衡量指标和统计检验对本文提出的 IFDSCN 模型进行了评估, 显而易见, IFDSCN 的模型二分类精度与本文所列的其他机器学习模型相比有着很大的优势. 与此同时, 我们也可以得出两点结论: (1) IFDSCN 相较于 SCN、DSCN 在二分类性能上有了较大的改进; (2) 深层的 IFDSCN 相较于浅层的 IFSCN 也有一定程度的优势, 可见深度扩展的 SCN 有利于提取更多更有效的特征信息, 从而获取更高的二分类精度.

3.4 加入噪声的数据集实验结果

为了全面评估 IFDSCN 模型的鲁棒性, 本节设计了两个对比实验, 分别在基准数据集及人造数据集上进行. 首先, 我们选取了表 1 中的两个数据集: *ecoli3* 和 *wdbc*, 并分别加入了 5% 和 10% 的噪声, 以模拟真实场景中的数据质量波动. 在此基础上, 我们分别在 SCN、DSCN、IFSCN 和本文提出的 IFDSCN 模型上进行了实验, 以比较它们在不同噪声水平下的性能表现. 图 3 展示了各模型在不同噪声比例下的分类精度曲线. 从实验结果可以看出, IFDSCN 在所有噪声设置下均取得了最优的分类性能, 其精度曲线始终高于其他模型. 更值得注意的是, 随着噪声水平的提高, IFDSCN 的精度下降幅度最小, 其折线走势最为平缓. 这一现象充分证明了 IFDSCN 较之其他模型具有更强的鲁棒性, 能够在噪声环境中维持稳定的分类性能.

然后, 我们在 Circle-in-the-Square 问题上加入不同的噪声比例对本文提出的 IFDSN 模型进行鲁棒性能评估. Circle-in-the-Square 是一个二分类问题, 它需要分类器识别正方形中的样本是位于圆形的内部还是外部. 圆的位置为中心, 占正方形的一半. 在本实验中, 我们挑选 10 000 个样本, 分别加入 0、5% 及 10% 的噪声, 实验结果如表 4 所示, 可以看出, IFDSCN 在有噪声的数据集中有着优异的表现, 在各种噪声比例下都有着最高的分类精度.

综合以上实验结果, 我们可以得出以下结论: 本文提出的 IFDSCN 模型在不同程度的噪声干扰下, 展现出了较好的分类精度和鲁棒性, 优于现有的 SCN、DSCN 和 IFSCN 等模型. 这表明, IFDSCN 能够有效地抑制噪声的影响, 在复杂多变的实际应用场景中具有较强的适应性和可靠性.

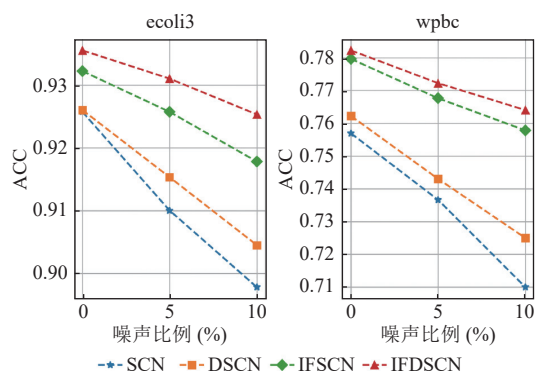


图3 不同噪声比例下的实验结果

表4 加入噪声的 Circle-in-the-Square 问题分类精度

噪声比例 (%)	IFTWSVM	RVFL	SCN	IFSCN	DSCN	IFDSCN
0	0.9944	0.9788	0.9945	0.9960	0.9963	0.9980
5	0.9845	0.9640	0.9841	0.9912	0.9912	0.9923
10	0.9840	0.9585	0.9832	0.9905	0.9908	0.9921

4 结束语

随机配置网络 (SCN) 能够利用监督机制自适应分配隐含层节点参数. 本文引入直觉模糊概念, 提出的 IFDSCN 能够通过构造基于隶属度和非隶属度函数的直觉模糊函数对数据样本进行加权, 并设计了一种新的监督机制分配隐含层节点参数. 通过复在二分类问题数据集上的实验结果表明, IFDSCN 提高了 SCN、DSCN 的二分类精度, 且具有更高的鲁棒性.

多分类问题作为分类问题的重要分支, 在未来的研究中, IFDSCN 将用于解决多分类问题, 以获得更高的分类精度. 除此之外, 受到一些安全学习框架研究的启发^[20,21], 未来将进行进一步研究, 以提高 IFDSCN 数据预测的安全性.

References:

- [1] Pao YH, Takefuji Y. Functional-link net computing: Theory, system architecture, and functionalities. *Computer*, 1992, 25(5): 76–79. [doi: 10.1109/2.144401]
- [2] Wang DH, Li M. Stochastic configuration networks: Fundamentals and algorithms. *IEEE Trans. on Cybernetics*, 2017, 47(10): 3466–3479. [doi: 10.1109/TCYB.2017.2734043]
- [3] Huang CQ, Huang QH, Wang DH. Stochastic configuration networks based adaptive storage replica management for power big data processing. *IEEE Trans. on Industrial Informatics*, 2020, 16(1): 373–383. [doi: 10.1109/TII.2019.2919268]
- [4] Wang QJ, Dai W, Ma XP, Shang ZG. Driving amount based stochastic configuration network for industrial process modeling. *Neurocomputing*, 2020, 394: 61–69. [doi: 10.1016/j.neucom.2020.02.029]
- [5] Wang W, Jia Y, Yu W, Pang HS, Cai KW. On-line ammonia nitrogen measurement using generalized additive model and stochastic configuration networks. *Measurement*, 2021, 170: 108743. [doi: 10.1016/j.measurement.2020.108743]
- [6] Liu JN, Hao RJ, Zhang TL, Wang XZ. Vibration fault diagnosis based on stochastic configuration neural networks. *Neurocomputing*, 2021, 434: 98–125. [doi: 10.1016/j.neucom.2020.12.080]
- [7] Li M, Wang DH. 2-D stochastic configuration networks for image data analytics. *IEEE Trans. on Cybernetics*, 2021, 51(1): 359–372. [doi: 10.1109/TCYB.2019.2925883]
- [8] Dai W, Li DP, Zhou P, Chai TY. Stochastic configuration networks with block increments for data modeling in process industries. *Information Sciences*, 2019, 484: 367–386. [doi: 10.1016/j.ins.2019.01.062]
- [9] Wang DH, Li M. Robust stochastic configuration networks with kernel density estimation for uncertain data regression. *Information Sciences*, 2017, 412–413: 210–222. [doi: 10.1016/j.ins.2017.05.047]
- [10] Li M, Huang CQ, Wang DH. Robust stochastic configuration networks with maximum correntropy criterion for uncertain data regression. *Information Sciences*, 2019, 473: 73–86. [doi: 10.1016/j.ins.2018.09.026]
- [11] Wang DH, Cui CH. Stochastic configuration networks ensemble with heterogeneous features for large-scale data analytics. *Information Sciences*, 2017, 417: 55–71. [doi: 10.1016/j.ins.2017.07.003]
- [12] Pratama M, Wang DH. Deep stacked stochastic configuration networks for lifelong learning of non-stationary data streams. *Information Sciences*, 2019, 495: 150–174. [doi: 10.1016/j.ins.2019.04.055]

- [13] Wang DH, Li M. Deep stochastic configuration networks with universal approximation property. In: Proc. of the 2018 Int'l Joint Conf. on Neural Networks. Rio de Janeiro: IEEE, 2018. 1–8. [doi: [10.1109/IJCNN.2018.8489695](https://doi.org/10.1109/IJCNN.2018.8489695)]
- [14] Rezvani S, Wang XZ, Pourpanah F. Intuitionistic fuzzy twin support vector machines. IEEE Trans. on Fuzzy Systems, 2019, 27(11): 2140–2151. [doi: [10.1109/TFUZZ.2019.2893863](https://doi.org/10.1109/TFUZZ.2019.2893863)]
- [15] Saunders C, Gammerman A, Vovk V. Ridge regression learning algorithm in dual variables. In Proc. of the 15th Int'l Conf. on Machine Learning (ICML 1998). Madison: Morgan Kaufmann Publishers Inc., 1998. 515–521.
- [16] Hazarika BB, Gupta D, Borah P. An intuitionistic fuzzy kernel ridge regression classifier for binary classification. Applied Soft Computing, 2021, 112: 107816. [doi: [10.1016/j.asoc.2021.107816](https://doi.org/10.1016/j.asoc.2021.107816)]
- [17] Atanassov KT. Intuitionistic fuzzy sets. 1983. https://biomed.bas.bg/bioautomation/2016/vol_20.s1/files/20.s1_02.pdf
- [18] Ha MH, Wang C, Chen JQ. The support vector machine based on intuitionistic fuzzy number and kernel function. Soft Computing, 2013, 17(4): 635–641. [doi: [10.1007/s00500-012-0937-y](https://doi.org/10.1007/s00500-012-0937-y)]
- [19] Wang QJ, Yang CY, Ma XP, Zhang CF, Peng SM. Underground airflow quantity modeling based on SCN. Acta Automatica Sinica, 2021, 47(8): 1963–1975 (in Chinese with English abstract). [doi: [10.16383/j.aas.c190602](https://doi.org/10.16383/j.aas.c190602)]
- [20] Guo LZ, Zhang ZY, Jiang Y, Li YF, Zhou ZH. Safe deep semi-supervised learning for unseen-class unlabeled data. In: Proc. of the 37th Int'l Conf. on Machine Learning (ICML 2020). JMLR.org, 2020. 365.
- [21] Li YF, Guo LZ, Zhou ZH. Towards safe weakly supervised learning. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2019, 43(1): 334–346. [doi: [10.1109/TPAMI.2019.2922396](https://doi.org/10.1109/TPAMI.2019.2922396)]

附中文参考文献:

- [19] 王前进, 杨春雨, 马小平, 张春富, 彭思敏. 基于随机配置网络的井下供给风量建模. 自动化学报, 2021, 47(8): 1963–1975. [doi: [10.16383/j.aas.c190602](https://doi.org/10.16383/j.aas.c190602)]



丁世飞(1963—), 男, 博士, 教授, 博士生导师, CCF 杰出会员, 主要研究领域为人工智能与模式识别, 机器学习, 数据挖掘, 大数据智能分析.



郭丽丽(1990—), 女, 博士, 副教授, CCF 专业会员, 主要研究领域为深度学习, 情感计算, 语音情感识别, 多模态情感识别.



朱姜兰(1996—), 男, 博士生, 主要研究领域为机器学习, 随机配置网络.



张健(1990—), 男, 博士, 副教授, CCF 专业会员, 主要研究领域为机器学习, 深度学习.



张成龙(1992—), 男, 博士生, 主要研究领域为随机配置网络, 智能优化算法.