



控制理论与应用
Control Theory & Applications
ISSN 1000-8152, CN 44-1240/TP

《控制理论与应用》网络首发论文

题目: 带有 Dropout 结构的贝叶斯近似宽度学习系统
作者: 陈滔, 王立杰, 刘洋, 徐丽莉, 于海生
收稿日期: 2023-02-26
网络首发日期: 2024-03-06
引用格式: 陈滔, 王立杰, 刘洋, 徐丽莉, 于海生. 带有 Dropout 结构的贝叶斯近似宽度学习系统[J/OL]. 控制理论与应用.
<https://link.cnki.net/urlid/44.1240.TP.20240229.1727.010>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

带有Dropout结构的贝叶斯近似宽度学习系统

陈 滔¹, 王立杰^{1†}, 刘 洋², 徐丽莉³, 于海生¹

(1. 青岛大学 自动化学院, 山东 青岛 266071; 2. 青岛科技大学 自动化与电子工程学院, 山东 青岛 266061;

3. 北京师范大学珠海校区 应用数学学院, 广东 珠海 519087)

摘要: 宽度学习系统(Broad Learning System, BLS) 及其改进算法均普遍存在一个问题, 即随着实际场景中数据复杂性的逐步增强, 网络结构变得极其复杂, 进一步导致计算资源的消耗也大幅度增加. 针对此问题, 本文提出了一种带有Dropout算法的贝叶斯近似宽度学习系统(Dropout-BABLS). 首先, 利用Dropout算法对宽度学习系统的隐藏层节点随机进行丢弃. 其次, 通过结合高斯回归过程和贝叶斯理论近似Dropout对输出结果的损失函数以确定Dropout-BABLS的目标函数, 进一步采用增广拉格朗日乘法对目标函数的输出权重进行优化求解. 最后, 通过UCI机器学习知识库的10组回归数据集和自建的6组时间序列数据集对算法进行分析评估. 结果表明, 本文所提出的Dropout-BABLS算法能保证相应的输出精度, 并减少25% ~ 50%的运行时间.

关键词: 宽度学习系统; Dropout; 高斯过程; 贝叶斯近似; 拉格朗日乘法; 回归分析

引用格式: 陈滔, 王立杰, 刘洋, 等. 带有Dropout结构的贝叶斯近似宽度学习系统. 控制理论与应用, 2024, 41(x): 1 – 10.

DOI: 10.7641/CTA.2023.30087

Bayesian approximate broad learning system with dropout structure

CHEN Tao¹, WANG Li-jie^{1†}, LIU Yang², XU Li-li³, YU Hai-sheng¹

(1. School of Automation, Qingdao University, Qingdao Shandong 266071, China;

2. School of Automation, Qingdao University of Science and Technology, Qingdao Shandong 266061, China;

3. School of Applied Mathematics, Beijing Normal University at Zhuhai, Zhuhai Guangdong 519087, China)

Abstract: The existing broad learning system (BLS) and its improved algorithms have a common problem, that is, with the increasing complexity of data in practical scenarios, the network structure becomes extremely complex, resulting in the consumption of computing resources increased greatly. To handle the problem, this paper proposes a Bayesian approximate broad learning system with dropout structure (Dropout-BABLS). Firstly, the dropout algorithm is used to randomly discard the hidden layer nodes of Broad Learning System. Secondly, by combining the Gaussian regression process and Bayesian theory to approximate the loss function of Dropout on the output results, the objective function of Dropout-BABLS is determined. Next, the augmented Lagrange multiplier method is used to optimize the output weight of the objective function. Finally, the analysis and evaluation of the algorithm 10 sets of regression data of UCI machine learning knowledge base and 6 sets of time series data built by ourselves. The results show that the developed algorithm by Dropout-BABLS can maintain the corresponding output accuracy and reduce the training time by 25% to 50%.

Key words: Broad Learning System; Dropout; Gaussian Process; Bayesian Approximate; Lagrange Multipliers; Regression Analysis

Citation: CHEN Tao, WANG Li-jie, LIU Yang, et al. Bayesian approximate broad learning system with dropout structure. *Control Theory & Applications*, 2024, 41(x): 1 – 10.

1 引言

随着人工智能技术的飞速发展, 深度神经网络已经成为人工智能领域最热门的研究方向之一, 在计算机视觉、自然语言处理、机器人技术等一系列具有

挑战性的领域取得重大创新和广泛应用^[1]. 深度神经网络通过不断增加网络层数来提升网络的泛化能力, 具有优秀的特征提取能力和良好的非线性逼近能力^[2]. 它是在反向传播过程中采用逐层求梯度的方式

收稿日期: 2023-02-26; 录用日期: 2023-12-27.

[†]通信作者. E-mail: wanglijie1@qdu.edu.cn.

本文责任编辑: 丛爽.

国家自然科学基金项目(62103214, 62373208, 62273189), 中国博士后科学基金项目(2021M700077, 2023T160348), 山东省青年泰山学者(tsqnz20221133, tsqn202306218).

Supported by the National Natural Science Foundation of China under Grants (62103214, 62373208, 62273189), the Postdoctoral Science Foundation of China under Grants (2021M700077, 2023T160348), the Shandong Province's Taishan Scholars under Grants (tsqnz20221133, tsqn202306218).

更新网络的权重,这使得深度神经网络模型容易陷入局部最优解、训练时间长等问题.针对此问题,Chen等人^[3]于2018年提出了宽度学习系统(Broad Learning System, BLS)并证明了其良好的性能^[4].相较于深度神经网络, BLS具有网络结构简单、参数少等优点,能够有效减少训练时间和节约计算资源.

近年来,宽度学习得到了国内外诸多学者的关注. Jin等人^[5]提出了一种基于L1范数和L2范数正则化方法的鲁棒性宽度学习系统,有效减少了BLS隐藏层节点的冗余,提高了算法的鲁棒性,但该方法最主要是为提升算法的鲁棒性,对BLS结构的精简程度仍稍显不足.为进一步精简BLS的网络结构,褚菲等人^[6]提出了一种基于lasso和elastic net的宽度学习系统,并证明了其在网络稀疏上的可应用性.考虑到BLS的随机映射对输出结果的影响, Wang等人^[7]通过改变BLS模型的初始随机映射权重分配方法提出了GBLS,进一步提高了模型的数据处理能力,但仍对数据的特征提取程度不够. Cao等人^[8]提出了一种带有混合特征的BLSHF算法,用不同的分布来初始化每个组中的映射特征节点,从而增加映射特征的多样性,可以实现比BLS更好的泛化能力,但也带来了冗余特征.相较于只向“宽度”方向构造的改进宽度学习系统,其“深层”结构有助于提高BLS的精确性. Liu等人^[9]为解决BLS的浅层结构对于其算法能力的限制,提出了一种具有动态结构的堆叠宽度学习系统Stacked BLS. Zhang等人^[10]深入分析了BLS快速增量学习的原理,根据深度神经网络的新结构设计理念在级联结构的基础上分别结合金字塔结构、Dropout操作和密集结构提出了四种全新的BLS变体及其增量实现,有效提高了BLS的输出准确性,但该方法只是“机械”的将级联宽度学习系统与Dropout操作结合起来,没有考虑到Dropout算法对节点丢弃后对输出精度的影响.

综合上述分析,为进一步解决宽度学习系统在面对真实场景应用下的复杂数据时容易出现网络结构过于复杂而导致模型训练成本增加的问题,本文提出了一种带有Dropout结构的贝叶斯近似宽度学习系统(Dropout-BABLS).该方法考虑了Dropout算法以概率 p 对BLS的隐藏层节点随机屏蔽时造成模型输出精度下降的情况,通过引入损失函数对Dropout-BABLS目标函数进行逼近.通过结合高斯回归过程和贝叶斯理论对Dropout-BABLS的目标函数进行近似,进一步采用增广拉格朗日乘子法来对目标函数进行优化改写,然后利用块坐标下降法求解出模型的输出权重.最后,分别结合10组回归数据集和6组真实应用场景下的复杂数据进行实证分析,进一步为BLS的应用与相关算法的改进提供一定的思路和理论支撑.

2 宽度学习系统

对于输入维度为 C ,输出维度为1的单输出回归训练数据集 $\{(X_i, Y_i) | X_i \in R^C, Y_i \in R\}$,其中, $i = 1, 2, \dots, N$.首先,将输入数据 X 随机映射到 n 个特征节点组,每个特征节点组包含 k 个特征节点,第 i 个特征映射组可以表示为:

$$Z_i = \phi(XW_{ei} + \beta_{ei}), i = 1, 2, \dots, n, \quad (1)$$

其中, W_{ei}, β_{ei} 是网络随机生成的第 i 个特征节点组的权重和偏置, $\phi(\cdot)$ 是特征映射函数.定义训练数据集的特征节点空间为 $Z^n = [Z_1, Z_2, \dots, Z_n]$,然后以 Z^n 作为输入通过非线性变换生成 m 个增强节点组,每个增强节点组包含 q 个特征节点,第 j 个增强节点组的输出定义为:

$$H_j = \xi(Z^n W_{hj} + \beta_{hj}), j = 1, 2, \dots, m, \quad (2)$$

其中, W_{hj}, β_{hj} 是网络随机生成的第 j 个增强节点组的权重和偏置, $\xi(\cdot) = \tanh(\cdot)$ 是非线性激活函数.定义增强节点空间的输出为 $H^m = [H_1, H_2, \dots, H_m]$,最后将所有特征节点组和增强节点组串联起来作为BLS的隐藏层,因此,宽度学习系统的输出 \hat{Y} 为:

$$\hat{Y} = AW, \quad (3)$$

其中, $A = [Z^n | H^m]$ 为BLS隐藏层的串节点空间, W 是连接隐藏层到输出层的输出权重.一般来说, A 是奇异的,甚至可能为病态矩阵^[5],因此,输出权重 W 可通过下式最小化目标函数优化求解:

$$\min L_{BLS} = \|Y - AW\|_2^2 + \lambda \|W\|_2^2, \quad (4)$$

其中, λ 是正则化参数,式(4)中第一项用以控制模型训练误差最小化,第二项用以防止BLS模型过拟合.通过岭回归方法^[11],对式(4)关于 W 求导并置导函数为0,求解得到:

$$W = (A^T A + \lambda I)^{-1} A^T Y. \quad (5)$$

训练结束后,随机生成的权重 W_{ei}, W_{hj} 和训练得到的输出权重是固定的 W .因此, BLS对样本量为 M 的测试数据集的响应为:

$$\hat{Y}_t = A_t W, \quad (6)$$

其中, A_t 是宽度学习系统对输入数据 X_t 生成的隐藏层.

3 基于Dropout的贝叶斯近似宽度学习系统

3.1 Dropout-BABLS结构

Dropout是Hinton等人^[12-13]提出的用于解决神经网络出现过拟合现象的方法,能够适用于宽度学习系统出现网络结构复杂、隐藏层节点数量较多的情况. Dropout在对宽度学习系统进行训练时,对节点组的激活值随机置为0或1,激活值为0的节点组不参与输出BLS的结果,其中,激活值的随机选取服从以概

率 p 为参数的伯努利分布^[14]. 本文将带有Dropout结构的宽度学习系统记为Dropout-BABLS, 其网络结构如图2所示.

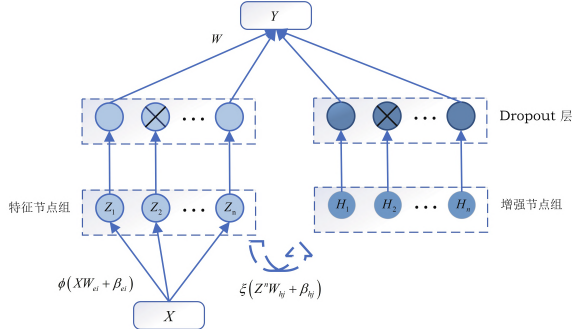


图 1 Dropout-BABLS的网络结构

Fig. 1 Network structure of Dropout-BABLS

记 r_k 为 A 中第 k 个节点组的激活值依概率 p 被置为0或1, $k = 1, 2, \dots, n + m$. 则由以下过程得到Dropout层输出 \tilde{A} 为:

$$\begin{aligned} r_k &\sim \text{Bernoulli}(p), \\ \tilde{A} &= r \circ A, \end{aligned} \quad (7)$$

其中, \circ 代表哈达玛积^[15]. 此时, 宽度学习系统的输出变为:

$$\hat{Y} = \tilde{A}W. \quad (8)$$

记输出结果的损失函数为 $E(\cdot, \cdot)$, 如欧氏损失函数. 定义 $e = Y - \tilde{A}W$ 为误差向量. 利用 L_2 范数正则化项, 可以得到基于Dropout结构的宽度学习系统的最小化目标优化函数^[16-17]为:

$$\min L_{\text{Dropout}} = E(Y, \hat{Y}) + \gamma (\|e\|_2^2 + \|W\|_2^2), \quad (9)$$

其中, γ 是权重衰减因子.

3.2 基于Dropout的贝叶斯近似

不同于传统的BLS采用岭回归的方法直接求解输出权重 W , 本节通过高斯回归过程^[18], 将统计学中的概率论方法与Dropout-BABLS结合起来, 引入贝叶斯理论优化Dropout-BABLS的目标函数, 再利用增广拉格朗日乘法对目标函数优化求解.

根据贝叶斯定理, 对于BLS随机生成的隐藏层 A , 模型参数 W 的后验分布可以表示为:

$$p(W|A, Y) \propto p(Y|A, W)p(W), \quad (10)$$

其中, 假设 $p(W) = N(0, I)$ 是 W 的先验分布. 根据Geller^[19]的理论, 可以定义一个线性协方差函数^[20]为:

$$K(x, y) = (W_k^T x + \tilde{e}_k)^T (W_k^T y + \tilde{e}_k). \quad (11)$$

式(11)中, \tilde{e} 是系统随机生成的误差, 服从某个未知分布 $p(\tilde{e})$. 通过高斯过程对函数空间 F 上的分布进

行建模, 可以在回归任务中分析评估其相应的后验^[21-22]. 因此, 宽度学习系统中的高斯过程为:

$$\begin{aligned} W &= [w_k]_{k=1}^K, \tilde{e} = [\tilde{e}_k]_{k=1}^K, \\ F|A, W, \tilde{e} &\sim N(0, K(A, A)), \\ Y|F &\sim N(F, \tau^{-1}I_N), \end{aligned} \quad (12)$$

其中, $w_k \sim p(W)$, τ 为模型精度的超参数, 本文将其值设置为0.9. I_N 为 $N \times N$ 维的单位矩阵. 基于高斯过程, 可以得到预测分布为:

$$\begin{aligned} p(Y|A) &= \iiint p(Y|F)p(F|W, A, \tilde{e}) \times \\ &p(W)p(\tilde{e})dWd\tilde{e}dF. \end{aligned} \quad (13)$$

对 F 进行解析积分, 得到:

$$\begin{aligned} p(Y|A) &= \iint N(Y; 0, K(A, A) + \\ &\tau^{-1}I_N)p(W)p(\tilde{e})dWd\tilde{e}. \end{aligned} \quad (14)$$

令 $\phi(a_n, W) = a_n W + \tilde{e}$ 为 $1 \times K$ 维行向量, 有 $N \times K$ 维特征矩阵 $\Phi = [\phi(a_n, W, \tilde{e})]_{n=1}^N$, 易得 $K(A, A) = \Phi^T \Phi$, 则式(14)可以重写为:

$$p(Y|A) = \iint N(Y; 0, \Phi^T \Phi + \tau^{-1}I_N)p(W)p(\tilde{e})dWd\tilde{e}. \quad (15)$$

根据式(13)-(15), 易得:

$$p(Y|A, W, \tilde{e}) = N(Y; 0, \Phi^T \Phi + \tau^{-1}I_N). \quad (16)$$

通过以上分析, 宽度学习系统对新的输入 x^* 的预测分布可以参数化为:

$$\begin{aligned} p(y^*|a^*, A, Y) &= \\ \iint p(y^*|a^*, W, \tilde{e})p(W, \tilde{e}|A, Y)dWd\tilde{e}. \end{aligned} \quad (17)$$

式(17)中, a^* 是输入 x^* 生成的新的BLS隐藏层, 后验分布 $p(W, \tilde{e}|A, Y)$ 是难以解析求解的, 需要通过变分推断的方法引入一个变分分布 $q(W, \tilde{e}) := q(W)q(\tilde{e})$ 来近似 $p(W, \tilde{e}|A, Y)$. 通过最小化变分分布 $q(W, \tilde{e})$ 与后验分布 $p(W, \tilde{e}|A, Y)$ 的Kullback-Leibler(KL)散度, 来确定最优变分分布近似实际的后验分布, KL散度计算如下:

$$KL(q(W, \tilde{e}) || p(W, \tilde{e}|A, Y)) = \quad (18)$$

$$\begin{aligned} &\iint q(W, \tilde{e}) \log \frac{q(W, \tilde{e})}{p(W, \tilde{e}|A, Y)} dW d\tilde{e} = \\ &\iint q(W, \tilde{e}) \log \frac{q(W, \tilde{e}) p(A, Y)}{p(W, \tilde{e}, A, Y)} dW d\tilde{e} = \\ &\log p(A, Y) - \iint q(W, \tilde{e}) \log \frac{p(W, \tilde{e}, A, Y)}{q(W, \tilde{e})} dW d\tilde{e}. \end{aligned}$$

对于式(18)中等式右边第二项, 有:

$$\iint q(W, \tilde{e}) \log \frac{p(W, \tilde{e}, A, Y)}{q(W, \tilde{e})} dW d\tilde{e} = \quad (19)$$

$$\begin{aligned} & \iint q(W, \tilde{e}) \log \frac{p(Y|A, W, \tilde{e})p(W, \tilde{e})p(A, Y)}{q(W, \tilde{e})p(Y|A)} dW d\tilde{e} = \\ & \iint q(W, \tilde{e}) \log p(Y|A, W, \tilde{e}) dW d\tilde{e} + \log p(A, Y) - \\ & KL(q(W, \tilde{e}) \| p(W, \tilde{e})) - \log p(Y|A). \end{aligned}$$

由于 $p(Y|A)$ 是一个常数,依照式(18)和式(19),可知最小化KL散度等价于最大化对数证据下界(the Log Evidence Lower Bound, ELBO),即:

$$\max L_{EBLO} = \iint q(W, \tilde{e}) \log p(Y|A, W, \tilde{e}) dW d\tilde{e} - KL(q(W, \tilde{e}) \| p(W, \tilde{e})). \quad (20)$$

利用蒙特卡洛积分进行对式(20)进行近似. 首先从 $q(W, \tilde{e})$ 采样 \hat{W} 和 \hat{e} ,然后针对 $q(W, \tilde{e})$ 进行优化,以最大化式(21)的目标函数. 接下来,更新 $q(W, \tilde{e})$,并从中抽取新的样本 \hat{W} 和 \hat{e} . 重复上述操作,最终收敛到一个最接近真实后验分布 $p(W, \tilde{e}|A, Y)$ 的 $q(W, \tilde{e})$.

$$\begin{aligned} \max L_{EBLO-MC} &= \sum_{n=1}^N \log p(y_n | a_n, \hat{W}, \hat{e}) - \\ & KL(q(W, \tilde{e}) \| p(W, \tilde{e})). \end{aligned} \quad (21)$$

在宽度学习系统的隐藏层与输出层之间添加一个Dropout层,取 $W \sim p(W)$,构造 $q(W)$ 如下:

$$\begin{aligned} \hat{W} &= W \cdot \text{diag}(r_k), \\ r_k &\sim \text{Bernoulli}(p). \end{aligned} \quad (22)$$

其中,隐藏层节点损失活概率 p 与矩阵 \hat{W} 为变分参数. 变分分布 $q(W)$ 是高度多模态的,在 \hat{W} 的行上诱导强联合相关性. 关于随机误差 \tilde{e} 的变分分布 $q(\tilde{e})$,可以取一个简单的高斯分布:

$$q(\tilde{e}) = N(e, \sigma^2 I_K). \quad (23)$$

对于回归任务,由式(16)可以将式(20)中右边第一项的被积函数重写为:

$$\begin{aligned} \log p(Y|A, W, \tilde{e}) &= \log N(Y; 0, \Phi^T \Phi + \tau^{-1} I_N) = \\ & -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\Phi^T \Phi + \tau^{-1} I_N) - \frac{Y^T Y}{2(\Phi^T \Phi + \tau^{-1} I_N)}. \end{aligned} \quad (24)$$

高斯过程的各输出是独立的,忽略系统随机生成的误差 \tilde{e} ,有 $\Phi \approx \hat{Y} = AW$. 于是,式(24)可以转化为:

$$\begin{aligned} \log p(Y|A, W, \tilde{e}) &\approx -\frac{1}{2} \log(2\pi) - \\ & \frac{1}{2} \log(\|AW\|_2^2 + \tau^{-1}) - \frac{\|Y\|_2^2}{2(\|AW\|_2^2 + \tau^{-1})}. \end{aligned} \quad (25)$$

根据文献[25]的研究理论, $q(W, \tilde{e})$ 和 $p(W, \tilde{e})$ 之间的KL散度可以近似为:

$$KL(q(W, \tilde{e}) \| p(W, \tilde{e})) \approx \frac{p}{2} \|W\|_2^2 + \frac{1}{2} \|e\|_2^2. \quad (26)$$

通过上述分析,式(20)可转化为最小化下式的目标函数:

$$\begin{aligned} \min L_{GP-MC} &\propto \log(\|AW\|_2^2 + \tau^{-1}) + \\ & \frac{\|Y\|_2^2}{\|AW\|_2^2 + \tau^{-1}} + p \|W\|_2^2 + \|e\|_2^2, \end{aligned} \quad (27)$$

其中, $E(Y, \hat{Y})$ 满足下式:

$$E(Y, \hat{Y}) = \log\left(\|\hat{Y}\|_2^2 + \tau^{-1}\right) + \frac{\|Y\|_2^2}{\|\hat{Y}\|_2^2 + \tau^{-1}}. \quad (28)$$

至此,从数学理论的角度证明了,最小化带有Dropout结构的宽度学习系统的目标函数,同最小化变分分布与高斯过程的后验分布之间的KL散度所达到的优化效果一致,即 L_{GP-MC} 等价于 $L_{Dropout}$.

3.3 Dropout-BABLS优化求解

由于Dropout-BABLS的目标函数 L_{GP-MC} 结构较为复杂,不易通过解析求解出权重 W . 本节利用增广拉格朗日乘子法(the Augmented Lagrange Multiplier, ALM)来对目标函数 L_{GP-MC} 进行优化求解,引入松弛变量 D 来替代 AW ,同时由 $e = Y - AW$,可将式(27)改写为:

$$\begin{aligned} \arg \min_{D, W} & \log(\|D\|_2^2 + \tau^{-1}) + \frac{\|Y\|_2^2}{\|D\|_2^2 + \tau^{-1}} + \\ & p \|W\|_2^2 + \|Y - AW\|_2^2, \\ \text{s.t. } & D = AW. \end{aligned} \quad (29)$$

依据ALM方法,式(29)的增广拉格朗日函数为:

$$\begin{aligned} L(D, W, C) &= \log(\|D\|_2^2 + \tau^{-1}) + \frac{\|Y\|_2^2}{\|D\|_2^2 + \tau^{-1}} + \\ & p \|W\|_2^2 + \|Y - AW\|_2^2 + C^T (D - AW) + \\ & \frac{\mu}{2} \|D - AW\|_2^2. \end{aligned} \quad (30)$$

其中, C 为拉格朗日乘子, $\mu \geq 0$ 是惩罚因子. 然后,利用块坐标下降法优化求解上式中的未知变量 D, W, C ,具体步骤如下:

1)求解 W : 令 $F = D + \frac{C}{\mu}$,保持变量 D, C 不变,

移除与 W 无关的项, 可得下式:

$$\arg \min_W p \|W\|_2^2 + \|Y - AW\|_2^2 + \frac{\mu}{2} \|F - AW\|_2^2. \quad (31)$$

上式是一个典型的正则化最小二乘问题, 其解为:

$$W = (2pI + (2 + \mu) A^T A)^{-1} (2A^T Y + \mu A^T F). \quad (32)$$

2) 求解 D : 令 $E = AW - \frac{C}{\mu}$, 保持变量 W, C 不变, 移除与 D 无关的项, 可得下式:

$$\arg \min_D \log \left(\|D\|_2^2 + \tau^{-1} \right) + \frac{\|Y\|_2^2}{\|D\|_2^2 + \tau^{-1}} + \frac{\mu}{2} \|E - D\|_2^2. \quad (33)$$

由式(33)可知, D 满足下式:

$$D \leftarrow \left(\frac{2\mu}{\|D\|_2^2 + \tau^{-1}} I - \frac{2\mu \|Y\|_2^2}{\left(\|D\|_2^2 + \tau^{-1} \right)^2} I + \mu^2 I \right)^{-1} E. \quad (34)$$

3) 拉格朗日乘子 C 迭代: 在每次迭代中, 参数 C 可以更新为:

$$C \leftarrow C + \mu (D - AW). \quad (35)$$

对于给定的拉格朗日乘子 C 和惩罚因子 μ , 通过迭代重复块坐标下降(32)、(34)所生成的递归序列 $\{D^l, W^l\}$ 最终会收敛于式(29)的最优解, 这是容易证明的, 本文不再对此重复赘述^[23-24].

4 实验与分析

4.1 实验设计

为了验证Dropout-BABLS算法的有效性, 本文选取传统机器学习算法(SVM和LSSVM)、神经网络领域前沿算法(自适应深度神经网络ADNN^[25]、非线性自回归神经网络NARNN^[26]和单层长短时神经网络LSTM^[27])、标准BLS、L1BLS、ENBLS、L1RBLS、ENRBLS和DKBLS^[28]进行对比实验, 每组实验均进行20次独立实验并选取最优实验结果进行比较分析, 以平均绝对百分比误差(MAPE)和均方根误差(RMSE)作为算法测试精度的评价指标。

算法的参数选取对学习效果具有重要的影响作用。因此, 为了公平的比较本文所提出的Dropout-BABLS算法与其他算法的回归效果, 需要通过合适的方式进行参数设置。对于SVM算法, 采用网格搜索法从 $[2^{-24}, \dots, 2^{25}]$ 来确定不同数据集的惩罚参数 C 和径向基函数核参数 γ , LSSVM则通过MATLAB工具箱 $LS - SVMlab$ 来确定 (C, γ) 。就本文而言, 3种神经网络前沿算法的输入层节点数 l 与输入变量数一

致, 输出层节点数为1。其中, 采用顺序搜索法来对自适应深度神经网络ADNN的隐层数 h 和各隐层节点数 q 进行搜索, 搜索范围分别为 $[1, 5]$ 和 $[1, l]$; 非线性自回归神经网络NARNN的延迟阶数 k 根据数据集特点分别设置为2或3^[26], 其隐层节点数 q 则根据经验公式 $q = \sqrt{1 + l} + s, s \in [2, 15]$ 采用顺序搜索的方法进行选取; 单层长短时记忆神经网络LSTM的隐层节点数则利用网格搜索法从 $[1, 100]$ 范围内以步长1进行搜索确定。同时, 还需要对Dropout-BABLS、标准BLS、L1BLS、ENBLS、L1RBLS和ENRBLS进行网格搜索, 包括每个特征节点组中的特征节点数量 k , 特征节点组数量 n 和每个增强节点组中的特征节点数量 q , 分别针对不同回归数据集从 $[1, 20] \times [1, 20] \times [1, 200]$ 中搜索, 步长设置为1。另外, 将Dropout-BABLS的隐藏层失活概率 p 和模型精度 τ 分别设置为0.5和0.9。对于DKBLS, 只需对高斯核函数参数 σ 进行调节即可, 本文采用顺序搜索法从 $[2^{-10}, 2^{-9}, \dots, 2^9, 2^{10}]$ 范围内选取 σ 。所有实验均采用统一的实验平台。具体配置信息为: CPU为AMD Ryzen5 5500U@2.10GHz 6Cores, 内存16GB, 图像处理单元(GPU)为AMD Radeon(TM) Graphics, 操作系统为Windows11终端, 其中, Friedman post-hoc test检验实验的编程语言及版本为R语言4.2.3版本, 其它所有实验的编程语言及版本均为MATLAB2022b。

4.2 回归数据集测试

4.2.1 数据集描述

本节实验选取10组回归数据集来自于美国加州大学欧文分校的UCI机器学习知识库^[5,29], 根据各组数据集的特点可划分为三类: 小尺寸低维度、中等尺寸中等维度和大尺寸低维度, 数据集的规格见表1。

4.2.2 不同算法实验结果分析

对每个算法的参数依照上述实验设计进行选取, 结果如表2所示。通过观察表2, 可以发现本文所提出的Dropout-BABLS算法相较于BLS、L1BLS、ENBLS、L1RBLS和ENRBLS, 所需的隐藏层节点数量更多。这是因为Dropout-BABLS对隐藏层节点进行了随机丢弃, 需要增加更多的节点以保证模型输出结果的精确度。

在选定所需的参数之后, 得到每个算法的回归测试结果见表3。从算法的训练时间来看, Dropout-BABLS在网络节点数量增加的情况下所需要的学习时长明显优于其他BLS改进算法, 相较于ADNN、NARNN和单层LSTM这三种神经网络前沿算法而言, Dropout-BABLS具有更高效的学习能力。从算法测试时间来看, Dropout-BABLS算法除了对Weather Izmir的测试集进行测试所需的时间略高于ENBLS算法外, 对其他9组回归数据集进行测试所需的时间均明显低于其

他算法. 从算法的输出结果来看, Dropout-BABLS对于10组回归数据集输出结果明显优于标准BLS算法和传统机器学习算法SVM与LSSVM; 同时, 相较于其他宽度学习系统的改进算法以及三种神经网络前沿算法, Dropout-BABLS的输出精度得到了明显地提高. Dropout-BABLS在Abalone和Quake两个数据集上取得了显著优势, 不仅训练成本也大幅降低, 其输出精度得到了略微提升. 此外, Dropout-BABLS在Bodyfat、Strike、Basketball、Cleveland和Pyrin这5个数据集上也取得了相对比较明显的优势. 特别地, 通过表3可以发现, Dropout算法对数据集Housing的影响相对较大的, 但通过本文提出的贝叶斯近似方法对BLS进行优化后, Dropout-BABLS与ENBLS在Housing数据集上输出精度MAPE和RMSE的差异也分别仅为1%和4%.

表 1 回归数据集的规格信息

Table 1 Specification information of regression data set

数据集	输入变量	样本量		目标值范围 [y_{\min} , y_{\max}]
		训练集	测试集	
Abalone	8	2784	1393	[1, 29]
Bodyfat	14	168	84	[0, 47.5]
Strike	6	416	209	[0, 7000]
Housing	13	337	169	[5, 50]
Mortgage	15	699	350	[0, 7000]
Weather Izmir	9	974	487	[0, 7000]
Basketball	4	64	32	[0.1593, 0.8291]
Cleveland	13	202	101	[126, 564]
Pyrin	27	49	25	[0.1, 0.9]
Quake	3	1452	726	[5.8, 6.9]

表 2 各算法的参数设置

Table 2 Parameter setting of each algorithm

数据集	SVM		LSSVM		ADNN				NARNN				单层LSTM	BLS			L1BLS			ENBLS			L1RBLS			ENRBLS			DKBLS	Dropout-BABLS		
	C	γ	C	γ	h	q	k	q	q	k	n	q	k	n	q	k	n	q	k	n	q	k	n	q	k	n	q	σ	k	n	q	
Abalone	2 ²	2 ⁻²	2.9382	9.1072	3	5	2	9	23	5	4	40	4	6	166	6	6	50	13	15	97	14	19	102	2 ⁻¹⁰	14	5	176				
Bodyfat	2 ⁰	2 ⁻¹	5276.0061	286.7154	5	13	3	8	54	5	5	16	3	14	64	14	10	25	5	4	93	7	19	200	2 ⁰	9	3	185				
Strike	2 ⁻⁴	2 ⁻²	0.4167	0.9938	2	6	2	5	18	4	8	76	5	16	155	5	20	117	3	15	181	5	17	126	2 ⁵	5	17	120				
Housing	2 ³	2 ¹	45.0965	7.1136	4	9	3	14	26	8	6	61	4	16	163	8	16	94	11	16	197	14	10	178	2 ⁻²	14	10	178				
Mortgage	2 ¹³	2 ⁻¹	76.7623	5.4322	5	11	3	12	41	12	5	197	3	20	200	4	14	191	9	7	134	16	16	193	2 ⁵	16	16	197				
Weather Izmir	2 ⁴	2 ⁻²	1652.0012	24.5381	3	7	2	6	20	7	4	91	3	14	85	18	8	164	3	17	151	8	18	138	2 ⁶	8	18	138				
Basketball	2 ⁰	2 ⁰	9.0367	27.3176	3	3	2	8	96	9	5	7	14	14	17	14	14	16	12	13	68	9	16	13	2 ⁰	10	17	145				
Cleveland	2 ⁻¹	2 ²	6.4432	44.0015	2	10	2	12	41	12	5	2	14	10	27	14	9	107	7	9	46	10	12	39	2 ⁻³	14	10	197				
Pyrin	2 ¹⁸	2 ³	66.3198	2.5455	1	26	3	9	25	7	9	69	2	16	44	6	4	28	3	19	157	4	6	19	2 ⁻¹	10	17	141				
Quake	2 ⁷	2 ⁻⁵	1.2287	4.2616	5	3	2	15	78	11	9	9	11	9	10	9	11	75	4	16	22	9	7	51	2 ⁷	10	16	169				

为了进一步比较Dropout-BABLS相较于其他算法的优越性, 本节对各算法的训练时间进行了Friedman post-hoc test检验, 结果如表4所示. 结果表明, 相较于传统机器学习算法和神经网络前沿算法, Dropout-BABLS在显著性水平0.01条件下具有显著差异, 可以认为Dropout-BABLS在节约训练成本上具有显著优势. 对于BLS及DKBLS两个算法来说, Dropout-BABLS也能在0.01水平条件下通过显著性检验, 说明Dropout-BABLS能极大程度减少宽度学习系统的隐层节点数量而不降低模型的输出精度. 对于L1BLS、ENBLS、L1RBLS和ENRBLS来说, Dropout-BABLS仅能在显著性水平0.05条件下通过检验, 这是由于这四种算法的目标函数均包含L1范数, 能够一定程度上稀疏BLS的网络结构, 但简化程度相较于Dropout-BABLS而言仍有较大差距.

4.3 时序数据预测

4.3.1 数据集构造

为了验证Dropout-BABLS对现实数据的学习能力, 本节利用天气后报网(<http://www.aqistudy.cn/historydata/>)提供的空气质量历史数据, 采集了山东省青岛市、济南市、烟台市、聊城市、临沂市和日照市等6个城市2022年01月01日-2022年10月31日的历史空气质量数据, 分别包含 $PM_{2.5}$ 、 AQI 、 PM_{10} 、 NO_2 、 SO_2 、 CO 和 O_3 等7个变量, 其中, $PM_{2.5}$ 与其他6个变量具有高度相关性^[30]. 为了消除各变量不同量级、量纲的影响, 需要分别依照式(36)对获取到的原始数据进行标准化处理. 在得到标准化数据之后, 以 AQI 、 PM_{10} 、 NO_2 、 SO_2 、 CO 和 O_3 作为输入变量 X , $PM_{2.5}$ 作为输出变量 Y , 将2022年01月01日-2022年08月31日的标准化数据作为训练集、2022年09月01日-2022年10月

表 3 各算法的输出结果
Table 3 Output results of each algorithm

数据集		SVM	LSSVM	ADNN	NARNN	单层LSTM	BLS	L1BLS	ENBLS	L1RBLS	ENRBLS	DKBLS	Dropout-BABLS
Abalone	训练时间(s)	0.0751	0.0743	0.0604	0.0782	0.0711	0.0619	0.0616	0.0577	0.0768	0.057	6.9836	0.0563
	测试时间(s)	0.0774	0.0781	0.0417	0.0399	0.0455	0.0518	0.0487	0.0376	0.0685	0.0427	4.1310	0.0326
	MAPE($\times 10^{-2}$)	13.79	9.13	6.01	5.65	8.99	6.32	5.36	5.53	5.81	5.28	5.72	5.28
	RMSE($\times 10^{-2}$)	15.25	12.17	9.79	9.13	12.35	11.39	7.63	7.82	8.25	7.58	7.97	7.57
Bodyfat	训练时间(s)	0.0199	0.0167	0.0113	0.0105	0.0148	0.0130	0.0198	0.0162	0.0132	0.0174	0.0186	0.0093
	测试时间(s)	0.0104	0.0185	0.0076	0.0067	0.0093	0.0070	0.0079	0.0138	0.0139	0.0090	0.0152	0.0057
	MAPE($\times 10^{-2}$)	2.94	1.09	0.67	0.64	0.91	0.86	0.52	0.63	0.69	0.70	0.76	0.57
	RMSE($\times 10^{-2}$)	1.89	2.31	0.68	0.93	1.75	1.52	0.66	0.71	1.49	0.82	0.94	0.62
Strike	训练时间(s)	0.1366	0.1134	0.0433	0.0501	0.0695	0.0735	0.0432	0.0538	0.0399	0.0486	0.1165	0.0388
	测试时间(s)	0.0579	0.0719	0.0319	0.0368	0.0413	0.0420	0.0386	0.0333	0.0403	0.0424	0.0868	0.0288
	MAPE($\times 10^{-2}$)	5.63	4.02	3.94	3.82	3.99	3.94	3.91	3.91	3.79	3.70	4.19	3.71
	RMSE($\times 10^{-2}$)	13.22	12.35	10.59	10.77	10.52	10.41	10.22	10.14	10.19	10.02	10.89	9.98
Housing	训练时间(s)	0.0714	0.0811	0.0466	0.0439	0.0598	0.0621	0.0559	0.0453	0.0471	0.0433	0.0761	0.0431
	测试时间(s)	0.0312	0.0746	0.0319	0.0231	0.0377	0.0419	0.0221	0.0289	0.350	0.0436	0.0614	0.0163
	MAPE($\times 10^{-2}$)	13.52	7.10	5.09	5.55	6.13	5.09	5.13	5.03	5.30	5.17	5.18	5.13
	RMSE($\times 10^{-2}$)	14.23	10.77	9.15	8.86	8.03	7.67	7.55	7.56	8.34	8.07	7.46	7.91
Mortgage	训练时间(s)	0.1043	0.0676	0.0473	0.0419	0.0574	0.0441	0.0405	0.0577	0.404	0.0613	0.9068	0.0385
	测试时间(s)	0.6187	0.0223	0.0222	0.0182	0.0276	0.0181	0.0231	0.0394	0.0270	0.0291	0.2448	0.0109
	MAPE($\times 10^{-2}$)	2.09	0.84	0.64	0.98	1.73	0.65	0.38	0.37	0.48	0.42	0.63	0.38
	RMSE($\times 10^{-2}$)	5.32	2.39	1.72	1.99	2.76	1.14	0.53	0.52	0.69	0.58	0.92	0.53
Weather Izmir	训练时间(s)	0.0799	0.0611	0.0457	0.0377	0.0410	0.0389	0.0354	0.0384	0.0402	0.0379	2.2156	0.0319
	测试时间(s)	0.0718	0.0392	0.0253	0.0356	0.0326	0.0251	0.0370	0.0193	0.0396	0.0305	0.5139	0.0294
	MAPE($\times 10^{-2}$)	4.27	6.39	1.87	2.11	3.79	2.08	1.39	1.37	1.44	1.39	1.67	1.38
	RMSE($\times 10^{-2}$)	12.52	10.13	2.66	3.95	4.14	2.75	1.90	1.89	1.99	1.94	2.29	1.93
Basketball	训练时间(s)	0.0735	0.0462	0.0548	0.0399	0.0429	0.0133	0.0325	0.0359	0.0344	0.0336	0.0511	0.0059
	测试时间(s)	0.0469	0.0233	0.0366	0.435	0.0322	0.0114	0.0260	0.0314	0.0294	0.0301	0.0355	0.0053
	MAPE($\times 10^{-2}$)	12.00	10.11	8.44	9.25	9.67	8.61	8.58	8.58	8.49	8.05	8.07	7.94
	RMSE($\times 10^{-2}$)	10.45	8.65	9.01	8.14	7.44	8.26	7.05	7.13	7.32	6.62	6.64	6.85
Cleveland	训练时间(s)	0.0685	0.0469	0.0443	0.0476	0.0297	0.0319	0.0217	0.0275	0.0244	0.0203	0.0610	0.0131
	测试时间(s)	0.0423	0.0310	0.0275	0.0254	0.0158	0.0263	0.0188	0.0214	0.0238	0.0251	0.0221	0.0092
	MAPE($\times 10^{-2}$)	11.58	10.12	9.33	8.55	9.03	8.75	8.20	8.40	8.19	8.18	8.24	8.24
	RMSE($\times 10^{-2}$)	15.69	15.23	12.97	12.58	13.25	12.76	11.45	12.21	11.81	12.07	12.02	11.35
Pyrim	训练时间(s)	0.0312	0.0298	0.0177	0.0365	0.0176	0.0312	0.0214	0.0193	0.0229	0.0355	0.0496	0.0075
	测试时间(s)	0.0155	0.0246	0.0181	0.0214	0.0097	0.0285	0.0196	0.0108	0.0170	0.0240	0.0449	0.0029
	MAPE($\times 10^{-2}$)	5.99	6.14	5.28	5.33	5.96	5.27	4.43	3.90	4.88	5.11	5.02	3.92
	RMSE($\times 10^{-2}$)	13.85	11.68	8.39	9.59	10.56	10.02	6.22	4.84	6.60	7.69	8.23	4.67
Quake	训练时间(s)	0.0753	0.0528	0.0291	0.0202	0.0339	0.0323	0.0288	0.0267	0.0273	0.0280	2.0894	0.0094
	测试时间(s)	0.0326	0.0227	0.0117	0.0135	0.0205	0.0233	0.0103	0.0281	0.0208	0.0203	0.2134	0.0038
	MAPE($\times 10^{-2}$)	19.25	16.82	13.96	13.55	13.76	13.57	13.54	13.56	13.52	13.55	13.53	13.52
	RMSE($\times 10^{-2}$)	23.11	21.59	19.02	17.39	18.20	18.43	18.02	17.29	17.22	17.01	17.22	17.01

表 4 Dropout-BABLS与其他算法的Friedman post-hoc test检验结果

Table 4 Friedman post-hoc test results for Dropout-BABLS and other algorithms

	SVM	LSSVM	ADNN	NARNN	单层LSTM	BLS	L1BLS	ENBLS	L1RBLS	ENRBLS	DKBLS	Dropout-BABLS
Dropout-BABLS	0.000	0.000	0.002	0.002	0.000	0.001	0.022	0.019	0.013	0.011	0.000	/

31日的标准化数据作为测试集,自建了6组时间序列数据集用于本节的实验测试与分析.

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}}. \quad (36)$$

式中, X' 为标准化数据, X 为原始数据, X_{\min} 和 X_{\max} 分别为数据的最小值和最大值.

4.3.2 预测结果比较与分析

利用各算法分别对六个城市2022年09月01日-2022年10月31日的 $PM_{2.5}$ 进行预测,如表5所示.可以发现ropout-BABLS对烟台市、济南市和聊城市3个城市的 $PM_{2.5}$ 预测效果是相对较好的,平均绝对百分比误差MAPE的值分别为4.3%、4.14%和4%. 本文所提出的Dropout-BABLS算法对烟台市、聊城市、临沂市和日照市的预测效果均明显优

于其他改进BLS算法和三种神经网络前沿算法、对青岛市的预测效果仅次于ENBLS、对济南市的预测效果仅次于ENRBLS 算法,整体预测精确度都相对较高. Dropout-BABLS与ENBLS对青岛市预测精度的MPAE和RMSE分别仅相差0.69%和0.37%情况下, 其进行训练和测试的时间足足相对减少了一半以上, 更进一步体现了Dropout-BABLS的优越性. 从各算法执行所需时间来看, 单层LSTM所需要的时间相对优于除Dropout-BABLS外的改进BLS 算法,其输出精度也略差于其他算法. ADNN和NAR-NN对6组时间序列数据的预测效果和其他BLS改进算法没有明显的差异, 而Dropout-BABLS在一定程度上提高算法的预测精度的同时大幅度节省了算法运行的时间.

表 5 时序数据预测效果对比
Table5 Comparison of prediction effects of time series data

数据集	ADNN	NARNN	单层LSTM	BLS	L1BLS	ENBLS	L1RBLS	ENRBLS	DKBLS	Dropout-BABLS
青岛	训练时间(s)	0.0513	0.0467	0.0573	0.0413	0.0407	0.0610	0.0433	0.0494	0.0923
	测试时间(s)	0.0342	0.0224	0.0419	0.0364	0.0314	0.0264	0.0229	0.0359	0.0300
	MAPE($\times 10^{-2}$)	6.24	6.96	7.10	6.01	6.06	5.24	6.27	6.17	6.24
	RMSE($\times 10^{-2}$)	4.09	5.17	5.99	3.30	3.16	2.80	3.53	3.28	4.20
济南	训练时间(s)	0.0463	0.0477	0.0579	0.0420	0.0427	0.0513	0.0474	0.0496	0.1050
	测试时间(s)	0.0301	0.0396	0.0331	0.0294	0.0356	0.0394	0.0265	0.0313	0.0294
	MAPE($\times 10^{-2}$)	4.19	4.24	5.57	4.30	4.33	4.23	4.56	4.08	4.55
	RMSE($\times 10^{-2}$)	2.88	2.87	3.16	2.91	2.88	2.85	2.88	2.91	3.12
烟台	训练时间(s)	0.0338	0.0294	0.0392	0.0366	0.0276	0.0289	0.0338	0.0290	0.1003
	测试时间(s)	0.0319	0.0277	0.0353	0.0312	0.0282	0.0268	0.0282	0.0272	0.0327
	MAPE($\times 10^{-2}$)	4.62	4.44	5.17	4.79	4.69	4.33	5.11	4.48	4.49
	RMSE($\times 10^{-2}$)	3.81	3.55	4.31	4.36	3.93	3.78	4.01	3.76	3.37
聊城	训练时间(s)	0.0475	0.0643	0.0528	0.0590	0.0435	0.0503	0.0428	0.0482	0.0843
	测试时间(s)	0.0398	0.0455	0.0442	0.0437	0.0401	0.0466	0.0429	0.0423	0.0419
	MAPE($\times 10^{-2}$)	4.95	4.72	5.11	4.26	4.11	4.06	4.14	4.09	4.54
	RMSE($\times 10^{-2}$)	3.63	3.56	3.68	3.56	3.72	3.53	3.60	3.59	3.94
临沂	训练时间(s)	0.0499	0.0578	0.0611	0.0574	0.0510	0.0505	0.0563	0.0639	0.0946
	测试时间(s)	0.0542	0.0413	0.0467	0.0483	0.0338	0.0384	0.0396	0.0430	0.0344
	MAPE($\times 10^{-2}$)	8.01	7.45	7.58	7.59	7.23	7.20	7.28	7.25	7.16
	RMSE($\times 10^{-2}$)	5.26	4.17	4.25	4.15	3.98	3.95	4.03	3.93	4.05
日照	训练时间(s)	0.0449	0.0534	0.0486	0.0492	0.0427	0.0335	0.0387	0.0394	0.0910
	测试时间(s)	0.0284	0.0438	0.0270	0.0353	0.0336	0.0194	0.0229	0.0270	0.0496
	MAPE($\times 10^{-2}$)	9.47	10.02	9.84	10.95	9.57	9.76	9.59	9.46	9.54
	RMSE($\times 10^{-2}$)	4.47	4.75	4.84	4.89	4.47	4.49	4.26	4.39	4.64

5 总结

为有效解决宽度学习系统在面对真实场景下时网络结构过于庞大导致计算资源损耗增加问题, 本文提出了一种带有Dropout结构的贝叶斯近似宽度学习系统(Dropout-BABLS). 通过对10组回归数据集和6组时间序列数据集进实证分析. 结果表明, 本

文所提出的算法在保证输出精度的条件下显著提高了算法的学习能力, 减少了计算资源的消耗, 同时, Dropout-BABLS对真实场景下的时序数据预测效果也取得了良好的效果. 本研究仅针对回归和预测问题进行了基于Dropout的贝叶斯近似, 后续研究将把Dropout-BABLS 扩展到模式识别分类领域, 提

升算法的普适性和可移植性.

参考文献:

- [1] BOULILA W, DRISS M, ALSHANQITI E, et al. Weight initialization techniques for deep learning algorithms in remote sensing: Recent trends and future perspectives. *Advances on Smart and Soft Computing*, 2022, 1399: 477 – 484.
- [2] LY H B, NGUYEN T A, TRAN V Q. Development of deep neural network model to predict the compressive strength of rubber concrete. *Construction and Building Materials*, 2021, 301: 124081.
- [3] 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*, 2017, 29(1): 10 – 24.
- [4] CHEN C L P, LIU Z L, FENG S. Universal approximation capability of broad learning system and its structural variations. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 30(4): 1191 – 1204.
- [5] JIN J W, CHEN C L P. Regularized robust broad learning system for uncertain data modeling. *Neurocomputing*, 2018, 322(12): 58 – 69.
- [6] ZHU Fei, SU Jiaming, LIANG Tao, et al. Sparsity method for network structure of broad learning system based on lasso and elastic net. *Control Theory & Applications*, 2020, 37(12): 2543 – 2550. (褚菲, 苏嘉铭, 梁涛, 等. 基于lasso和elastic net的宽度学习系统网络结构稀疏方法. 控制理论与应用, 2020, 37(12): 2543 – 2550.)
- [7] WANG Y, JIA P F, CUI H, et al. A novel regression prediction method for electronic nose based on broad learning system. *IEEE Sensors Journal*, 2021, 21(17): 19374 – 19381.
- [8] CAO W P, LI D C, ZHANG X J, et al. BLSHF: broad learning system with hybrid features. *The 15th International Conference on Knowledge Science, Engineering and Management*. Singapore: Springer, 2022: 655 – 666.
- [9] LIU Z, CHEN C L P, FENG S, et al. Stacked broad learning system: From incremental flatted structure to deep model. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021, 51(1): 209 – 222.
- [10] ZHANG L, LI J, LU G, et al. Analysis and variants of broad learning system. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2022, 52(1): 334 – 344.
- [11] GONG X, ZHANG T, CHEN C L P, et al. Research review for broad learning system: algorithms, theory, and applications. *IEEE Transactions on Cybernetics*, 2021, 52(9): 8922 – 8950.
- [12] HINTON G E, SRIVASTAVA N, KRIZHEVSKY A, et al. Improving neural networks by preventing co-adaptation of feature detectors. *Computer Science*, 2012, 3(4): 212 – 223.
- [13] BALDI P, SADOWSKI P. The Dropout learning algorithm. *Artificial Intelligence*, 2014, 210(1): 78 – 122.
- [14] SHI L, COPOT C, VANLANDUIT S. Evaluating dropout placements in bayesian regression resnet. *Journal of Artificial Intelligence and Soft Computing Research*, 2021, 12(1): 61 – 73.
- [15] ARVIND V, CHATTERJEE A, DATTA R, et al. Fast exact algorithms using hadamard product of polynomials. *Algorithmica*, 2022, 84(2): 436 – 463.
- [16] RONALD S. *Qualitative analysis of monte carlo dropout*. (2019-11-13) [2023-02-25]. <https://arxiv.org/pdf/2007.01720.pdf>.
- [17] SADR MA, GANTE J, CHAMPAGNE B, et al. Uncertainty estimation via monte carlo dropout in CNN-Based mmWave MIMO localization. *IEEE Signal Processing Letters*, 2021, 29(11): 269 – 273.
- [18] LIN M X. *Research on commodity sales forecast model based on gaussian process regression under large-scale data with uncertain demand*. Guangdong: Jinan University, 2020.
- [19] GELLER D K. Linear covariance techniques for orbital rendezvous analysis and autonomous onboard mission planning. *Journal of Guidance, Control, and Dynamics*, 2006, 29(6): 1404 – 1414.
- [20] GAL Y, GHAHRAMAN Z. Dropout as a bayesian approximation: representing model uncertainty in deep learning. *Proceedings of The 33rd International Conference on Machine Learning*. New York, America: PMLR, 2016: 1050 – 1059.
- [21] GAL Y, GHAHRAMAN Z. Bayesian convolutional neural networks with bernoulli approximate variational inference. *International Conference on Learning Representations*. San Juan, Puerto Rico: arXiv Press, 2016: 1 – 12.
- [22] GAL Y, TURNER R. Improving the gaussian process sparse spectrum approximation by representing uncertainty in frequency inputs. *Proceedings of the 32nd International Conference on International Conference on Machine Learning*. Lille, France: PMLR, 2015: 655 – 664.
- [23] LIN Z, CHEN M, MA Y. *The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices*. (2010-10-13)[2023-02-25]. <https://people.eecs.berkeley.edu/~yima/matrix-rank/Files/ALM-v5.pdf>.
- [24] ZHANG Z, LAI Z, XU Y, et al. Discriminative elastic-net regularized linear regression. *IEEE Transactions on Image Processing*, 2017, 29(3): 1466 – 1481.
- [25] RAO Zhangyu, LIU Chunyuan, PENG Zhen, et al. Calculation model of detent force of permanent magnet synchronous linear motor based on adaptive deep neural network. *Computer Era*, 2023(2): 1 – 6 + 10. (饶章宇, 刘春元, 彭珍, 等. 基于自适应深度神经网络的永磁同步直线电机定位力计算模型. 计算机时代, 2023(2): 1–6 + 10.)
- [26] XU X, YU H, WANG X, et al. Prediction of temperature in firing zone of cement rotary kiln with nonlinear autoregressive neural network model. *2022 37th Youth Academic Annual Conference of Chinese Association of Automation (YAC)*. Beijing, China: IEEE, 2022: 1060 – 1064.
- [27] SUN Kai, SUI Lin, ZHANG Fangfang, et al. Dynamic soft sensor algorithm based on nonnegative garrote and LSTM neural network. *Control Theory & Applications*, 2023, 40(1): 83-93. (孙凯, 隋琳, 张芳芳, 等. 基于非负绞杀与长短期记忆神经网络的动态软测量算法. 控制理论与应用, 2023, 40(1): 83-93.)
- [28] YU Z W, LAN K K, LIU Z L. Progressive Ensemble Kernel-Based Broad Learning System for Noisy Data Classification. *IEEE Transactions on Cybernetics*, 2022, 52(9): 9656 – 9669.
- [29] DAVID A, PATRICK M, CHRISTOPHER M, et al. *U-CI machine learning repository*. (1995-12-01) [2022-02-25]. <http://archive.ics.uci.edu/ml/datasets.php>.
- [30] LIU Zuoyun, LIU Henlin. Analysis on the characteristics of pollution and correlation of influencing factors of PM2.5 in Hengyang, China. *Journal of Hunan Ecological Science*, 2021, 8(3): 69 – 75. (刘作云, 刘衡林. 衡阳市PM2.5污染特征及影响因素相关性分析. 湖南生态科学学报, 2021, 8(3): 69 – 75.)

作者简介:

陈滔 硕士研究生, 目前研究方向为宽度学习系统、信息处理的研究与应用, E-mail: qdu.chentao@qdu.edu.cn;

王立杰 副教授, 硕士生导师, 目前研究方向为复杂系统的智能优化控制理论与应用研究, E-mail: wanglijie1@qdu.edu.cn;

刘洋 特聘教授, 硕士生导师, 目前研究方向为多智能体协同控制、车辆队列控制, E-mail: 03878@qust.edu.cn;

徐丽莉 讲师, 硕士生导师, 目前研究方向为统计学习、贝叶斯统计方向的研究, E-mail: xulili@bnu.edu.cn;

于海生 教授, 博士生导师, 目前研究方向为智能控制与机器人技术、工程系统控制与优化的研究, E-mail: yhsh_qd@qdu.edu.cn.