

基于图正则化的贝叶斯宽度学习系统

段俊伟¹, 许林灿¹, 全渝娟¹, 陈龙², 陈俊龙³

(1. 暨南大学信息科学技术学院, 广东 广州 510632;

2. 澳门大学科技学院, 澳门 999078;

3. 华南理工大学计算机科学与工程学院, 广东 广州 510006)

摘要: 作为一种前馈神经网络, 宽度学习系统因其精度高、训练速度快且能有效代替深度学习方法而备受研究者的关注。然而, 宽度学习系统存在对网络中的特征节点个数比较敏感且求伪逆方式易使模型出现过拟合等问题。为此, 在宽度学习系统中引入贝叶斯推断和图正则化。一方面, 通过引入先验知识进行贝叶斯学习可以有效提高权重的稀疏性, 提高模型的稳定性; 另一方面, 加入图正则化可充分考虑数据内在的图信息, 进一步提高模型的泛化能力。在 UCI 数据集和 NORB 数据集上对所提模型进行性能评估, 实验结果表明, 所提的基于图正则化的贝叶斯宽度学习系统模型能进一步提高宽度学习系统的分类精度且具有更好的稳定性。

关键词: 宽度学习系统; 贝叶斯推断; 图正则化; 模式识别

中图分类号: TP181

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202203

Graph-regularized Bayesian broad learning system

DUAN Junwei¹, XU Lincan¹, QUAN Yujuan¹, CHEN Long², CHEN C.L. Philip³

1. College of Information Science and Technology, Jinan University, Guangzhou 510632, China

2. Faculty of Science and Technology, University of Macau, Macau 999078, China

3. School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China

Abstract: As a feed forward neural network, broad learning system (BLS) has attracted much attention because of its high accuracy, fast training speed, and the ability to effectively replace deep learning methods. However, it is sensitive to the number of feature nodes and the pseudo-inverse method is likely to result in the problem of over fitting for BLS model. To address the above issues, Bayesian inference and graph regularization was introduced in to the BLS model. By introducing the prior knowledge for Bayesian learning, the sparsity of the weights and the stability of the model could be effectively improved; while the graph information mining from the data could be fully considered to improve the generalization ability of the model by regularization. The UCI and NORB dataset were adopted for evaluating the performance of the proposed model. The experiment results demonstrated that the proposed graph-regularized Bayesian broad learning system model can further improve the accuracy of classification and has better stability.

Key words: board learning system, Bayesian inference, graph regularization, pattern recognition

收稿日期: 2021-02-01; 修回日期: 2021-04-17

通信作者: 全渝娟, Tquanyj@jnu.edu.cn

基金项目: 国家重点研发计划基金资助项目 (No.2018YFC2002500); 广东省基础与应用基础研究基金资助项目 (No. 2021A1515011999); 广州市科技创新发展专项资金项目 (No.201902010041)

Foundation Items: The National Key Research and Development Program of China (No.2018YFC2002500), The Guangdong Basic and Applied Basic Research Foundation (No.2021A1515011999), The Guangzhou Science and Technology Innovation and Development Special Fund Project (No.201902010041)

0 引言

目前,利用人工智能技术快速准确地获取数据并进行分析与处理已成为被广泛关注的问题之一^[1-2]。神经网络(neural network, NN)可以从海量数据中提取有效的特征,常见的模型有径向基函数(radial basis function, RBF)网络^[3]、自组织映射(self-organizing map, SOM)网络^[4]、递归神经网络(recurrent neural network, RNN)^[5]、玻尔兹曼机(Boltzmann machine, BM)^[6]、深度置信网络(deep belief network, DBN)^[7]、卷积神经网络(convolutional neural network, CNN)^[8]等。然而,为了有效完成更复杂的任务,在神经网络模型中有大量超参数需要调节,这会导致模型的训练时间较长且因结构过于复杂而缺乏解释性,这些问题都限制了深度神经网络(deep neural network, DNN)理论的发展^[9-11]。

为了有效解决上述问题,Chen C L P 等人^[12]提出了宽度学习系统(broad learning system, BLS),BLS 将输入数据映射到特征层形成特征节点,然后利用非线性激活函数产生所需的增强节点,最后将特征节点与增强节点相连,通过求伪逆方法获得与输出层的权值矩阵。扩展模型时,BLS 的增量学习方法可以有效实现对模型的快速重建且不需要消耗大量的训练时间。参考文献[13]证明了 BLS 的收敛性。BLS 的基本结构如图 1 所示。

近年来 BLS 的理论正在不断被完善且在不同领域中获得了较好的应用。郑云飞等人^[14]在 BLS 模型的基础上提出将最小 p -范数作为损失函数,使

得 BLS 模型能够较好地处理不同噪声的干扰问题。Yang F 等人^[15]将 CNN 和主成分分析(principal component analysis, PCA)与 BLS 模型结合,使得 BLS 模型在图像分类上的能力进一步提高。Feng S 等人^[16]将模糊逻辑引入 BLS,将 BLS 的特征层替换为多个模糊子系统,提出了一种模糊宽度学习系统(fuzzy broad learning system, FBLS)。基于 FBLS 结构简单、网络参数少、训练效率高等特点,李润泽等人^[17]将 FBLS 应用在光伏发电功率的预测任务中,实验结果显示,与传统的预测模型相比,FBLS 具有更优良的预测性能。Xu L L 等人^[18]在 BLS 中引入贝叶斯推断(Bayesian inference),利用贝叶斯学习方式代替原本的伪逆求解权重方式,进而提高模型的泛化能力。Jin J W 等人^[19]提出了图正则化(graph regularization),用于约束 BLS 的损失函数,使其能够学习更多的数据几何结构信息,进而提高分类能力。贾晨等人^[20]将两个 BLS 进行结合,提出了双宽度学习系统,并将其应用于多模态信息融合,实验结果表明,该模型在保持较高准确率的同时具备快速性和更好的稳定性。褚菲等人^[21]将 BLS 目标函数中的 ℓ_2 范数替换为 ℓ_1 范数和弹性网络(elastic net),目的是提高 BLS 的稀疏性,回归任务实验显示,该模型在保持良好预测精度的同时简化了网络结构。Yang L 等人^[22]将 BLS 与迁移学习(transfer learning)结合,提出了源域自适应宽度学习系统(BLS based source domain adaption, BLS-SDA)和目标域自适应宽度学习系统(BLS based target domain adaptation, BLS-TDA)两种模型。在域适应基准数据集 office+Caltech-256 的实验

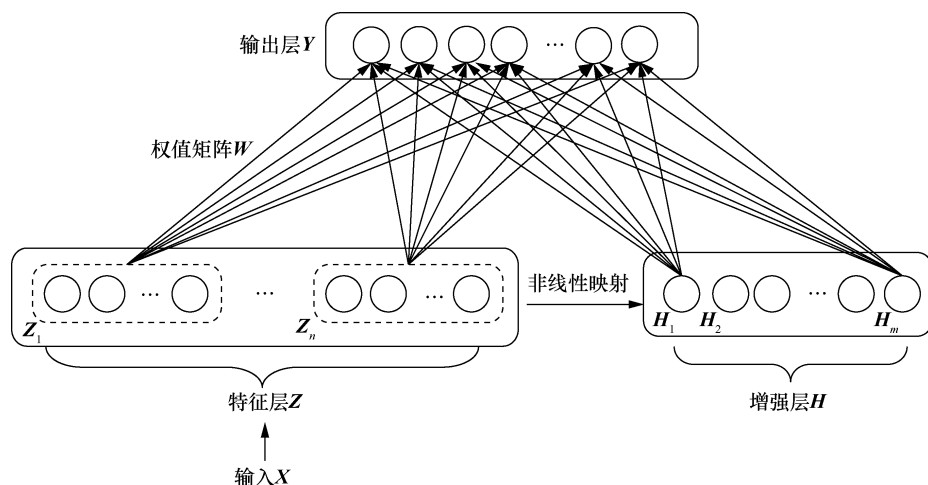


图1 BLS 的基本结构

中显示, 与支持向量机 (support vector machine, SVM)、极限学习机 (extreme learning machine, ELM)、迁移成分分析 (transfer component analysis, TCA) 等方法相比, BLS-SDA 和 BLS-TDA 拥有更高的准确率。李旺等人^[23]将 BLS 应用于蘑菇毒性判别中, 以蘑菇的气味和颜色为特征训练 BLS, 与其他方法相比, 基于 BLS 的蘑菇毒性判别方法具有高效性、快速性、易扩展性。

为了进一步提高模型的分类能力和泛化能力, 本文在 BLS 的基础上, 将贝叶斯推断和图正则化同时引入 BLS, 提出一种基于图正则化的贝叶斯宽度学习系统 (graph-regularized Bayesian broad learning system, GBBLS)。通过引入贝叶斯推断^[18,24-25], 输出值的概率分布可以被更好地估计, 而不仅是对数据进行拟合, 在一定程度上也有效地避免了过拟合问题; 引入图正则化^[19]的主要目的是使模型能够结合数据的几何结构, 在拟合过程中增加更多的约束, 从而进一步提高模型的泛化能力。本文提出的 GBBLS 模型如下: 首先随机初始化权重矩阵和偏置形成特征层和增强层; 接着对特征层和增强层映射到输出层的权值矩阵进行零初始化, 并且将权值矩阵的先验分布设置为服从零均值的高斯分布, 高斯分布的方差初始化为 1; 然后基于训练样本计算对数似然函数, 同时在该函数中添加图正则化; 最后利用拉普拉斯近似 (Laplace approximation)^[26]和阻尼牛顿法^[27], 对基于样本的权值矩阵的后验分布进行优化, 并将此后验分布的均值作为权值矩阵的最终结果。

本文其余部分主要内容如下: 第 1 节介绍 BLS 的主要结构, 并将贝叶斯推断引入 BLS; 第 2 节在贝叶斯宽度学习系统 (Bayesian broad learning system, BBLs) 的基础上添加图正则化; 第 3 节基于 UCI 数据集和 NORB 数据集对 GBBLS 进行性能评估, 主要包括准确率和增强节点数目稳定性分析; 第 4 节对全文内容进行总结, 并对将来的研究方向进行分析与展望。

1 贝叶斯宽度学习系统

1.1 宽度学习系统

BLS 是一种前馈神经网络, 具有很强的非线性映射能力, 同时其训练过程具有快速性和高效性, 具体过程如下。

假设有 N 个样本 $\{X, Y\}$, 其中每个样本

$x_i (i=1, 2, 3, \dots, N)$ 是一个 d 维向量, 且每个维度均为连续型特征, 则 $X \in \mathbb{R}^{N \times d}$ 。每个样本对应的标签 $y_i (i=1, 2, 3, \dots, N)$ 是经过独热编码的 c 维向量, 则 $Y \in \mathbb{R}^{N \times c}$ 。因此 $X=[x_1, x_2, \dots, x_N]^T, Y=[y_1, y_2, \dots, y_N]^T$ 。

首先, 输入 X 经过 n 组特征映射形成特征层 $Z=[Z_1, Z_2, \dots, Z_n]$ 。每个特征节点矩阵 Z_i 满足:

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

其中, $\phi(\cdot)$ 是特征映射函数, W_{ei} 是随机初始化的权重矩阵, β_{ei} 是随机初始化的偏置。

其次, 将特征层 Z 通过非线性激活函数进一步扩展, 得到增强层 $H=[H_1, H_2, \dots, H_m]$ 。每个增强节点矩阵 H_j 满足:

$$H_j = \xi(ZW_{hj} + \beta_{hj}), j=1, 2, \dots, m \quad (2)$$

其中, $\xi(\cdot)$ 是非线性激活函数, W_{hj} 是随机初始化的正交矩阵, β_{hj} 是随机初始化的偏置。

则 BLS 的输出可以表示为 $\hat{Y}=A_x W$, 其中 $A_x=[Z, H]$, 是由特征层 Z 和增强层 H 拼接形成的矩阵。 W 是将 A_x 映射到输出层 $\hat{Y}$ 的权值矩阵。可通过基于 ℓ_2 范数的岭回归^[28]进行求解, 表示为:

$$W = \operatorname{argmin}(\|\hat{Y} - A_x W\|_2^2 + \lambda \|W\|_2^2) = (A_x^T A_x + \lambda I)^{-1} A_x^T \hat{Y} \quad (3)$$

其中, λ 为正则化参数, I 为单位矩阵。

1.2 在宽度学习系统中引入贝叶斯推断

不同于 BLS 中通过岭回归直接求解 W , 在引入贝叶斯推断之后, BLS 通过贝叶斯学习获得 W 。先从二分类开始引入贝叶斯推断, 则每个样本都可被看作一个独立的伯努利事件, 其概率 $p(r|x)$ 满足伯努利分布。因此可得似然函数:

$$p(r|w, A_{xi}) = \prod_{i=1}^N \sigma(A_{xi} w)^{r_i} [1 - \sigma(A_{xi} w)]^{1-r_i} \quad (4)$$

其中, $\sigma(\cdot)$ 为 Sigmoid 函数:

$$\sigma(A_{xi} w) = \frac{1}{1 + e^{-A_{xi} w}} \quad (5)$$

同时, $r=[r_1, r_2, \dots, r_N]^T$, $r_i \in \{0, 1\}$, 表示样本标签; $w=[w_1, w_2, \dots, w_{n+m}]^T$ 。对于 w 中的每个元素 w_k , 其先验分布均满足零均值的高斯分布, 则可得如下计算式:

$$w_k \sim \mathcal{N}(0, \gamma_k^{-1}) \quad (6)$$

$$p(\mathbf{w}|\boldsymbol{\gamma}) = \prod_{k=1}^{n+m} \frac{\gamma_k}{\sqrt{2\pi}} e^{-\frac{\gamma_k \mathbf{w}_k^2}{2}} \quad (7)$$

式(7)中, $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \dots, \gamma_k, \dots, \gamma_N]^\top$, 为超参数。

下一步是求解 $\mathbf{r}$ 在 $\mathbf{A}_x$ 和 $\boldsymbol{\gamma}$ 上的边缘似然, 它被定义为如下的积分形式:

$$p(\mathbf{r}|\boldsymbol{\gamma}, \mathbf{A}_x) = \int p(\mathbf{r}|\mathbf{w}, \mathbf{A}_x) p(\mathbf{w}|\boldsymbol{\gamma}) d\mathbf{w} \quad (8)$$

求解式(8)的核心目的是在建立 $p(\mathbf{r}|\boldsymbol{\gamma}, \mathbf{A}_x)$ 的分布的基础上将 $p(\mathbf{r}|\boldsymbol{\gamma}, \mathbf{A}_x)$ 最大化, 进而确定 $\boldsymbol{\gamma}$ 的取值。在此, 首先需要最大化 $p(\mathbf{r}|\mathbf{w}, \mathbf{A}_x) p(\mathbf{w}|\boldsymbol{\gamma})$ 的值。根据高斯分布的性质, 因为 $p(\mathbf{r}|\mathbf{w}, \mathbf{A}_x)$ 和 $p(\mathbf{w}|\boldsymbol{\gamma})$ 均满足高斯分布, 所以 $p(\mathbf{r}|\mathbf{w}, \mathbf{A}_x) p(\mathbf{w}|\boldsymbol{\gamma})$ 也满足高斯分布, 可通过拉普拉斯近似来求解, 具体过程如下。

先定义对数似然函数 J :

$$J = \ln(p(\mathbf{r}|\mathbf{w}, \mathbf{A}_x) p(\mathbf{w}|\boldsymbol{\gamma})) = \sum_{i=1}^N [r_i \ln s_i + (1-r_i) \ln(1-s_i)] - \frac{1}{2} \mathbf{w}^\top \boldsymbol{\Gamma} \mathbf{w} + \mathbf{c} \quad (9)$$

其中, $s_i = \sigma(\mathbf{A}_{xi} \mathbf{w})$, $\boldsymbol{\Gamma} = \text{diag}(\boldsymbol{\gamma})$, $\text{diag}(\cdot)$ 表示取矩阵对角线上的元素组成向量, $\mathbf{c}$ 为常数项。这里采用阻尼牛顿法优化参数 $\mathbf{w}$ 。利用式(9)分别对 $\mathbf{w}$ 求一阶导数和二阶导数得到 $\boldsymbol{\phi}$ 和 $\boldsymbol{\psi}$:

$$\boldsymbol{\phi} = \frac{\partial J}{\partial \mathbf{w}} = \sum_{i=1}^N (r_i - s_i) \mathbf{A}_{xi} - \boldsymbol{\Gamma} \mathbf{w} = \mathbf{A}_x^\top (\mathbf{r} - \mathbf{s}) - \boldsymbol{\Gamma} \mathbf{w} \quad (10)$$

$$\boldsymbol{\psi} = \frac{\partial J}{\partial^2 \mathbf{w}} = -(\mathbf{A}_x^\top \mathbf{Q} \mathbf{A}_x + \boldsymbol{\Gamma}) \quad (11)$$

其中, $\mathbf{Q}$ 为对角矩阵, 对角上的元素为 $s_i(1-s_i), i=1, 2, \dots, N$ 。

则 $\mathbf{w}$ 的优化表达式为:

$$\mathbf{w}_{\text{new}} = \mathbf{w}_{\text{old}} - \lambda \boldsymbol{\psi}^{-1} \boldsymbol{\phi} \quad (12)$$

其中, $\mathbf{w}_{\text{old}}$ 表示更新前的 $\mathbf{w}$, $\mathbf{w}_{\text{new}}$ 表示更新后的 $\mathbf{w}$ 。

因此, 由拉普拉斯近似可得到 $p(\mathbf{r}|\mathbf{w}, \mathbf{A}_x) \cdot p(\mathbf{w}|\boldsymbol{\gamma}) \sim \mathcal{N}(\mathbf{w}_{\text{best}}, \boldsymbol{\psi}^{-1})$, $\mathbf{w}_{\text{best}}$ 为通过阻尼牛顿法找到的 $\mathbf{w}$ 的最优解。

然后, 对式(8)的对数边缘似然 $\mathcal{N}_{\text{LC}}(\boldsymbol{\gamma})$ 进行求解:

$$\mathcal{N}_{\text{LC}}(\boldsymbol{\gamma}) = \ln(p(\mathbf{r}|\boldsymbol{\gamma}, \mathbf{A}_x)) = -\frac{1}{2} [N \ln(2\pi) + \ln|\mathbf{E}| + \mathbf{R}^\top \mathbf{E}^{-1} \mathbf{R}] \quad (13)$$

其中, $\mathbf{E} = \mathbf{Q} + \mathbf{A}_x \boldsymbol{\Gamma} \mathbf{A}_x^\top$, $\mathbf{R} = \mathbf{A}_x \mathbf{w} + \mathbf{Q}^{-1}(\mathbf{r} - \mathbf{s})$ 。

最后利用式(13)对 γ_k 求导, 并令其导数为 0, 得到:

$$\frac{\partial \mathcal{N}_{\text{LC}}(\boldsymbol{\gamma})}{\partial \gamma_k} = \frac{1}{2\gamma_k} - \frac{1}{2} \boldsymbol{\psi}_{kk}^{-1} - \frac{1}{2} \mathbf{w}_{\text{best}}^2 = 0 \quad (14)$$

求解式(14)得到:

$$\gamma_k^{\text{new}} = \frac{1 - \gamma_k \boldsymbol{\psi}_{kk}^{-1}}{\mathbf{w}_{\text{best}}^2} \quad (15)$$

基于 $\mathbf{w}$ 和 $\boldsymbol{\gamma}$ 的初始值, 利用式(10)、式(11)、式(12)、式(15)进行优化迭代, 直到 $\boldsymbol{\gamma}$ 收敛到给定的阈值, 优化迭代结束。最后, 将优化后的参数 $\mathbf{w}$ 带入式(5), 即可得到样本属于某个类别的概率(预测值)。

2 基于图正则化的贝叶斯宽度学习系统

传统的 BLS 在非线性映射过程中会破坏数据间的距离信息, 基于流形学习在目标函数中引入图正则化项, 神经网络模型能够更好地保持数据的局部一致性^[19], 即如果两个样本彼此接近, 则它们可能具有相似的条件概率。本文利用参考文献[19]中的内在图(intrinsic graph)在 BBLS 中引入图信息。内在图首先构造一个对称矩阵 $\mathbf{V}$ (式(17)), 它在一定程度上可以反映同类样本之间的相似性; 再利用式(20)得到图正则化项 $\mathbf{E}_g$; 最后将 $\mathbf{E}_g$ 加入目标函数即可利用数据几何信息对输出矩阵 $\mathbf{W}$ 进行约束。充分考虑数据的局部一致性可在一定程度上提高模型的性能, 而在之前的传统 BLS 线性回归过程中, 数据的基本几何结构往往会被忽略。因此, 基于流形正则化的理论, 本文将含有数据几何结构信息的图正则化项加入对数似然函数 J , 则式(9)对数似然函数 J 重写为:

$$J = \sum_{i=1}^N [r_i \ln s_i + (1-r_i) \ln(1-s_i)] - \frac{1}{2} \mathbf{w}^\top \boldsymbol{\Gamma} \mathbf{w} - \lambda_G \mathbf{E}_g + \mathbf{c} \quad (16)$$

其中, λ_G 为正则化参数, $\mathbf{E}_g$ 为图正则化项。

$\mathbf{E}_g$ 被用于表达数据之间的内在关系。为了得出 $\mathbf{E}_g$ 的表达式, 需要定义几个重要的参数。

首先是对称矩阵 $\mathbf{V}$, 它包含样本标签的信息, 有如下定义:

$$V_{ij} = \begin{cases} \frac{1}{N_i}, & \text{label}(x_i) = \text{label}(x_j) \\ 0, & \text{其他} \end{cases} \quad (17)$$

其中, N_i 表示第 i 类样本中的样本个数, $\text{label}(x)$ 表示样本 x 的标签。

定义对角矩阵 $\mathbf{D}$ ，对角线上的元素 D_{ii} 满足：

$$D_{ii} = \sum_j V_{ij} \quad (18)$$

由 $\mathbf{V}$ 和 $\mathbf{D}$ 可得图拉普拉斯矩阵(graph Laplacian matrix) $\mathbf{L}_g$ ：

$$\mathbf{L}_g = \mathbf{D} - \mathbf{V} \quad (19)$$

则有：

$$\mathbf{E}_g = \mathbf{s}^\top \mathbf{L}_g \mathbf{s} \quad (20)$$

此时需要利用式 (16) 重新对 $\mathbf{w}$ 求一阶导数 $\boldsymbol{\varphi}$ 和二阶导数 $\boldsymbol{\psi}$ ：

$$\boldsymbol{\varphi} = \mathbf{A}_x^\top (\mathbf{r} - \mathbf{s}) - \Gamma \mathbf{w} - \lambda_G \mathbf{A}_x^\top \mathbf{Q} \mathbf{L}_g \mathbf{s} \quad (21)$$

$$\boldsymbol{\psi} = -(\mathbf{A}_x^\top \mathbf{Q} \mathbf{A}_x + \Gamma + \mathbf{D}_g) \quad (22)$$

式 (22) 中， $\mathbf{D}_g = \text{Get}_{\text{diag}}(\mathbf{L}_g \mathbf{s} \mathbf{s}^\top) + \text{Get}_{\text{diag}}(\mathbf{S}_d \mathbf{L}_g \mathbf{s}) + \mathbf{Q} \mathbf{L}_g$ ， $\mathbf{S}_d = \text{Get}_{\text{diag}}(\mathbf{s} \odot (1 - \mathbf{s}))$ ， $\text{Get}_{\text{diag}}(\cdot)$ 表示取对角线上的元素组成对角矩阵。

此时，利用重写之后的式 (21)、式 (22)，再结合第 1 节给出的优化算法，即可对模型进行优化求解。

GBBLS 的结构如图 2 所示，其算法如算法 1 所示。

算法 1 基于图正则化的贝叶斯宽度学习算法 (GBBLS)。

输入：数据集 $\mathbf{D} = \{\mathbf{X}, \mathbf{Y}\}$ ， $\mathbf{X}$ 为输入样本， $\mathbf{Y}$ 为样本对应的标签，特征层的特征节点矩阵个数 n ，增强层的增强节点矩阵个数 m 。

输出：数据样本的预测值 $\mathbf{Y}_{\text{pre}}$ ，模型的分类精度 η 。

1. 标准化输入样本 $\mathbf{X}$
2. for $i = 1, 2, \dots, n$ do
3. 随机初始化 $\mathbf{W}_{ei}$ 和 $\boldsymbol{\beta}_{ei}$
4. 利用式 (1) 计算特征节点矩阵 $\mathbf{Z}_i$
5. 得到特征层 $\mathbf{Z} = [\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n]$
6. end for
7. for $j = 1, 2, \dots, m$ do
8. 随机初始化 $\mathbf{W}_{hj}$ 和 $\boldsymbol{\beta}_{hj}$
9. 利用式 (2) 计算增强节点矩阵 $\mathbf{H}_j$
10. 得到增强层 $\mathbf{H} = [\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_m]$
11. end for
12. 得到 $\mathbf{A}_x = [\mathbf{Z}, \mathbf{H}]$ ，作为贝叶斯推断的输入
13. 初始化参数 γ 为 0， $\mathbf{w}$ 为由 0 组成的矩阵
14. 利用式 (19) 计算图拉普拉斯矩阵 $\mathbf{L}_g$
15. for $k=1, 2, \dots, \text{its}$ do
16. 利用式 (21) 计算一阶导数 $\boldsymbol{\varphi}$
17. 利用式 (22) 计算二阶导数 $\boldsymbol{\psi}$
18. 利用式 (12) 更新 $\mathbf{w}$
19. 利用式 (16) 计算对数似然函数 J
20. if 对数似然函数 J 不再减小 then
21. break
22. end if
23. end for
24. 利用式 (15) 更新 γ
25. 重复步骤 15 到 24
26. until γ 收敛
27. $\mathbf{Y}_{\text{pre}} = \text{argmax}(\mathbf{A}_x \mathbf{W})$
28. $\eta = \text{sum}(\mathbf{Y}_{\text{pre}} == \mathbf{Y}) / \text{len}(\mathbf{Y})$

3 实验结果与分析

本节采用 UCI 数据集^[29]和 NORB 数据集^[30]分

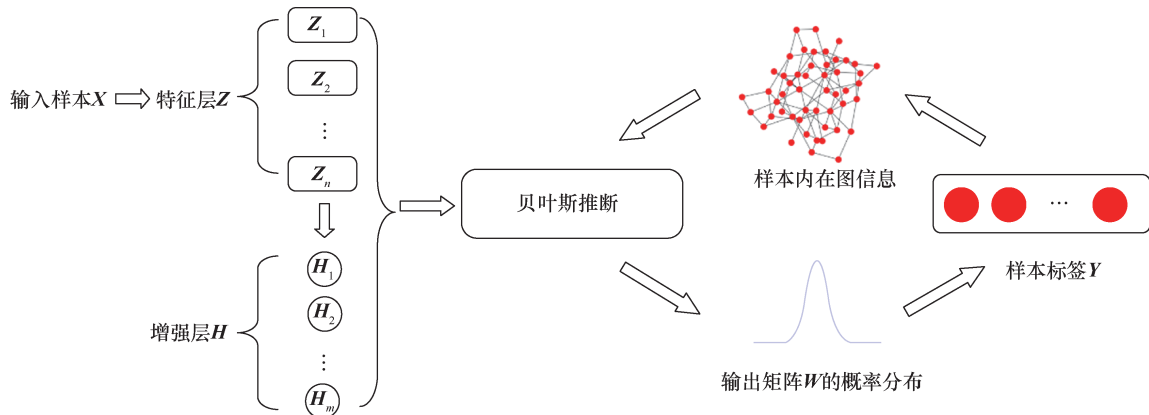


图 2 GBBLS 的结构

别从分类精度和稳定性两方面来评估 GBBLs 算法在处理分类任务时的性能。实验中, BBLs^[18]、BLS、贝叶斯极限学习机 (Bayesian extreme learning machine, BELM)^[24]算法被用来与 GBBLs 算法进行性能对比; 在稳定性实验中, 首先固定 GBBLs 特征层的参数和增强层中增强节点矩阵的个数, 再逐渐增加增强节点的个数, 观察 GBBLs 稳定性的变化情况。所有的实验均在基于 Intel-i5 的 1.60 GHz CPU 和 8 GB 内存的计算机上运行, 软件环境为 MATLAB R2018a。

3.1 数据集介绍

实验所需的数据集为 NORB 数据集和在 UCI 中引用量相对较高的数据集中随机挑选的 10 个数据集 (其中包含 5 个二分类数据集和 5 个多分类数据集), 具体见表 1。

表 1 UCI 中的 10 个数据集信息

数据集	样本数/个	特征数/个	类别/个
Breast Cancer	683	10	2
Climate Model	540	18	2
Fertility	100	10	2
Liver Disorders	345	6	2
King-Rook	3 196	6	2
Forest	523	27	3
Glass	214	10	6
Segmentation	210	19	7
Seeds	210	7	3
Wine	178	13	3

NORB 数据集是一个 3D 物体图像识别数据集, 它从不同的角度对 50 个玩具进行拍摄, 这些玩具分别属于动物、人偶、飞机、卡车、汽车类别。拍摄过程中采用了 2 个照相机、6 种不同的光照条件、9 个特定的拍摄角度和 18 个仰角。每幅图像包含 $32 \times 32 \times 2$ 个像素。训练集和测试集中分别包含了每个类别的 5 个实例。NORB 数据集的部分实例如图 3 所示。

3.2 算法性能评估

第 2 节提出的 GBBLs 模型是基于二分类的。在多分类的问题中, 采用成对耦合的一对一策略^[31], 参考文献[32]表明成对耦合比其他方法能够更好地平衡精度和效率。每次实验均采用 3 重交叉验证。第 3.2.1 节利用 UCI 中的 10 个数据集和 NORB 数据集分别对 GBBLs、BBLs、BLS 和 BELM 进行训练, 并将准确率作为模型分类精度的评估指标; 第

3.2.2 节在保持 GBBLs 的特征节点和特征窗口个数不变的基础上, 逐渐增加增强节点的个数, 观察 GBBLs 在不同数据集上的稳定性变化情况。

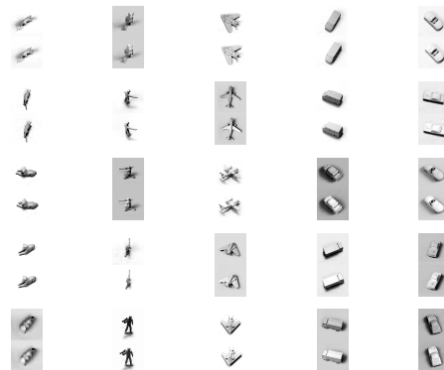


图 3 NORB 数据集的部分实例

3.2.1 精度评估

在 UCI 中的 10 个数据集中寻找 GBBLs、BBLs、BLS、BELM 的最佳准确率, 对于 GBBLs、BBLs 和 BLS, 采用增量学习的方法, 固定增强节点矩阵个数为 1, 在 20 内选择最佳的特征节点 (Nf)、特征节点矩阵 (Nm) 和增强节点 (Ne) 的个数; 对于 BELM, 在 20 内选择最佳的隐藏层节点 (Hn) 个数, 结果见表 2。

从表 2 可看出, 在二分类情况中 (前 5 个数据集), GBBLs 的准确率均高于 BLS 和 BELM。除了 Fertility 数据集, GBBLs 在其余 4 个二分类数据集上的准确率均高于 BBLs。在多分类情况中 (后 5 个数据集), 相比 BLS 和 BELM, GBBLs 在准确率方面有一定的提高, 而与 BBLs 相比, GBBLs 除了在 Segmentation 数据集的分类精度提高 0.9% 外, 在其余多分类数据集上的准确率均和 BBLs 持平。综上所述, 添加图正则化的 BBLs 可在二分类中进一步提高分类能力, 而在多分类中的效果不明显。

在 NORB 数据集上进行实验时, 为了减少计算量, 本文从 NORB 数据集的每个类别中分别随机筛选 1 000 张图像组成含有 5 000 张图像样本的训练集, 再从除去训练集后剩余的 NORB 数据集中每个类别分别随机筛选 200 张图像组成含有 1 000 张图像样本的测试集。在本次实验中, 对比 GBBLs、BBLs、BLS 和 BELM 的准确率。固定 GBBLs、BBLs 和 BLS 的特征节点矩阵和特征节点个数为 10, 同时保持增强节点矩阵的个数为 1, 逐渐增加增强节点的个数, 实验结果见表 3。

表 2

各模型在 UCI 中的 10 个数据集上的准确率

数据集	GBBLS 参数				BBLS 参数				BLS 参数				BELM 参数	
	准确率	Nf	Nm	Ne	准确率	Nf	Nm	Ne	准确率	Nf	Nm	Ne	准确率	Hn
Breast Cancer	79.59%	8	1	6	78.91%	2	6	6	78.91%	2	5	7	74.15%	10
Climate Model	94.44%	5	4	13	94.07%	5	6	8	12.22%	2	3	19	91.85%	15
Fertility	91.22%	8	1	13	91.22%	9	3	1	36.84%	8	2	18	80.7%	5
Liver Disorders	69.83%	6	10	4	68.71%	6	10	2	53.07%	1	4	7	65.92%	7
King-Rook	96.03%	9	9	8	95.96%	10	5	8	94.95%	10	7	18	91.94%	20
Forest	91.5%	5	3	6	91.5%	3	8	1	36.29%	1	1	4	90.35%	20
Glass	92.79%	10	10	3	92.79%	10	10	3	86.49%	8	7	8	83.78%	12
Segmentation	90.9%	8	2	7	90%	8	3	2	90%	3	5	15	80.91%	20
Seeds	96.36%	10	8	1	96.36%	10	8	1	92.27%	1	7	1	94.55%	9
Wine	100%	5	10	5	100%	5	10	5	100%	5	8	3	96.94%	12

表 3 各模型在 NORB 数据集上的准确率

模型	准确率
GBBLS	91.82%
BBLS	91.55%
BLS	92.01%
BELM	68.61%

从表 3 可看出, GBBLS、BBLS 和 BLS 的准确率均在 91% 以上, 而 BELM 的准确率只有 68.61%, BLS 在 NORB 数据集上表现优秀, 而 BELM 则表

现不佳。GBBLS 的准确率比 BLS 略低, 但是比 BBLS 高 0.27%, 说明添加了图正则化后的 BBLS 在图像分类方面的能力有所提高。

3.2.2 稳定性评估

对 GBBLS、BBLS 和 BLS 的稳定性进行分析, 固定特征节点矩阵和特征节点个数为 10, 同时保持增强节点矩阵的个数为 1, 逐渐增加增强节点的个数, 观察 3 种模型在不同数据集上的准确率变化, 结果如图 4 所示。

由图 4 可知, 除了在 Breast Cancer 和 Wine 两个数据集中, BLS 的最高准确率与 BBLS 和 GBBLS

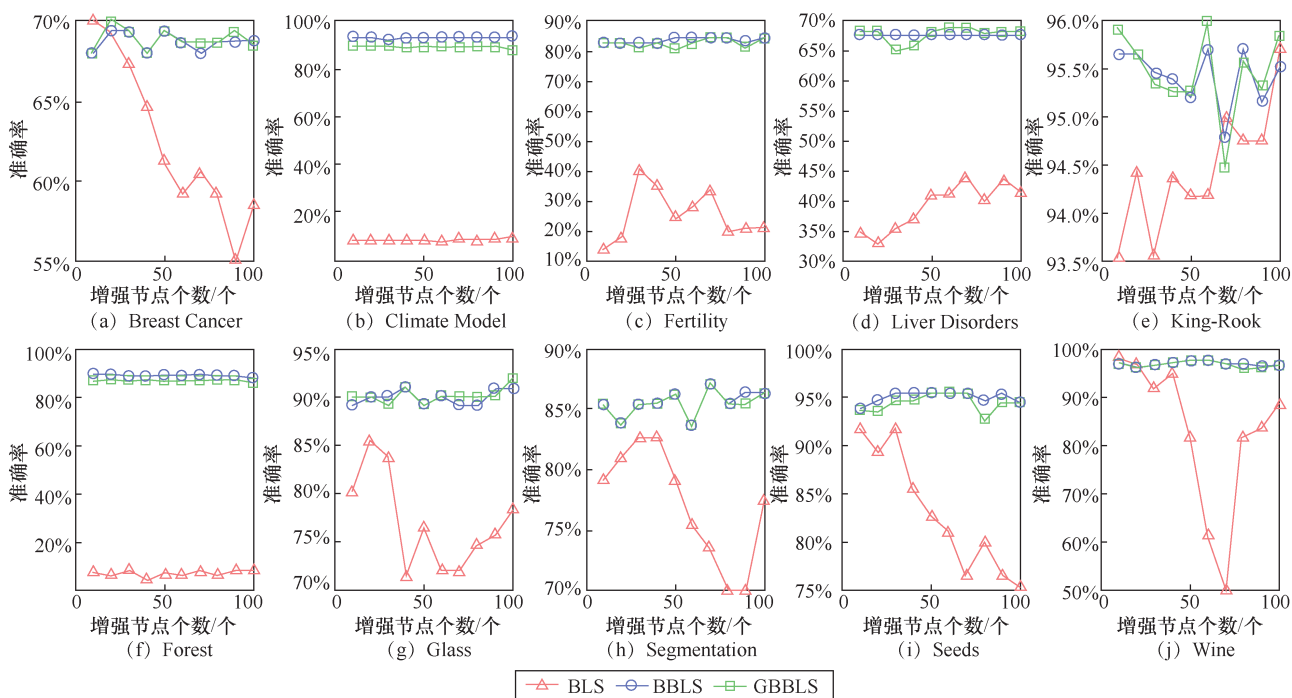


图 4 各模型在 UCI 中的 10 个数据集上的稳定性

相同外,在其他数据集中 BBLs 和 GBBLs 的最高准确率均高于 BLS。在绝大多数数据集中, BBLs 和 GBBLs 的最高准确率和准确率变化趋势几乎相同。无论是在二分类问题还是多分类问题中,与 BLS 相比, GBBLs 和 BBLs 的准确率在增强节点个数不同时均保持在相对较高水平。随着增强节点个数的递增, GBBLs 和 BBLs 在各个数据集上的准确率变化幅度基本保持在 3% 以内,而与 GBBLs 和 BBLs 相比, BLS 在各数据集上准确率的变化幅度较大。由此可见,随着增强节点个数的增加, BLS 出现过拟合现象,导致泛化性能下降,因此准确率出现较大的波动。而 BBLs 在增强节点个数递增的过程中准确率变化幅度较小,进一步验证了贝叶斯推断在缓解 BLS 过拟合现象和提高 BLS 泛化能力方面的有效性。同时, GBBLs 在保持较高的分类精度的基础上,能够维持 BBLs 原有的稳定性。

4 结束语

本文将贝叶斯推断和图正则化同时引入 BLS 中,提出一种基于图正则化的贝叶斯宽度学习系统 GBBLs,有效地解决了标准 BLS 模型中过拟合问题,并提高了 BLS 模型的分类精度。贝叶斯推断可以估计输出的概率分布,防止模型的过拟合;图正则化在损失函数中引入数据的几何结构信息,使模型在拟合中考虑更多的因素。本文在 UCI 中的 10 个数据集和 NORB 数据集上对模型进行了性能评估,实验结果表明, GBBLs 具有与 BBLs 相同的稳定性,同时在准确率上优于 BBLs。在后续的研究工作中,将考虑采用其他的特征工程来代替原有特征层的随机映射过程,提出新模型并将其应用在多模态信息融合领域。

参考文献:

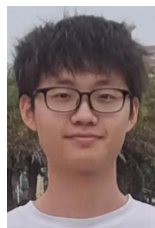
- [1] 张钊. 人工智能进入后深度学习时代[J]. 智能科学与技术学报, 2019, 1(1): 4-6.
ZHANG B. Artificial intelligence is entering the post deep-learning era[J]. Chinese Journal of Intelligent Science and Technology, 2019, 1(1): 4-6.
- [2] 于璠. 新一代深度学习框架研究[J]. 大数据, 2020, 6(4): 69-80.
YU F. Research on the next-generation deep learning framework[J]. Big Data Research, 2020, 6(4): 69-80.
- [3] BROOMHEAD D S, LOWE D. Radial basis functions, multi-variable functional interpolation and adaptive networks[R]. 1988.
- [4] KOHONEN T. Self-organized formation of topologically correct feature maps[J]. Biological Cybernetics, 1982, 43(1): 59-69.
- [5] ELMAN J L. Finding structure in time[J]. Cognitive Science, 1990, 14(2): 179-211.
- [6] ACKLEY D H, HINTON G E, SEJNOWSKI T J. A learning algorithm for boltzmann machines[J]. Cognitive Science, 1985, 9(1): 147-169.
- [7] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.
- [8] YANN L, YOSHUA B. Convolutional networks for images, speech, and time-series[J]. The Handbook of Brain Theory and Neural Networks, 1995: 255-258.
- [9] 黄宜华. 大数据机器学习系统研究进展[J]. 大数据, 2015, 1(1): 35-54.
HUANG Y H. Research progress on big data machine learning system[J]. Big Data Research, 2015, 1(1): 35-54.
- [10] 郑南宁. 人工智能新时代[J]. 智能科学与技术学报, 2019, 1(1): 1-3.
ZHENG N N. The new era of artificial intelligence[J]. Chinese Journal of Intelligent Science and Technology, 2019, 1(1): 1-3.
- [11] 孙长银, 穆朝絮. 多智能体深度强化学习的若干关键科学问题[J]. 自动化学报, 2020, 46(7): 1301-1312.
SUN C Y, MU C X. Important scientific problems of multi-agent deep reinforcement learning[J]. Acta Automatica Sinica, 2020, 46(7): 1301-1312.
- [12] CHEN C L P, LIU Z L. Broad learning system: an effective and efficient incremental learning system without the need for deep architecture[J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(1): 10-24.
- [13] CHEN C L P, LIU Z L, FENG S. Universal approximation capability of broad learning system and its structural variations[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(4): 1191-1204.
- [14] 郑云飞, 陈霸东. 基于最小 p -范数的宽度学习系统[J]. 模式识别与人工智能, 2019, 32(1): 51-57.
ZHENG Y F, CHEN B D. Least p -norm based broad learning system[J]. Pattern Recognition and Artificial Intelligence, 2019, 32(1): 51-57.
- [15] YANG F. A CNN-based broad learning system[C]//Proceedings of 2018 IEEE 4th International Conference on Computer and Communications. Piscataway: IEEE Press, 2018: 2105-2109.
- [16] FENG S, CHEN C L P. Fuzzy broad learning system: a novel neuro-fuzzy model for regression and classification[J]. IEEE Transactions on Cybernetics, 2020, 50(2): 414-424.
- [17] 李润泽, 郭黎, 李保平, 等. 基于模糊宽度学习模型的光伏发电预测方法[J]. 控制工程, 2020, 27(11): 2016-2022.
LI R Z, GUO L, LI B P, et al. A photovoltaic power prediction method based on fuzzy broad learning system[J]. Control Engineering of China, 2020, 27(11): 2016-2022.
- [18] XU L L, CHEN C L P, HAN R Z. Sparse Bayesian broad learning system for probabilistic estimation of prediction[J]. IEEE Access, 2020, 8: 56267-56280.
- [19] JIN J W, LIU Z L, CHEN C L P. Discriminative graph regularized broad learning system for image recognition[J]. Science China Information Sciences, 2018, 61(11): 112209.
- [20] 贾晨, 刘华平, 续欣莹, 等. 基于宽度学习方法的多模态信息融合[J]. 智能系统学报, 2019, 14(1): 150-157.
JIA C, LIU H P, XU X Y, et al. Multi-modal information fusion based on broad learning method[J]. CAAI Transactions on Intelligent Systems, 2019, 14(1): 150-157.

- [21] 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[J]. 控制理论与应用, 2020, 37(12): 2543-2550.
CHU F, SU J M, LIANG T, et al. Sparsity method for network structure of broad learning system based on lasso and elastic net[J]. Control Theory & Applications, 2020, 37(12): 2543-2550.
- [22] YANG L, SONG S J, CHEN C L P. Transductive transfer learning based on broad learning system[C]//Proceedings of 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Piscataway: IEEE Press, 2018: 912-917.
- [23] 李旺, 俞祝良. 宽度学习系统在蘑菇毒性判别中的应用[J]. 现代食品科技, 2019, 35(7): 267-272, 54.
LI W, YU Z L. Application of broad learning system in discrimination of mushroom toxicity[J]. Modern Food Science and Technology, 2019, 35(7): 267-272, 54.
- [24] SORIA-OLIVAS E, GOMEZ-SANCHIS J, MARTIN J D, et al. BELM: Bayesian extreme learning machine[J]. IEEE Transactions on Neural Networks, 2011, 22(3): 505-509.
- [25] LUO J H, VONG C M, WONG P K. Sparse Bayesian extreme learning machine for multi-classification[J]. IEEE Transactions on Neural Networks and Learning Systems, 2014, 25(4): 836-843.
- [26] BISHOP C. Pattern recognition and machine learning[M]. New York: Springer, 2006.
- [27] DIRKSE S P, FERRIS M C. A pathsearch damped Newton method for computing general equilibria[J]. Annals of Operations Research, 1996, 68(2): 211-232.
- [28] BANERJEE K S. Generalized inverse of matrices and its applications[J]. Technometrics, 1973, 15(1): 197.
- [29] KHAN M M R, ARIF R B, SIDDIQUE M Q B, et al. Study and observation of the variation of the variation of accuracies of KNN, SVM, LMNN, ENN algorithms on eleven different datasets from UCI machine learning repository[C]//Proceedings of 2018 4th International Conference on Electrical Engineering and Information & Communication Technology. [S.l.:s.n.], 2018: 124-129..
- [30] LECUN Y, HUANG F J, BOTTOU L. Learning methods for generic object recognition with invariance to pose and lighting[C]//Proceedings of 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2004.
- [31] DEBNATH R, TAKAHIDE N, TAKAHASHI H. A decision based one-against-one method for multi-class support vector machine[J]. Pattern Analysis and Applications, 2004, 7(2): 164-175.
- [32] LI R, WANG X D, LEI L, et al. Probabilities modeling of multi-class based on relevance vector machine[C]//Proceedings of 2016 12th World Congress on Intelligent Control and Automation. Piscataway: IEEE Press, 2016: 2755-2759.

[作者简介]



段俊伟 (1984-), 男, 博士, 暨南大学信息科学技术学院讲师, 主要研究方向为人工智能与大数据、信息融合、图像处理、智能系统及应用。



许林灿 (1999-), 男, 暨南大学信息科学技术学院在读, 主要研究方向为模式识别、深度学习。



全渝娟 (1968-), 女, 博士, 暨南大学信息科学技术学院副教授, 主要研究方向为人工智能与推荐系统、物联网。



陈龙 (1978-), 男, 博士, 澳门大学科技学院副教授, 主要研究方向为计算智能、机器学习方法及应用。



陈俊龙 (1959-), 男, 博士, 欧洲科学院院士, IEEE Fellow, 华南理工大学计算机科学与工程学院院长, 主要研究方向为智能系统与控制、计算智能、混合智能、无人系统、数据科学。