

# 基于自编码器和超图的半监督宽度学习系统

王雪松<sup>1,2</sup>, 张翰林<sup>1,2</sup>, 程玉虎<sup>1,2</sup>

(1. 地下空间智能控制教育部工程研究中心, 江苏徐州 221116; 2. 中国矿业大学信息与控制工程学院, 江苏徐州 221116)

**摘要:** 常规宽度学习系统(Broad Learning System, BLS)通常采用的线性稀疏特征提取方法难以对数据的复杂非线性特征进行有效表征. 此外, 当标记样本量较少时, BLS的泛化性能难以得到保证. 为此, 提出一种基于自编码器和超图的半监督宽度学习系统(Autoencoder and Hypergraph-based Semi-supervised BLS, AH-SBLS). 主要步骤为: 首先, 使用包括标记样本和无标记样本在内的全部样本训练自编码器, 利用训练好的自编码器自动提取数据的复杂非线性特征; 其次, 将自编码器特征层中的特征作为AH-SBLS的特征节点并对其进行宽度拓展; 然后, 构造半监督超图以挖掘标记样本和无标记样本间的高阶流形关系, 并将超图正则项引入宽度学习系统的目标函数中; 最后, 利用岭回归对目标函数进行求解, 实现对无标记样本的类别预测. 在图像分类实验上的结果表明, AH-SBLS能够实现半监督分类且获得较高的分类精度.

**关键词:** 半监督; 宽度学习系统; 自编码器; 超图; 图像分类

中图分类号: TP181

文献标识码: A

文章编号: 0372-2112(2022)03-0533-07

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20210105

## Autoencoder and Hypergraph-Based Semi-Supervised Broad Learning System

WANG Xue-song<sup>1,2</sup>, ZHANG Han-lin<sup>1,2</sup>, CHENG Yu-hu<sup>1,2</sup>

(1. Engineering Research Center of Intelligent Control for Underground Space, Ministry of Education, Xuzhou, Jiangsu 221116, China;

2. School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China)

**Abstract:** The linear sparse feature extraction method used in the classical broad learning system(BLS) is difficult to extract the complex nonlinear features of data effectively. In addition, when the number of labeled samples is small, the generalization ability of BLS cannot be guaranteed. To solve these problems, a novel autoencoder and hypergraph-based semi-supervised BLS(AH-SBLS) is proposed. The main steps of AH-SBLS are described as follows. Firstly, we use all labeled and unlabeled samples to train the autoencoder, and then the trained autoencoder is used to extract the features of input data automatically. Secondly, the extracted features are viewed as the feature nodes of AH-SBLS and are further broadened. In the third step, a semi-supervised hypergraph is constructed to express the high-order manifold relationship between labeled and unlabeled samples, and the hypergraph regularization term is introduced into the objective function of AH-SBLS. Finally, the objective function of AH-SBLS is solved by ridge regression and thus the labels of unlabeled samples can be predicted. Experimental results of image classification show that AH-SBLS can achieve higher classification accuracy in semi-supervised classification tasks.

**Key words:** semi-supervised; broad learning system; autoencoder; hypergraph; image classification

### 1 引言

Chen等<sup>[1]</sup>提出的宽度学习系统(Broad Learning System, BLS), 亦称宽度网络, 是一种由输入层、中间层(包括映射特征和增强节点两个部分)和输出层构成的前馈神经网络. 因其简洁灵活的网络结构、高效的学习

过程等优点, 而被广泛应用于众多领域. Sui等<sup>[2]</sup>使用BLS识别动力学未知的不确定系统, 提出了一种随机有限时间量化控制方法. Chu等<sup>[3]</sup>使用惩罚系数来约束每个样本对模型的贡献, 提出了一种加权BLS并将其用以解决工业过程中的噪声和离群值问题. Han等<sup>[4]</sup>提出

了一种用于大规模混沌时间序列建模的宽度学习系统,解决了动力学系统建模中利用混沌系统演化信息进行时间序列预测的问题。

然而,上述方法均为监督型宽度学习方法,在标记样本有限的情况下,往往存在泛化性不足的问题.半监督学习是一种可以同时使用少量标记样本和大量无标记样本进行学习的方法.因此,半监督宽度学习的研究引起了学者的广泛关注.Kong等<sup>[5]</sup>将常规BLS进行半监督拓展,提出了一种基于类别概率框架<sup>[6]</sup>的半监督BLS(Semi-supervised BLS, SBLS),用于解决标记样本较少情况下的高光谱图像分类问题.但是,SBLS非常依赖伪标签的质量,被错误标记的样本将严重影响算法的性能.Zhao等<sup>[7]</sup>将BLS和流形正则化框架<sup>[8]</sup>相结合,提出了一种基于图的半监督BLS(ss-BLS).ss-BLS通过构造标记样本和无标记样本的拉普拉斯矩阵,将流形正则化项加入BLS的目标函数中,并使用岭回归算法求解构造的目标函数.传统的图模型<sup>[9]</sup>虽然可以在低维空间中保持数据原有的流形结构,但是该模型仅考虑了两个数据间的成对关系,无法准确地表达数据间的复杂高阶流形关系.为此,本文引入超图模型<sup>[10]</sup>,提出一种基于超图的半监督宽度学习系统.此外,常规BLS中通常采用的线性稀疏特征提取方法难以挖掘数据的本质特征,从而限制了BLS性能的发挥.自编码器<sup>[11]</sup>因其在无监督情况下,仍能够学习到数据的复杂非线性特征而被广泛应用于入侵检测<sup>[12]</sup>、目标检测<sup>[13]</sup>和辐射源识别<sup>[14]</sup>等领域.众所周知,良好的特征提取方法对于提高模型的性能至关重要.为此,本文将自编码器与BLS相结合,用以提高BLS对复杂非线性特征的提取能力。

综上,本文提出了一种基于自编码器和超图的半监督宽度学习系统(Autoencoder and Hypergraph-based Semi-supervised Broad Learning System, AH-SBLS),主要工作包括:(1)将超图模型引入到常规监督型BLS中,以充分挖掘包括标记样本和未标记样本在内的所有样本之间的复杂流形关系;(2)利用多层自编码器提取数据的复杂非线性特征,从而进一步提升AH-SBLS的泛化性能。

## 2 基于自编码器和超图的半监督宽度学习系统

如图1所示,AH-SBLS主要包括4个部分:(1)基于自编码器的特征提取.首先,使用含标记样本和无标记样本在内的全部样本训练自编码器;然后,将自编码器特征层中的特征作为AH-SBLS的特征节点;(2)特征增强.对特征节点进行非线性映射以实现宽度拓展,并将得到的增强节点与特征节点进行堆叠;(3)超图构造.同样利用全部样本构造半监督超图,并计算超图的拉普拉斯矩阵;(4)目标函数构造.根据得到的拉普拉斯

矩阵进一步构造超图正则项,并与常规BLS的目标函数相结合.通过对该目标函数进行求解,可以得到输出层权重,进而实现对无标记样本的类别预测。

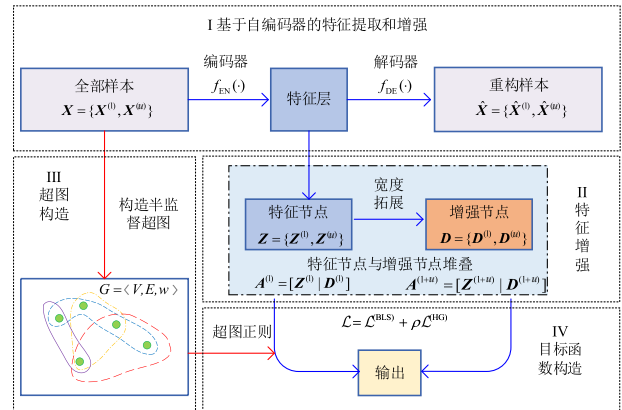


图1 AH-SBLS模型结构图

### 2.1 基于自编码器的特征提取

图1中的自编码器由编码器 $f_{EN}(\cdot)$ 和解码器 $f_{DE}(\cdot)$ 两部分组成,其中,编码器旨在将全部样本映射到特征层以得到编码特征 $Z$ ,解码器的目标为对编码特征进行重构.给定输入样本矩阵 $X$ ,则编码器的计算过程为:

$$Z = f_{EN}(X) = \zeta(XW_{EN} + b_{EN}) \quad (1)$$

其中, $\zeta(\cdot)$ 为编码器的激活函数, $W_{EN}$ 和 $b_{EN}$ 分别为编码器的权重和偏置.类似地,令解码器的权重和偏置分别为 $W_{DE}$ 和 $b_{DE}$ ,激活函数为 $\delta(\cdot)$ ,则解码器的输出可以表示为:

$$\hat{X} = f_{DE}(Z) = \delta(ZW_{DE} + b_{DE}) \quad (2)$$

其中, $\hat{X}$ 为重构样本.自编码器的学习目标为最小化重构样本 $\hat{X}$ 与原始样本 $X$ 之间的差异,即:

$$\arg \min_{W, b} \mathcal{L}(\hat{X}, X) \quad (3)$$

一般而言,自编码器的损失函数通常为均方误差损失.均方误差损失函数可以表示为:

$$\mathcal{L}(\hat{X}, X) = \frac{1}{2} \|\hat{X} - X\|_2^2 \quad (4)$$

其中,参数 $W, b$ 可以使用梯度下降法<sup>[15]</sup>来更新,其更新规则为:

$$W \leftarrow W - \alpha \frac{\partial \mathcal{L}(W, b)}{\partial W} \quad (5)$$

$$b \leftarrow b - \alpha \frac{\partial \mathcal{L}(W, b)}{\partial b} \quad (6)$$

其中, $\alpha$ 为学习率,用来控制参数更新的步长.梯度下降法常用的优化算法有SGD(Stochastic Gradient Descent)、RMSprop(Root Mean Square prop)和Adam(Adaptive moment estimation)<sup>[16]</sup>等,鉴于Adam具有较快的收敛速度,这里将其用于自编码器参数的更新.在

完成自编码器参数的更新后,将自编码器的编码特征  $\mathbf{Z}$  作为 AH-SBLS 模型的特征节点.

## 2.2 特征增强

AH-SBLS 的第 II 部分为特征增强,旨在利用随机生成的权重对特征节点进行映射,实现快速非线性宽度拓展,其计算过程为:

$$\mathbf{d} = \zeta(\mathbf{Z}\mathbf{w}_d + \mathbf{b}_d) \quad (7)$$

其中,  $\zeta(\cdot)$  为 tansig 激活函数,  $\mathbf{w}_d$  为随机的线性稀疏权重,  $\mathbf{b}_d$  为增强节点的偏置. 设共有  $m$  组增强节点, 则  $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_m]$ , 将特征节点和增强节点堆叠, 得到输入样本的特征矩阵  $\mathbf{A} = [\mathbf{Z}|\mathbf{D}]$ .

## 2.3 超图构造

在半监督学习中, 虽然无标记的样本没有标记信息, 但是无标记样本与标记样本包含相同的数据分布信息, 所以利用无标记样本能够帮助建立更具泛化性的分类模型. 为了利用无标记样本, 必须假设无标记样本与类别标记之间的联系. 基于光滑度假设, 图模型被提出用于解决半监督学习问题. 然而, 常规图模型只能描述数据之间简单的二元关联关系, 无法描述数据中复杂的单对多或多对多的多元关联关系, 从而难以对数据的复杂流形结构进行充分描述.

相对于常规图, 超图可以更加准确地描述存在多元关联的对象之间的关系. 在常规图中, 一条边包含两个顶点, 而在超图中, 一条超边可以包含多个顶点, 从而对多个数据点之间的关系进行描述. 给定  $N$  个输入样本, 则输入样本矩阵可以表示为  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$ . AH-SBLS 将每个样本  $\mathbf{x}$  作为超图顶点和一条超边的中心, 使该超边连接该样本最近邻的  $k$  个样本, 以此来构造样本超图模型. 将超图模型表示为三元组  $G = \langle V, E, w \rangle$ , 其中,  $V = \{v_1, v_2, \dots, v_N\}$  为顶点的集合,  $E = \{e_1, e_2, \dots, e_N\}$  为超边的集合,  $w$  为超边的权重集合. 超图模型可以通过构造点边关联矩阵  $\mathbf{H}_{ve} \in \mathbb{R}^{N \times N}$  来建立顶点和超边的关系,  $\mathbf{H}_{ve}$  中的元素可通过下式计算:

$$h(v_i, e_j) = \begin{cases} 1, & \text{if } v_i \in e_j \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

样本之间的距离度量函数通过下式计算:

$$S(i, j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{2\sigma^2}\right) \quad (9)$$

其中,  $\sigma^2 = \frac{1}{N} \sum_{v_i \in e_j} \|\mathbf{x}_i - \mathbf{x}_j\|_2^2$  是控制函数  $S(i, j)$  径向作用范围的超参数. 定义超边权重向量为  $\mathbf{w}_e \in \mathbb{R}^{1 \times N}$ , 则每条超边的权重可以通过下式来计算:

$$\mathbf{w}(e_j) = \sum_{i=1}^N S(i, j), \quad v_i \in e_j \quad (10)$$

定义顶点的度向量为  $\mathbf{d}_v \in \mathbb{R}^{1 \times N}$ , 用于描述每个顶

点的超边权重. 则顶点的度的计算公式为:

$$\mathbf{d}(v_i) = \sum_{j=1}^N w(e_j) h(v_i, e_j), \quad v_i \in e_j \quad (11)$$

定义超边的度向量  $\delta_e \in \mathbb{R}^{1 \times N}$ , 用于描述每条超边包含的顶点个数. 则超边的度计算公式如下:

$$\delta(e_j) = \sum_{i=1}^N h(v_i, e_j), \quad v_i \in e_j \quad (12)$$

通过上述讨论, 可以定义超边的权重矩阵  $\mathbf{W}_e = \text{diag}(\mathbf{w}_e) \in \mathbb{R}^{N \times N}$ , 顶点的度矩阵  $\mathbf{D}_v = \text{diag}(\mathbf{d}_v) \in \mathbb{R}^{N \times N}$  和超边的度矩阵  $\mathbf{D}_e = \text{diag}(\delta_e) \in \mathbb{R}^{N \times N}$  三个对角矩阵. 根据光滑度假设, 距离接近的样本应该拥有相同的标签, 亦即同一条超边应连接相同标签的样本. 构造超图损失函数:

$$\mathcal{L}^{(\text{HG})} = \frac{1}{2} \sum_{e \in E} \sum_{(i, j) \in e} \frac{\mathbf{w}_e}{\delta_e} \|\hat{\mathbf{y}}_i - \hat{\mathbf{y}}_j\|_2^2 \quad (13)$$

其中,  $\hat{\mathbf{y}}$  为模型的预测样本标签. 将式 (13) 表示为矩阵形式:

$$\begin{aligned} \mathcal{L}^{(\text{HG})} &= \frac{1}{2} (\text{Tr}(\hat{\mathbf{Y}}^T \mathbf{D}_v \hat{\mathbf{Y}}) - \text{Tr}(\hat{\mathbf{Y}}^T \mathbf{S} \hat{\mathbf{Y}})) \\ &= \frac{1}{2} \text{Tr}(\hat{\mathbf{Y}}^T \mathbf{L}_{\text{hyper}} \hat{\mathbf{Y}}) \end{aligned} \quad (14)$$

其中, 超图拉普拉斯矩阵  $\mathbf{L}_{\text{hyper}}$  的计算公式为:

$$\mathbf{L}_{\text{hyper}} = \mathbf{D}_v - \mathbf{S} = \mathbf{D}_v - \mathbf{H}_{ve} \mathbf{W}_e \mathbf{D}_e^{-1} \mathbf{H}^T \quad (15)$$

## 2.4 目标函数构造及求解

给定  $l$  个标记样本  $\mathbf{X}^{(l)} \in \mathbb{R}^{l \times r}$  和  $u$  个无标记样本  $\mathbf{X}^{(u)} \in \mathbb{R}^{u \times r}$ , 则全部样本为  $\mathbf{X}^{(l+u)} \in \mathbb{R}^{(l+u) \times r}$ . 其中,  $r$  为样本的维数. 将标记样本的标签表示为矩阵形式  $\mathbf{Y}^{(l)}$ , 模型的预测标签矩阵记为  $\hat{\mathbf{Y}}$ , 标记样本的特征矩阵记为  $\mathbf{A}^{(l)}$ , 全部样本的特征矩阵记为  $\mathbf{A}^{(l+u)}$ . 常规 BLS 模型的目标函数为:

$$\mathcal{L}^{(\text{BLS})} = \left\| \mathbf{Y}^{(l)} - \mathbf{A}^{(l)} \mathbf{W}_o \right\|_2^2 + \lambda \left\| \mathbf{W}_o \right\|_2^2 \quad (16)$$

其中,  $\mathbf{W}_o$  为输出层权重,  $\lambda \in [0, 1]$  为结构风险项系数. 将式 (14) 作为超图正则项与式 (16) 相结合, 得到 AH-SBLS 的目标函数:

$$\begin{aligned} \mathcal{L}(\mathbf{W}_o) &= \mathcal{L}^{(\text{BLS})} + \rho \mathcal{L}^{(\text{HG})} \\ &= \left\| \mathbf{Y}^{(l)} - \mathbf{A}^{(l)} \mathbf{W}_o \right\|_2^2 + \lambda \left\| \mathbf{W}_o \right\|_2^2 + \rho \text{Tr}(\hat{\mathbf{Y}}^T \mathbf{L}_{\text{hyper}} \hat{\mathbf{Y}}) \\ &= \left\| \mathbf{Y}^{(l)} - \mathbf{A}^{(l)} \mathbf{W}_o \right\|_2^2 + \lambda \mathbf{W}_o^T \mathbf{W}_o \\ &\quad + \rho \text{Tr}(\mathbf{W}_o^T \mathbf{A}^{(l+u)T} \mathbf{L}_{\text{hyper}} \mathbf{A}^{(l+u)} \mathbf{W}_o) \\ &= (\mathbf{Y}^{(l)} - \mathbf{A}^{(l)} \mathbf{W}_o)^T (\mathbf{Y}^{(l)} - \mathbf{A}^{(l)} \mathbf{W}_o) + \lambda \mathbf{W}_o^T \mathbf{W}_o \\ &\quad + \rho (\mathbf{W}_o^T \mathbf{A}^{(l+u)T} \mathbf{L}_{\text{hyper}} \mathbf{A}^{(l+u)} + \mathbf{W}_o^T \mathbf{A}^{(l+u)T} \mathbf{L}_{\text{hyper}} \mathbf{A}^{(l+u)}) \\ &= \mathbf{Y}^{(l)T} \mathbf{Y}^{(l)} - \mathbf{Y}^{(l)T} \mathbf{A}^{(l)} \mathbf{W}_o - \mathbf{W}_o^T \mathbf{A}^{(l)T} \mathbf{Y}^{(l)} \\ &\quad + \mathbf{W}_o^T \mathbf{A}^{(l)T} \mathbf{A}^{(l)} \mathbf{W}_o + \lambda \mathbf{W}_o^T \mathbf{W}_o + \rho \mathbf{W}_o^T \mathbf{A}^{(l+u)T} \mathbf{L}_{\text{hyper}} \mathbf{A}^{(l+u)} \end{aligned} \quad (17)$$

其中,  $\rho \in [0, 1]$  为超图正则项的缩放因子. 令  $\frac{\partial \mathcal{L}(\mathbf{W}_o)}{\partial \mathbf{W}_o} =$

0,即可求解式(17),得到 AH-SBLS 模型输出层权重的计算公式:

$$\mathbf{W}_o = (\mathbf{A}^{(0)\top} \mathbf{A}^{(0)} + \lambda \mathbf{I} + \rho \mathbf{A}^{(l+u)\top} \mathbf{L}_{\text{hyper}} \mathbf{A}^{(l+u)})^{-1} \mathbf{A}^{(0)\top} \mathbf{Y}^{(0)} \quad (18)$$

## 2.5 算法流程

所提 AH-SBLS 的流程如算法 1 所示:

### 算法 1 基于自编码器和超图的半监督宽度学习系统

**输入:** 有标记样本  $\mathbf{X}^{(0)}$  及对应标签  $\mathbf{Y}^{(0)}$ , 无标记样本  $\mathbf{X}^{(u)}$ , 测试集样本  $\mathbf{X}_{\text{te}}$ , 超参  $\alpha, k, \lambda, \rho$  以及宽度学习系统的特征节点和增强节点数.

**输出:** 测试集样本标签

步骤 1: 根据式(1), 计算标记样本和无标记样本的特征节点  $\mathbf{Z}^{(l+u)}$ ;

步骤 2: 根据式(7), 计算标记样本和无标记样本的增强节点  $\mathbf{D}^{(l+u)}$ ;

步骤 3: 将特征节点与增强节点进行堆叠, 得到输出特征  $\mathbf{A}^{(l+u)}$ ;

步骤 4: 根据式(8)和(10)~(12), 构造超图模型;

步骤 5: 根据式(15), 计算超图的拉普拉斯矩阵  $\mathbf{L}_{\text{hyper}}$ ;

步骤 6: 根据式(17), 构造 AH-SBLS 模型的目标函数;

步骤 7: 根据式(18), 计算 AH-SBLS 模型的输出层权重  $\mathbf{W}_o$ ;

步骤 8: 计算测试样本  $\mathbf{X}_{\text{te}}$  在 AH-SBLS 中的输出特征  $\mathbf{A}_{\text{te}}$ ;

步骤 9: 根据  $\hat{\mathbf{Y}}_{\text{te}} = \mathbf{A}_{\text{te}} \mathbf{W}_o$ , 得到测试样本的标签  $\hat{\mathbf{Y}}_{\text{te}}$ .

## 3 实验与分析

### 3.1 数据集

为验证 AH-SBLS 的性能,选择人手写数字数据集 MNIST (Mixed National Institute of Standards and Technology) 和三维玩具模型数据集 NORB (New York University Object Recognition Benchmark) 进行实验. MNIST 数据集<sup>[17]</sup>由来自 250 位不同的人手写的 0~9 共十个数字 70000 幅图像构成, 每张均为 28×28 像素的灰度图像. NORB 数据集<sup>[18]</sup>是一个比 MNIST 特征更为多样的数据集, 它包含 50 种玩具模型的图像, 共 5 类: 四足动物、人物、飞机、卡车和汽车. NORB 数据集中的所有 48600 幅图像是用两个摄像机在 6 个光照条件, 9 个仰角和 18 个方位角对玩具模型拍摄的灰度图像, 由 2×32×32 个像素组成.

### 3.2 实验结果及分析

为分别考查自编码器和超图对 AH-SBLS 性能的影响, 将 AH-SBLS 及其 3 个特例包括 A-BLS1 (仅使用单层自编码器进行特征提取的监督型学习)、A-BLS (仅使用多层自编码器进行特征提取的监督型学习)、AH-SBLS1 (使用单层自编码器进行特征提取且利用超图实现半监督学习) 与下述 6 种方法进行对比, 包括: 堆栈自编码器 (SAE)<sup>[11]</sup>、抗噪堆栈自编码器 (SDA)<sup>[19]</sup>、深度置信网络 (DBN)<sup>[20]</sup>、深度玻尔兹曼机 (DBM)<sup>[21]</sup>、BLS<sup>[1]</sup> 和 ss-BLS<sup>[7]</sup>. 实验设置: (1) 对于监督型方法, 分别从 MNIST 和 NORB 数据集中选择 6000 个和 4860 个样本作为训练样本, 剩余样本用于测试; (2) 对于半监督型方法, 从 MNIST 数据集中每类选择 100 个标记样本和 900 个无标

记样本, 从 NORB 数据集中每类选择 600 个标记样本和 4260 个无标记样本, 其余样本用于测试; (3) SAE、SDA、DBN、DBM 和 BLS 的超参数及实验结果直接取自文献 [1]. 所有宽度学习系统的超参数均采用网格搜索法确定, 具体设置情况如表 1、表 2 所示, 其中“-”表示该参数不取值. 在 AH-SBLS 中, 参数  $\lambda$  和  $\rho$  分别控制岭回归正则项和超图正则项在目标函数中的权重. 如果  $\lambda=0$ , 岭回归方法就退化为了最小二乘. 如果  $\rho=0$ , 超图正则项就失去了作用.  $k$  为超图的最近邻样本数. A-BLS 和 AH-SBLS 均采用一个 5 层自编码器, 在 MNIST 和 NORB 数据集上每层的节点个数分别为 600-500-50-500-600、1800-550-300-550-1800; (4) 所有实验均在配备有 Intel-i7 2.4 GHz CPU, GTX1080Ti GPU 的计算机上进行. 为消除随机因素的影响, 所有实验重复 5 次并取平均值.

表 1 宽度学习系统的超参数设置 (MNIST 数据集)

模型	$\lambda$	$\rho$	$k$	$\alpha$	特征节点数	增强节点数
BLS	$10^{-8}$	-	-	-	100	11000
A-BLS1	$10^{-14}$	-	-	0.005	100	5000
A-BLS	1	-	-	$10^{-3}$	2250	3000
ss-BLS	1	$10^{-2}$	-	-	100	5000
AH-SBLS1	$10^{-3}$	$10^{-2}$	7	0.005	100	5000
AH-SBLS	1	$10^{-2}$	7	$10^{-3}$	2250	3000

表 2 宽度学习系统的超参数设置 (NORB 数据集)

模型	$\lambda$	$\rho$	$k$	$\alpha$	特征节点数	增强节点数
BLS	$10^{-12}$	-	-	-	1000	9000
A-BLS1	1	-	-	$10^{-4}$	900	3000
A-BLS	1	-	-	0.003	5000	3000
ss-BLS	$10^{-3}$	$10^{-4}$	-	-	900	3000
AH-SBLS1	$10^{-2}$	$10^{-2}$	5	$10^{-4}$	900	3000
AH-SBLS	$10^{-2}$	$10^{-2}$	5	0.003	5000	3000

表 3、表 4 给出了各方法在 MNIST 和 NORB 数据集上的图像分类实验结果对比, 可以看出:

(1) 在所有监督型宽度学习方法中, A-BLS 的分类精度最高, A-BLS1 次之. 这印证了 BLS 线性稀疏特征提取方法存在表征瓶颈, 并且由自编码器提取的非线性特征要比线性稀疏特征更能反映出样本在特征空间中的真实分布. 在所有监督型学习方法中, A-BLS 同样取得了最高的分类精度. 这是因为多层的自编码器可以学习到样本不同层次的特征表示, 通过组合不同层次的特征, A-BLS 能够实现特征空间与标记空间之间更准确的映射;

(2) 从时间上分析, 在所有监督型宽度学习方法中, BLS、A-BLS1 和 A-BLS 的耗时均相对较少. 这是因为宽度网络的输出层参数可以直接通过广义逆矩阵的计算得到, 且模型的结构较为简洁. 同为宽度学习方法, BLS 在 MNIST 数据集上的耗时最短, 其次为 A-BLS1. 这

表 3 图像分类结果对比(MNIST数据集)

模型	SAE	SDA	DBN	DBM	BLS
精度(%)	98.60	98.72	98.87	99.05	98.74
耗时(s)	36448.40	37786.03	53219.77	121455.69	<b>29.92</b>
模型	A-BLS1	A-BLS	ss-BLS	AH-SBLS1	AH-SBLS
精度(%)	98.93	<b>99.22</b>	92.54	93.43	<b>94.57</b>
耗时(s)	55.29	63.04	238.41	255.88	264.04

表 4 图像分类结果对比(NORB数据集)

模型	SAE	SDA	DBN	DBM	BLS
精度(%)	86.26	87.62	88.47	89.65	89.27
耗时(s)	60504.34	65747.69	87280.42	182183.53	41.47
模型	A-BLS1	A-BLS	ss-BLS	AH-SBLS1	AH-SBLS
精度(%)	92.37	<b>92.86</b>	73.25	87.56	<b>89.29</b>
耗时(s)	<b>32.36</b>	67.86	226.97	264.58	311.95

是因为训练自编码器比直接提取线性稀疏特征需要更多的时间. 在 NORB 数据集上, A-BLS1 为最快速的模型, 比 BLS 耗时更少. 这是因为 A-BLS1 使用了更少的特征节点和增强节点.

(3)同为半监督宽度学习方法,相较于 ss-BLS, AH-SBLS1(特征节点数与增强节点数与 ss-BLS 相同)能够以很少的训练时间为代价,取得更高的分类精度. 同时, AH-SBLS 模型仍是精度最高的模型,这一结果充分说明了超图模型比常规图模型更能反映样本之间的真实联系.

为分析标记样本的数量对 AH-SBLS 分类精度的影响,通过设置不同标记样本个数进行实验,并选择 ss-BLS 和 AH-SBLS1 作为对比方法. 在 MNIST 数据集中,标记样本的数量由每类 10 逐渐递增至 100. 在 NORB 数据集中标记样本的数量由每类 100 逐渐递增至 600. 实验结果如图 2 所示,由图可知:

(1)在两个数据集上,随着标记训练样本量的增加,ss-BLS、AH-SBLS1 和 AH-SBLS 的分类精度均逐渐提高. 这是因为对于半监督学习方法来说,利用更多的标记样本能够帮助学习到泛化能力更强的分类模型;

(2)标记训练样本数量相同的情况下,AH-SBLS 能够取得最高的分类精度,而且拥有相同特征节点和增强节点数的 AH-SBLS1 的分类精度仍高于 ss-BLS. 这印证了相比于常规图,超图能够通过标记样本和无标记样本之间的复杂流形关系的建模,帮助获取更高的分类精度.

### 3.3 特征 t-SNE 可视化

为了直观地展示 AH-SBLS 模型中自编码器的特征提取效果,在 MNIST 和 NORB 两个数据集上使用 t-SNE 方法<sup>[22]</sup>对自编码器提取的特征进行了可视化. 该实验在两个数据集上每类均随机选取 200 个样本. 对于拥

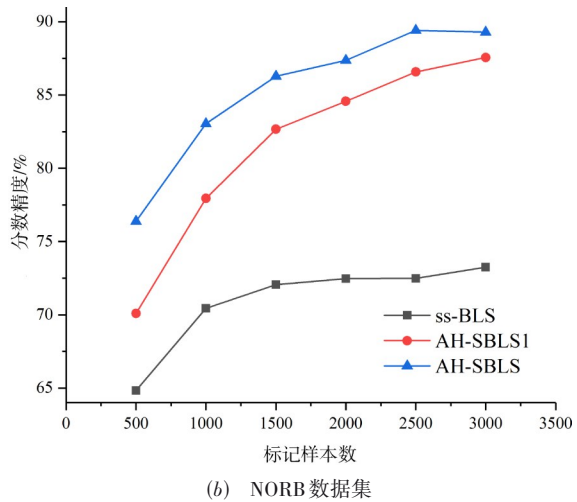
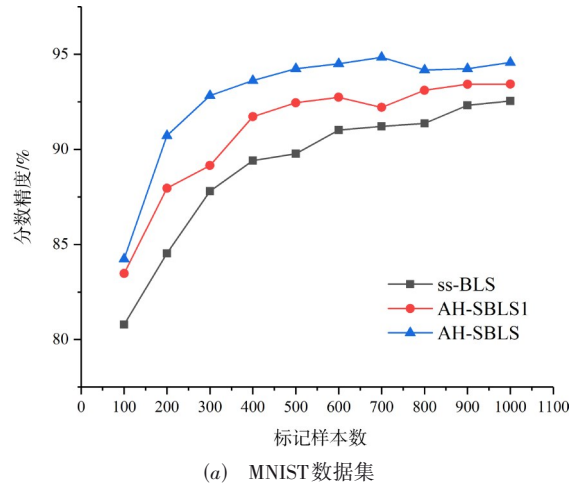


图 2 标记样本数量对分类精度的影响

有多层特征层的 AH-SBLS 模型,只显示中间维数最少的特征层的可视化结果. 实验结果如图 3、图 4 所示.

可以通过比较原始图像的分布和模型提取特征的分布来评价模型的特征提取效果. 由图 3、图 4 可知:(1)在 MNIST 数据集上,虽然大多数数字都可以被 AH-SBLS 模型准确地地区分开,但无论是原始数据还是模型提取的特征,数字“4”和“9”在图中的分布均非常接近,说明这两个数字具有非常相似的特征,将它们区分开来的难度较大. 同样在图 4 中,汽车和卡车的分布也具有这样的特点. 汽车和卡车均属于“车”,拥有较多的相似特征,将它们区分开来的难度较大也是符合直觉的;(2)相较于原始数据和 BLS 提取的特征, AH-SBLS 提取得到的特征不同类别之间有了更大的间距,同类特征分布地更加密集,并且离群的数据点较少. 这说明利用 AH-SBLS 提取的多层非线性特征具有更强的可分性,进而帮助宽度学习系统实现更加准确的类别预测.

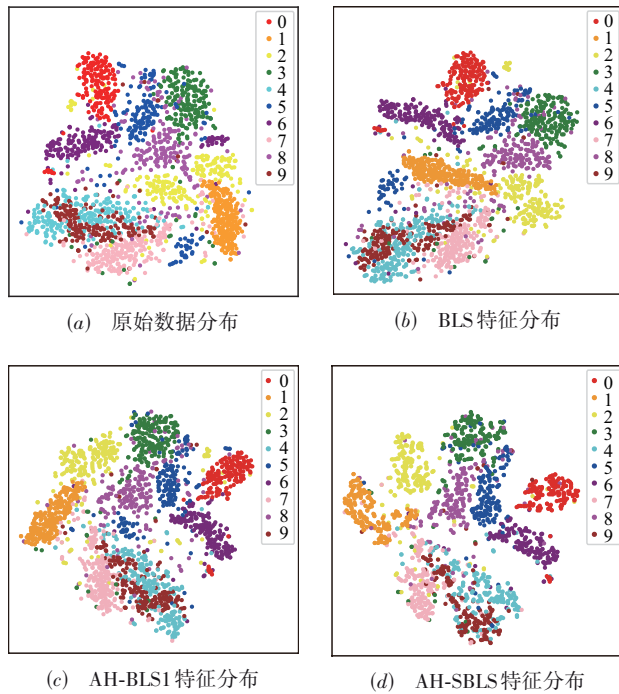


图3 特征t-SNE可视化(MNIST数据集)

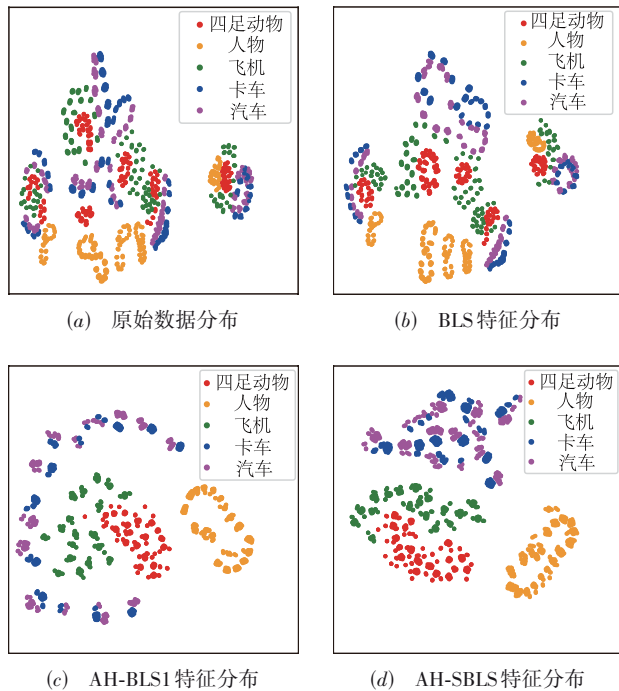


图4 特征t-SNE可视化(NORB数据集)

## 4 结论

随着计算机技术的迅猛发展,人们将获取越来越多的无标记数据,而有标记样本需要人工对样本进行标记,因而有标记样本的获取相对比较困难而且代价昂贵.在标记样本极少的情况下,常规的宽度学习系统

难以取得良好的表现,很容易陷入对少量样本的过拟合或者及对目标任务的欠拟合.因此,将少量的标记样本信息和大量的无标记样本信息加以综合利用的半监督学习有着广泛的需求.本文提出的AH-SBLS模型将超图结构引入到常规监督型BLS中,不仅能够实现半监督学习,而且充分考虑了数据之间的复杂多元关系.此外,多层自编码器的使用有助于提取到输入数据的本征特征,从而进一步提升半监督BLS的泛化性能.MNIST和NORB两个数据集上的实验结果验证了AH-SBLS的可行性和有效性.

## 参考文献

- [1] CHEN C L P, LIU Z. 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(99): 10-24.
- [2] SUI S, CHEN C L P, TONG S, FENG S. Finite-time adaptive quantized control of stochastic nonlinear systems with input quantization: a broad learning system based identification method[J]. IEEE Transactions on Industrial Electronics, 2020, 67(10): 8555-8565.
- [3] CHU F, LIANG T, CHEN C L P, WANG X, MA X. Weighted broad learning system and its application in nonlinear industrial process modeling[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(8): 3017-3031.
- [4] HAN M, LI W, FENG S, QIU T, CHEN C L P. Maximum information exploitation using broad learning system for large-scale chaotic time-series prediction[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(6): 2320-2329.
- [5] KONG Y, WANG X, CHENG Y, CHEN C L P. Hyperspectral imagery classification based on semi-supervised broad learning system[J]. Remote Sensing, 2018, 10(5): 685.
- [6] SHAO Y, SANG N, GAO C, MA L. Spatial and class structure regularized sparse representation graph for semi-supervised hyperspectral image classification[J]. Pattern Recognition, 2018, 81: 81-94.
- [7] ZHAO H, ZHENG J, DENG W, SONG Y. Semi-supervised broad learning system based on manifold regularization and broad network[J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2020, 67(3): 983-994.
- [8] BELKIN M, NIYOGI P, SINDHWANI V. Manifold regularization: a geometric framework for learning from labeled and unlabeled examples[J]. Journal of Machine

- Learning Research, 2006, 7: 2399-2434.
- [9] ANIS A, EL G A, AVESTIMEHR A S, ORTEGA A. A sampling theory perspective of graph-based semi-supervised learning[J]. IEEE Transactions on Information Theory, 2019, 65(4): 2322-2342.
- [10] 冀中, 樊帅飞. 基于超图排序算法的视频摘要[J]. 电子学报, 2017, 45(5): 1035-1043.  
JI Zhong, FAN Shuai-fei. Video summarization based on hypergraph ranking[J]. Acta Electronica Sinica, 2017, 45(5): 1035-1043. (in Chinese)
- [11] HINTON G E, SALAKHUTDINOV R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.
- [12] 高妮, 高岭, 贺毅岳, 王海. 基于自编码网络特征降维的轻量级入侵检测模型[J]. 电子学报, 2017, 45(3): 730-739.  
GAO Ni, GAO Ling, HE Yi-yue, WANG Hai. A light-weight intrusion detection model based on autoencoder network with feature reduction[J]. Acta Electronica Sinica, 2017, 45(3): 730-739. (in Chinese)
- [13] 翟颖, 陈渤. 基于稳健变分自编码模型的雷达高分辨距离像目标识别算法[J]. 电子学报, 2020, 48(6): 1149-1155.  
ZHAI Ying, CHEN Bo. Study of radar high range resolution profiles target recognition based on auto-encoder[J]. Acta Electronica Sinica, 2020, 48(6): 1149-1155. (in Chinese)
- [14] 李东瑾, 杨瑞娟, 李晓柏, 董睿杰. 基于栈式稀疏降噪自编码网络的辐射源调制识别[J]. 电子学报, 2020, 48(6): 1198-1204.  
LI Dong-jin, YANG Rui-juan, LI Xiao-bo, DONG Rui-jie. Emitter signal modulation recognition based on stacked sparse denoising auto-encoders[J]. Acta Electronica Sinica, 2020, 48(6): 1198-1204. (in Chinese)
- [15] RUMMELHART D E, HINTON G E, WILLIAMS R J. Learning internal representations by error propagation[J]. Readings in Cognitive Science, 1988, 323(2): 399-421.
- [16] QIU D, ZHENG L, ZHU J, et al. Multiple improved residual networks for medical image super-resolution[J]. Future Generation Computer Systems, 2021: 200-208.
- [17] LECUN Y, BOTTOU L. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [18] LECUN Y, HUANG F J, BOTTOU L. Learning methods for generic object recognition with invariance to pose and lighting[C]//Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004. Washington, DC, USA: IEEE, 2004: 97-104.
- [19] VINCENT P, LAROCHELLE H, BENGIO Y, MANZAGOL P A. Extracting and composing robust features with denoising autoencoders[C]//Proceedings of the 25th International Conference on Machine Learning. New York: Academic Press, 2008: 1096-1103.
- [20] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.
- [21] SALAKHUTDINOV R, HINTON G E. Deep boltzmann machines[J]. Journal of Machine Learning Research, 2009, 5(2): 1967-2006.
- [22] HAN X, GHAEMI M S, ANDO K, PETERSON L S, GAUDILLIERE B, et al. Differential dynamics of the maternal immune system in healthy pregnancy and pre-eclampsia[J]. Frontiers in Immunology, 2019, 10: 1305. DOI: 10.3389/fimmu.2019.01305.

#### 作者简介



**王雪松** 女, 1974年生, 安徽泗县人. 现为地下空间智能控制教育部工程研究中心主任, 中国矿业大学信息与控制工程学院教授、博士生导师. 研究领域为机器学习及模式识别.  
E-mail: wangxuesongcumt@163.com



**张翰林** 男, 1998年生, 山东聊城人. 现为中国矿业大学信息与控制工程学院硕士研究生. 研究领域为图像处理与分析.  
E-mail: 18252111368@163.com



**程玉虎(通讯作者)** 男, 1973年生, 安徽淮南人. 现为中国矿业大学信息与控制工程学院教授、博士生导师. 研究领域为机器学习及模式识别.  
E-mail: chengyuhu@163.com