

随机权分布对极限学习机性能影响的实验研究

翟俊海¹ 臧立光² 张素芳³

(河北省机器学习与计算智能重点实验室河北大学数学与信息科学学院 保定 071002)¹

(河北大学计算机科学与技术学院 保定 071002)² (中国气象局气象干部培训学院河北分院 保定 071000)³

摘要 极限学习机是一种训练单隐层前馈神经网络的算法,它随机初始化输入层的权值和隐层节点的偏置,用分析的方法确定输出层的权值。极限学习机具有学习速度快、泛化能力强的特点。很多研究都用服从 $[-1,1]$ 区间均匀分布的随机数初始化输入层权值和隐层节点的偏置,但没有对这种随机初始化合理性的研究。用实验的方法对这一问题进行了研究,分别研究了随机权服从均匀分布、高斯分布和指数分布对极限学习机性能的影响。研究发现随机权的分布对极限学习机的性能的确有影响,对于不同的问题或不同的数据集,服从 $[-1,1]$ 区间均匀分布的随机权不一定是最优的选择。研究结论对从事极限学习机研究的人员具有一定的借鉴作用。

关键词 随机权分布,极限学习机,均匀分布,高斯分布,指数分布

中图分类号 TP181

文献标识码 A

DOI 10.11896/j.issn.1002-137X.2016.12.022

Experimental Research on Effects of Random Weight Distributions on Performance of Extreme Learning Machine

ZHAI Jun-hai¹ ZANG Li-guang² ZHANG Su-fang³

(Hebei Province Key Laboratory of Machine Learning and Computational Intelligence, College of Mathematics and Information Science, Hebei University, Baoding 071002, China)¹

(College of Computer Science and Technology, Hebei University, Baoding 071002, China)²

(Hebei Branch of Meteorological Cadres Training Institute, China Meteorological Administration, Baoding 071000, China)³

Abstract Extreme learning machine (ELM) is an algorithm for training single-hidden layer feed-forward neural networks (SLFNs). ELM firstly employs randomization method to generate the input weights and hidden nodes biases, and then determines the output weights analytically. ELM has fast learning speed and good generalization ability. All methods published in literatures usually initialize the weights of input layer and biases of hidden nodes with a uniform distribution over the interval $[-1,1]$. However, there are no studies on the rationality of this setting in literatures. This paper investigated this problem by experimental approach. Specifically, the effects of random weights with uniform distribution, Gaussian distribution and exponential distribution were studied. We found that the random weight distributions do have impact on the performance of the extreme learning machine. For different problems or different data sets, the random weights with uniform distribution in $[-1,1]$ are not necessarily optimal. The results of this paper can be used for reference by the researchers engaged in the study of ELM.

Keywords Random weight distributions, Extreme learning machine, Uniform distribution, Gaussian distribution, Exponential distribution

1 引言

极限学习机(Extreme Learning Machine, ELM)是 Huang 等人^[1]于2004年提出的用于训练单隐层前馈神经网络(Single-hidden Layer Feed-forward Neural Networks, SLFNs)的学习算法。ELM用随机化的方法确定输入层和隐层之间的权值及隐层节点的偏置,用分析的方法确定隐层和

输出层之间的权值。ELM不需要迭代调整权参数,具有学习速度快、泛化能力强等特点,是近几年机器学习领域的研究热点,已成功应用于模式识别、生物信息处理、决策支持等领域^[2,3]。

自2004年Huang等人提出ELM以来,作为一种新的学习理论,ELM引起了很多科研人员的研究兴趣,已成为近几年机器学习领域的一个研究热点。研究人员从不同角度对

到稿日期:2016-01-27 返修日期:2016-05-27 本文受国家自然科学基金项目(71371063),河北省自然科学基金项目(F2013201220),河北省高等学校科学技术研究重点项目(ZD20131028)资助。

翟俊海(1964-),男,博士,教授,CCF会员,主要研究方向为数据挖掘与模式识别, E-mail: mczjh@126.com; 臧立光(1990-),男,硕士生,主要研究方向为数据挖掘; 张素芳(1966-),女,硕士,副教授,主要研究方向为机器学习。

ELM进行了扩展,提出了许多扩展的ELM学习模型。2006年,Huang等人^[4]将ELM由批处理学习模式推广到适用于模式学习的增量ELM(Incremental ELM, I-ELM),并在文献[4]中证明了随机生成隐含层结点的单隐含层前馈神经网络具有精确插值能力和一致逼近能力,该结论为ELM奠定了理论基础,也使ELM的发展进入了一个崭新的阶段。在文献[4]的基础上,Huang等人又提出了增强型ELM(Enhanced Incremental-ELM, EI-ELM)^[5]和凸增量ELM(Convex Incremental-ELM, CI-ELM)^[6]。Liang等人^[7]于2006年提出了适用于在线序列学习的ELM(Online Sequential-ELM, OS-ELM)。Liu等人^[8]于2008年提出了基于核函数的ELM(Extreme Support Vector Machine, ESVM),ESVM将传统SVM中的核函数替换成ELM核函数,这一工作具有重要的理论价值,它首次将ELM和SVM联系起来。在此工作基础上,Huang等人^[9]针对分类问题,研究了ELM中的优化方法,得出了重要的结论:“在ELM学习框架下,SVM的最大间隔性质和前馈神经网络的连接权最小范数性质是一致的”。Huang等人^[10]还进一步研究了ELM、最小范数支持向量机和近似支持向量机3种模型之间的关系,并得出结论:“ELM提供了一个统一的学习框架,可将最小范数支持向量机和近似支持向量机统一到这一框架下,而且从优化的观点,与最小范数支持向量机和近似支持向量机相比,ELM需要的约束条件更弱”。Emilio等人^[11]将贝叶斯线性回归模型与ELM相结合提出了贝叶斯ELM(Bayes-ELM, B-ELM),B-ELM能自动产生网络输出的置信区间,可以降低学习模型过拟合的风险。结合ELM的模型选择,Feng等人^[12]提出了误差最小化ELM(Error Minimized-ELM, EM-ELM);Rong等人^[13]提出了剪枝ELM(Pruned-ELM, P-ELM);Miche等人^[14]提出了最优剪枝ELM(Optimally Pruned-ELM, OP-ELM)。

近几年,ELM的应用研究也引起了研究人员的浓厚兴趣。Mohammed等人^[15]成功地将其应用于人脸识别,提出了基于ELM和多维主成分分析的人脸识别方法,获得了非常好的识别效果。Iosifidis等人^[16]提出了基于最小类方差的ELM人为识别。

文献[15]对近几年ELM的国内外发展现状作了全面的分析和综述,但没有随机权分布对ELM性能影响的研究。研究中都用服从 $[-1, 1]$ 区间均匀分布的随机数初始化输入层权和隐含层结点的偏置,但没有对这种随机初始化合理性的研究。本文用实验的方法对这一问题进行了研究。具体地,分别研究了随机权服从均匀分布、高斯分布和指数分布对极限学习机性能的影响。研究发现随机权的分布对极限学习机的性能的确有影响,对于不同的问题或不同的数据集,服从 $[-1, 1]$ 区间均匀分布的随机权不一定是最佳的选择。

2 极限学习机

极限学习机(Extreme Learning Machine, ELM)是训练如图1所示的单隐含层前馈神经网络(SLFNs)的算法。给定训练集 $D = \{(x_i, y_i) | x_i \in R^d, y_i \in R^k, i = 1, 2, \dots, n\}$,含有 m 个隐含层结点的SLFNs可用式(1)表示。

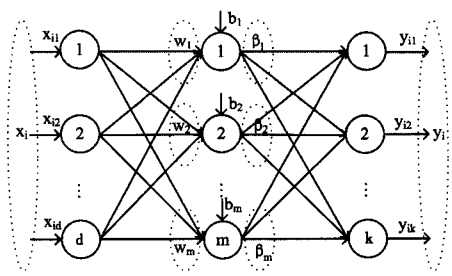


图1 单隐含层前馈神经网络

$$f(x_i) = \sum_{j=1}^m \beta_j g(w_j \cdot x_i + b_j), i = 1, 2, \dots, n \quad (1)$$

其中, $w_j = (w_{j1}, w_{j2}, \dots, w_{jd})^T$ 是连接输入层结点和隐含层第 j 个结点的权向量, b_j 是隐含层第 j 个结点的偏置。在ELM中, w_j 和 b_j ($j = 1, 2, \dots, m$)是用随机化的方法产生的。 $\beta_j = (\beta_{j1}, \beta_{j2}, \dots, \beta_{jk})^T$ 是连接隐含层第 j 个结点和输出层结点的权向量, β_j ($j = 1, 2, \dots, m$)是用分析的方法确定的。具体地,可以用给定的训练集 D 通过最小二乘拟合来估计,即 β_j 满足式(2):

$$f(x_i) = \sum_{j=1}^m \beta_j g(w_j \cdot x_i + b_j) = y_i, i = 1, 2, \dots, n \quad (2)$$

式(2)可写成下面的矩阵形式。

$$H\beta = Y \quad (3)$$

其中,

$$H = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \dots & g(w_m \cdot x_1 + b_m) \\ \vdots & & \vdots \\ g(w_1 \cdot x_n + b_1) & \dots & g(w_m \cdot x_n + b_m) \end{bmatrix} \quad (4)$$

$$\beta = (\beta_1^T, \beta_2^T, \dots, \beta_m^T)^T \quad (5)$$

$$Y = (y_1^T, y_2^T, \dots, y_n^T)^T \quad (6)$$

H 是SLFNs隐含层输出矩阵,它的第 j 列是相对于输入样例 x_1, x_2, \dots, x_n 第 j 个隐含层结点的输出向量,它的第 i 行是相对于输入样例 x_i 的隐含层的输出向量。如果隐含层结点数 m 等于训练集中包含的不同样例个数 n ,那么矩阵 H 是一个可逆方阵。在这种情况下,SLFNs可以以零误差逼近这 n 个样例。但一般情况下,隐含层结点数 m 远远小于训练集中包括的样例个数 n 。因此,矩阵 H 不是方阵,方程(3)一般很难得到精确解。但是,方程(3)可以用式(7)的解来逼近。

$$\min_{\beta} \| H\beta - Y \| \quad (7)$$

方程(7)的最小范数最小二乘解可表示为式(8)。

$$\hat{\beta} = H^+ Y \quad (8)$$

其中, H^+ 是矩阵 H 的Moore-Penrose伪逆矩阵。ELM算法可描述如下^[1]。

算法 极限学习机

输入:训练集 D ,激活函数 g ,隐含层结点数 m

输出:权矩阵 $\hat{\beta}$

Step1 随机初始化输入层权 w_j 和隐含层偏置 b_j ($j = 1, 2, \dots, m$);

Step2 计算隐含层输出矩阵 H ;

Step3 计算输出层权矩阵 $\hat{\beta} = H^+ Y$ 。

3 随机权分布对极限学习机性能的影响

在极限学习机及其各种变体中,输入层的随机权和隐含层结点的偏置都用服从 $[-1, 1]$ 区间均匀分布的随机数初始

化,目前还没有对这种随机初始化合理性的研究。这种设置是最优选择吗?采用其他分布的随机权初始化是否可行?如正态分布或指数分布的随机权,如果可行,对于不同的数据集,用哪种分布的随机权初始化是最优选择呢?本文用实验的方法对这些问题进行了研究。具体地,分别研究了随机权服从均匀分布、高斯分布和指数分布对极限学习机性能的影响。实验环境是 Intel(R) Core(TM) i5-2400 CPU @ 3.10 GHz 处理器,4GB 内存,32 位 Windows 操作系统,Matlab R2013a。实验所用数据集包括 16 个 UCI 数据集^[17]和 2 个真实数据集。2 个真实数据集分别是 CT 和 RenRu。CT 数据集是从河北大学附属医院得到的 212 幅脑 CT 图像经特征提取后得到的,212 幅图像中正常 CT 图像 170 幅,病变 CT 图像 42 幅。用于表示图像的特征有 35 个,其中 10 个对称特征,9 个纹理特征以及包括均值、方差、熵等在内的 16 个统计特征。RenRu 数据集是由河北大学智能图文实验室创建的,由 92 个汉字“人”和 56 个汉字“入”构成,每个汉字用 26 个特征来描述。实验所用数据集的基本信息如表 1 所列。

表 1 实验所用数据集的基本信息

数据集	样例数	属性数	类别数
Banknote	1372	5	2
Cancer	699	11	2
Fertility	100	10	2
Forest	325	28	4
Heart	2126	22	3
Image	194	19	4
Ionosphere	337	34	2
Iris	150	4	3
Mammo	961	6	2
Optical	1797	64	10
Parkinsons	195	22	2
Pen	7494	16	10
Seeds	210	8	3
Vertebral	308	7	2
Vowel	990	14	11
Wine	130	13	3
CT	221	36	2
RenRu	148	26	2

实验中,70%的样例用作训练集,30%的样例用作测试集。对于不同的数据集,用文献[18]中的方法确定 ELM 网络隐含层结点的个数。选择的 ELM 网络隐含层结点数及相应的最优测试精度具体如表 2 所列。

表 2 ELM 网络隐含层结点的个数及测试精度

数据集	结点数	测试精度
Banknote	150	0.99
Cancer	60	0.95
Fertility	20	0.80
Forest	80	0.81
Heart	80	0.88
Image	35	0.90
Ionosphere	50	0.82
Iris	40	0.92
Mammographic	100	0.79
Optical	100	0.93
Parkinsons	60	0.88
Pen	400	0.99
Seeds	60	0.92
Vertebral	400	0.91
Vowel	400	0.91
Wine	40	0.95
CT	40	0.89
RenRu	40	0.89

对于均匀分布,令随机变量在区间 $[-w, w]$ 取值, w 是一

个正实数。对于正态分布,令随机变量在区间 $[-w, w]$ 取值,并令均值 $\mu=0$,方差 σ 取不同值,以观察方差对极限学习机性能的影响。对于指数分布,令随机变量在区间 $[0, w]$ 取值,参数 θ 取不同值。对于上述 3 种分布, w 均从 0.25 变化到 3.0,步长为 0.25。

实验 1 概率密度函数中的参数对结果的影响

在本实验中,仅以正态分布为例,研究了概率密度函数中参数对极限学习机性能的影响。设随机变量 $X \sim N(\mu, \sigma^2)$, X 的概率密度函数为 $p(x)$ 。众所周知, $p(x)$ 的图形是关于均值 μ 对称的钟形图形,如图 2 所示。实际上,对于正态分布,均值决定了图形的位置,方差决定了图形的扁平程度。不失一般性,在实验中令均值 $\mu=0$,这样权值分布在以原点为中心的对称区间上。方差 σ 分别取值 1.0、1.5 和 2.0。在实验中,分别用区间 $[-w, w]$ 上服从 3 种正态分布的权值初始化单隐含层前馈神经网络,记录不同 w 所对应的 ELM 的测试精度,可得不同数据集上的性能曲线。由于篇幅所限,只列出了在两个 UCI 数据集(Pen 和 Pima)和一个真实数据集(CT)上的实验结果,如图 3—图 5 所示。

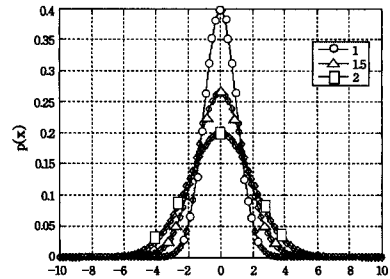


图 2 概率密度函数 $p(x)$ 的图形

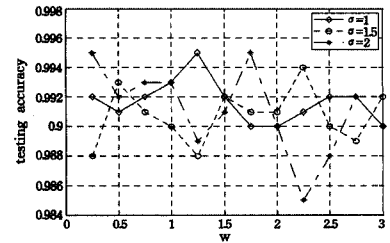


图 3 在数据集 Pen 上的实验结果

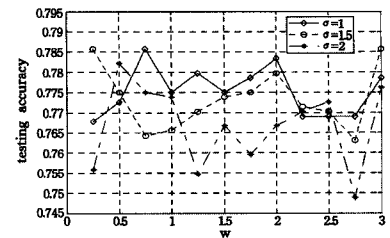


图 4 在数据集 Pima 上的实验结果

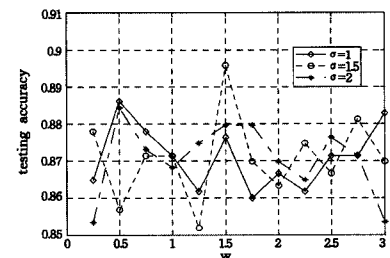


图 5 在数据集 CT 上的实验结果

从图3—图5的实验结果可以看出,方差变化对精度影响不大,与其他大多数数据集上的实验结果类似。对于指数分布,参数 θ 取不同的值对测试精度的影响也不大。由于篇幅有限,不再列出指数分布的实验结果。

实验2 3种分布对极限学习机性能的影响

在实验2中,研究了3种分布对极限学习机性能的影响。在正态分布中,选择参数 $\mu=0, \sigma=1$ 进行实验。在指数分布 $p(x)=\frac{1}{\theta}e^{-\frac{x}{\theta}}(x>0)$ 中,选择参数 $\theta=1$ 进行实验。对于3种分布, w 均从0.25变化到3.0,步长为0.25。分别用区间 $[-w, w]$ 上服从3种不同分布的权值初始化单隐层前馈神经网络,记录不同 w 所对应的ELM的测试精度,可得不同数据集上的性能曲线。实验结果如图6—图23所示。

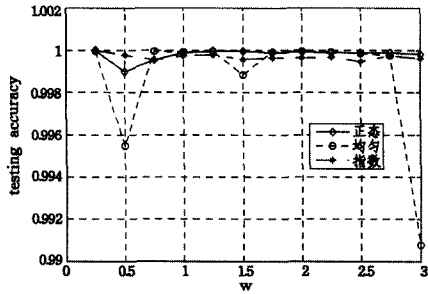


图6 在数据集 Banknote 上的实验结果

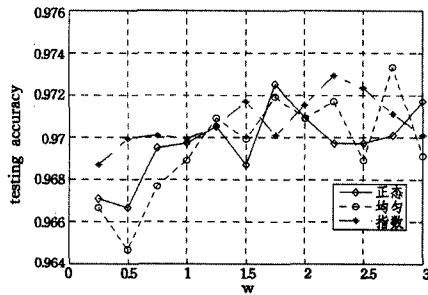


图7 在数据集 Cancer 上的实验结果

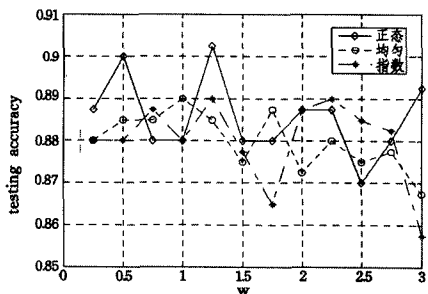


图8 在数据集 Fertility 上的实验结果

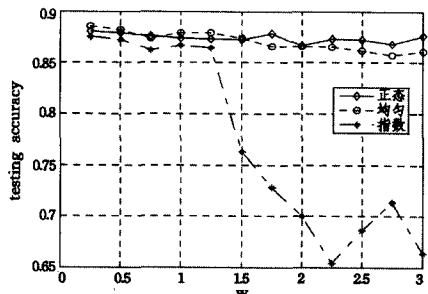


图9 在数据集 Forest 上的实验结果

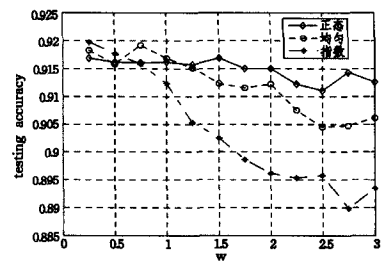


图10 在数据集 Heart 上的实验结果

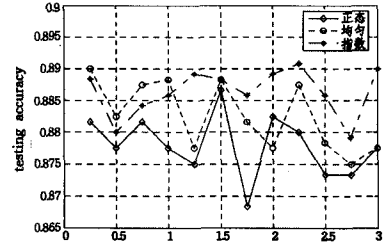


图11 在数据集 Image 上的实验结果

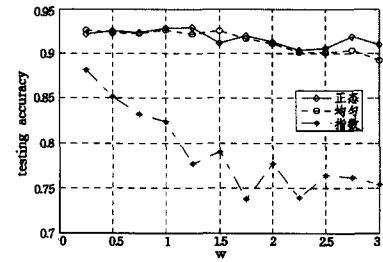


图12 在数据集 Ionosphere 上的实验结果

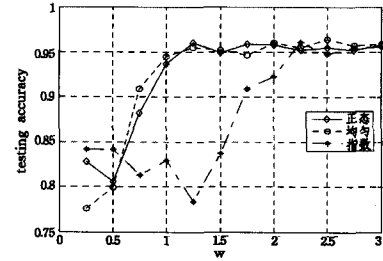


图13 在数据集 Iris 上的实验结果

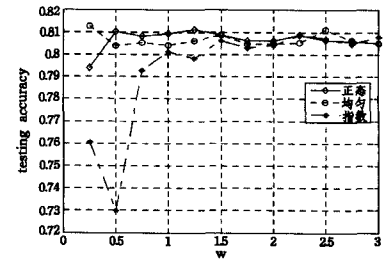


图14 在数据集 Mammo 上的实验结果

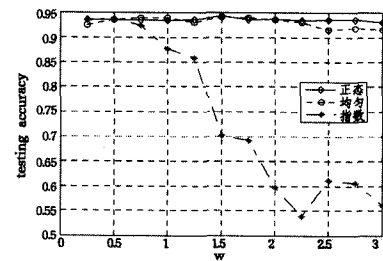


图15 在数据集 Optical 上的实验结果

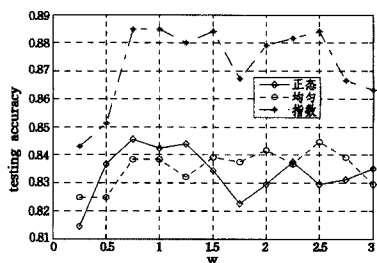


图 16 在数据集 Parkinsons 上的实验结果

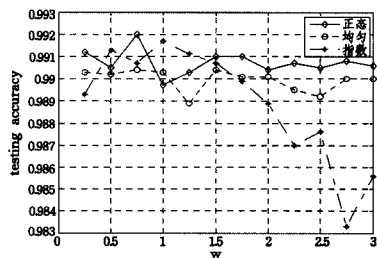


图 17 在数据集 Pen 上的实验结果

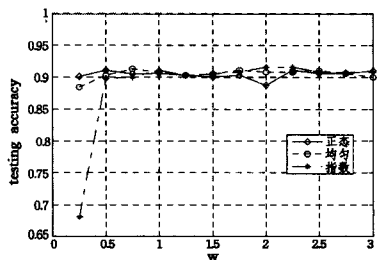


图 18 在数据集 Seeds 上的实验结果

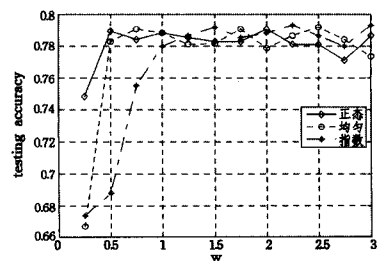


图 19 在数据集 Vertebral 上的实验结果

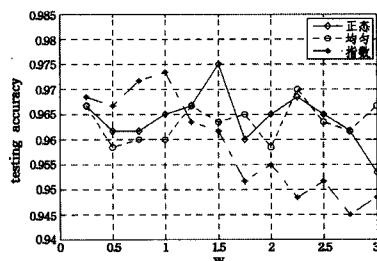


图 20 在数据集 Vowel 上的实验结果

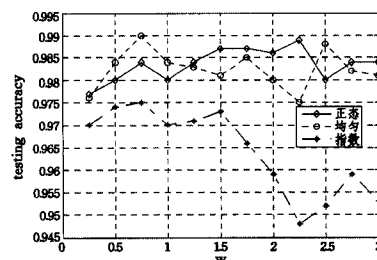


图 21 在数据集 Wine 上的实验结果

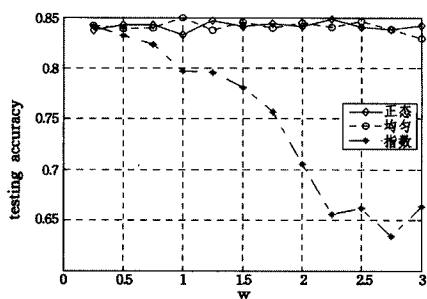


图 22 在数据集 CT 上的实验结果

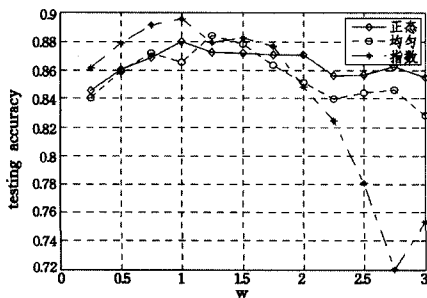


图 23 在数据集 RenRu 上的实验结果

由图 6—图 23 所示的实验结果可以看出:1)服从正态分布随机权和服从均匀分布随机权对测试精度的影响差别不大,但服从指数分布随机权对测试精度的影响差别较大,如在 Forest、Optical、CT、RenRu 4 个数据集上的实验结果就是如此。2)对于不同的数据集,服从 $[-1, 1]$ 区间均匀分布的随机权不一定是最佳的选择。例如,对于数据集 Fertility,用 $[-1.25, +1.25]$ 区间中服从高斯分布的随机权初始化得到的测试精度最高。3)采用服从其他分布(如高斯分布和指数分布)的随机权初始化也是可行的。例如,对于数据集 Banknote, w 取 $[0.75, 2.75]$ 区间中的任何值,对于 3 种分布均能得到类似的测试精度。但对于数据集 Forest, w 取 $[0.25, 3.00]$ 区间中的任何值,对于均匀分布和高斯分布都能得到类似的测试精度。但对指数分布, w 取 $[0.25, 1.25]$ 区间中的值是合适的,取大于 1.25 的值就不合适了。

为了进一步验证结论的可信性,下面用成对 T-检验对结论 1)进行统计分析,其它结论可类似分析。在实验中,显著性水平设置为 0.05。对 4 个数据集 (Forest、Optical、CT、RenRu)中的每一个数据集,运用服从不同概率分布的随机数初始化输入层权值和隐含层节点的偏置,用 ELM 训练单隐含层前馈神经网络。分别用成对 T-检验分析了 10-折交叉验证运行 5 次和 5-折交叉验证运行 10 次的测试精度,分析的结果分别列于表 3 和表 4 中。在表 3 和表 4 中,P-值 1 表示用服从均匀分布和高斯分布的随机数初始化输入层权值和隐含层节点的偏置得到的结果;P-值 2 表示用服从均匀分布和指数分布的随机数初始化输入层权值和隐含层节点的偏置得到的结果。对于不同的数据集, w 的取值也列于表 3 和表 4 中。

表 3 运用 10-折交叉验证运行 5 次对结论 1)的统计分析

数据集	均匀-高斯		均匀-指数	
	P-值 1	w	P-值 2	w
Forest	0.3170	1	5.1752E-12	2
Optical	0.2809	1	9.9790E-05	1
CT	0.1071	1	5.1084E-10	1
RenRu	0.2822	1	4.1511E-08	3

[34] Vapnik V N. 统计学习理论[M]. 许建华, 张学工, 译. 北京: 电子工业出版社, 2009

[35] Ding Shi-fei, Qi Bing-juan, Tan Hong-yan. An overview on theory and algorithm of support vector machines [J]. Journal of University of Electronic Science and Technology of China, 2011, 40(1): 2-10 (in Chinese)
丁世飞, 齐丙娟, 谭红艳. 支持向量机理论与算法研究综述[J]. 电子科技大学学报, 2011, 40(1): 2-10

[36] Platt J C, Cristianini N, Shawe-Taylor J. Large Margin DAGs for multiclass classification[J]. Advances in Neural Information Processing Systems, 2000, 12(3): 547-553

[37] Manivannan K, Aggarwal P, Devabhaktuni V, et al. Particulate

matter characterization by gray level co-occurrence matrix based support vector machines [J]. Journal of hazardous materials, 2012, 223-224(2): 94-103

[38] Hsu C W, Lin C J. A Comparison of Methods for Multiclass Support Vector Machines [J]. IEEE Transactions on Neural Networks, 2002, 13(3): 415-425

[39] Kre U, et al. Pairwise classification and support vector machines [C]//Advances in Kernel Methods. MIT Press, 1999: 255-268

[40] Larochelle H, Bengio Y. Classification using Discriminative Restricted Boltzmann Machines [C]//Proceedings of the 25th International Conference on Machine Learning, 2008. Helsinki, Finland, 2008: 1-8

(上接第 129 页)

表 4 运用 5-折交叉验证运行 10 次对结论 1) 的统计分析

数据集	均匀-高斯		均匀-指数	
	P-值 1	w	P-值 2	w
Forest	0.1428	1	5.9547E-08	2
Optical	0.3272	1	2.3619E-04	1
CT	0.1820	1	1.6185E-05	1
RenRu	0.2936	1	1.2497E-07	3

从表 3 和表 4 的 P-值 1 可以看出, 用服从均匀分布和高斯分布的随机数初始化输入层权值和隐含层节点的偏置得到的测试精度没有本质的区别, 这一结论成立的概率至少为 0.95。从表 3 和表 4 的 P-值 2 可以看出, 用服从均匀分布和高斯分布的随机数初始化输入层权值和隐含层节点的偏置得到的测试精度有本质的区别, 这一结论成立的概率至少为 0.95。因此, 本文得出的结论 1) 具有可信性。

结束语 本文研究了在极限学习机中随机权分布对学习系统测试精度的影响。具体地, 用实验方法分别研究了随机权服从均匀分布、高斯分布和指数分布对极限学习机性能的影响。在 18 个数据集上进行了大量的实验研究, 通过对实验结果的分析发现了有趣的结论。例如, 用服从 $[-1, +1]$ 区间均匀分布的随机数初始化单隐含层前馈神经网络不一定是最优的选择。本文得出的结论对从事极限学习机网络研究的人员具有一定的借鉴作用。

参 考 文 献

[1] Huang Guang-bin, Zhu Qin-yu, Siew C K. Extreme learning machine: A new learning scheme of feedforward neural networks [C]//Proceedings of International Joint Conference on Neural Networks(IJCNN2004). 2004, 985-990

[2] Huang Guang-bin, Wang Dian-hui, Lan Yuan. Extreme learning machines: a survey [J]. International Journal of Machine Learning and Cybernetics, 2011, 2(2): 107-122

[3] Huang Guang-bin, Zhu Qin-yu, Siew C K. Extreme learning machine: Theory and applications [J]. Neurocomputing, 2006, 70(1-3): 489-501

[4] Huang Guang-bin, Chen L, Siew C K. Universal approximation using incremental constructive feedforward networks with random hidden nodes [J]. IEEE Transactions on Neural Networks, 2006, 17(4): 879-892

[5] Huang Guang-bin, Chen Lei. Enhanced random search based incremental extreme learning machine [J]. Neurocomputing, 2008, 71(16-18): 3460-3468

[6] Huang Guang-bin, Chen Lei. Convex incremental extreme learn-

ing machine [J]. Neurocomputing, 2007, 70(16): 3056-3062

[7] Liang Nan-ning, Huang Guang-bin, Saratchandran P, et al. A fast and accurate on-line sequential learning algorithm for feed-forward networks [J]. IEEE Transactions on Neural Networks, 2006, 17(6): 1411-1423

[8] Liu Qiu-ge, He Qing, Shi Zhong-zhi. Extreme support vector machine classifier [C]//Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining. Springer-Verlag, 2008: 222-233

[9] Huang Guang-bin, Ding Xiao-jian, Zhou Hong-ming. Optimization method based extreme learning machine for classification [J]. Neurocomputing, 2010, 74(1-3): 155-163

[10] Huang Guang-bin, Zhou Hong-ming, Ding Xiao-jian, et al. Extreme Learning Machine for Regression and Multiclass Classification [J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B, 2012, 42(2): 513-529

[11] Emilio S O, Juan G S, Martin J D, et al. BELM: Bayesian Extreme Learning Machine [J]. IEEE Transactions on Neural Networks, 2011, 22(3): 505-509

[12] Feng Guo-rui, Huang Guang-bin, Lin Qing-ping, et al. Error Minimized extreme learning machine with growth of hidden nodes and incremental learning [J]. IEEE Trans. Neural Networks, 2009, 20(8): 1352-1357

[13] Rong Hai-jun, Ong Y S, Tan A H, et al. A fast pruned-extreme learning machine for classification problem [J]. Neurocomputing, 2008, 72(1-3): 359-366

[14] Miche Y, Sorjamaa A, Bas P, et al. OP-ELM: Optimally pruned extreme learning machine [J]. IEEE Transactions on Neural Networks, 2010, 21(1): 158-162

[15] Mohammed A A, Minhas R, Jonathan Q M, et al. Human face recognition based on multidimensional PCA and extreme learning machine [J]. Pattern Recognition, 2011, 44(10/11): 2588-2597

[16] Iosifidis A, Tefas A, Pitas I. Minimum Class Variance Extreme Learning Machine for Human Action Recognition [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2013, 23(11): 1968-1979

[17] Frank A, Asuncion A. UCI machine learning repository [OL]. <http://archive.ics.uci.edu/ml>

[18] Zhai Jun-hai, Li Ta, Zhai Meng-yao, et al. Experimental Research on Random Mapping Function in ELM Algorithm [J]. Computer Engineering, 2012, 38(20): 164-168 (in Chinese)
翟俊海, 李塔, 翟梦尧, 等. ELM 中随机映射作用的实验研究 [J]. 计算机工程, 2012, 38(20): 164-168