

# ELMs 和 SVMs 在多分类问题上的泛化性能比较

卢欣欣<sup>1,2</sup> 潘丽平<sup>3</sup>

<sup>1</sup>(周口师范学院计算机科学与技术学院 河南 周口 466001)

<sup>2</sup>(农产品质量安全追溯技术河南省工程实验室 河南 周口 466001)

<sup>3</sup>(周口科技职业学院 河南 周口 466001)

**摘要** 多分类问题是机器学习、数据挖掘领域的重要研究内容。在文本分类、语音识别、图像识别、基因检测等方面有广泛的应用。通过在 UCI 数据集对极限学习机算法 ELMs (ELM, KELM) 和支持向量机算法 SVMs (SVM, LSSVM) 在多分类问题上的表现进行详细比较,得出以下结论: ELMs 相较于 SVM 在多分类问题上有更高的分类准确率,而且随着分类数目的增加,ELMs 的泛化能力相较于 SVM 提高越多,但是 ELMs 对于 LSSVM 并没有得到上述结论;ELMs 相较于 SVMs 对数据的类别数目不敏感,分类准确率随类别数目增加下降不明显;ELMs 相较于 SVMs 在多分类问题上所需计算代价更小,且拥有更快的学习和训练速度,适用于多分类问题。

**关键词** 极限学习机(ELM) 核极限学习机(KELM) 支持向量机(SVM) 最小二乘支持向量机(LSSVM)  
多分类问题 泛化能力

中图分类号 TP311

文献标识码 A

DOI:10.3969/j.issn.1000-386x.2019.10.045

## COMPARING ELMs AND SVMs GENERALIZATION PERFORMANCE ON MULTI-CLASS CLASSIFICATION PROBLEM

Lu Xinxin<sup>1,2</sup> Pan Liping<sup>3</sup>

<sup>1</sup>(School of Computer Science and Technology, Zhoukou Normal University, Zhoukou 466001, Henan, China)

<sup>2</sup>(Traceability Technology of Agricultural Product Quality and Safety, Henan Engineering Laboratory, Zhoukou 466001, Henan, China)

<sup>3</sup>(Zhoukou Vocational College of Science and Technology, Zhoukou 466001, Henan, China)

**Abstract** Multi-class classification problem is an important issue in machine learning and data mining. It is widely used in text classification, speech recognition, image recognition and gene detection. In this paper, we compared the performance of ELMs (ELM, KELM) and SVMs (SVM, LSSVM) on multi-class classification problem in detail through UCI dataset. We draw the following conclusion: compared with SVM, ELMs has higher classification accuracy on multi-class classification problem, and with the increase of classification number, the generalization ability of ELMs is more improved than that of SVM, but ELMs does not get the above conclusion for LSSVM; compared with SVMs, ELMs is insensitive to the number of classes, and the classification accuracy decreases little with the increase of the number of classes; compared with SVMs, ELMs has less computational cost, and has faster learning and training speed, which is suitable for multi-class classification problem.

**Keywords** ELM KELM SVM LSSVM Multi-class classification problem Generalization performance

## 0 引言

由于科技的飞速发展,猛增的数据量和大幅提升

的计算性能使得多分类问题越来越多地代替二分类问题,并在文本分类识别<sup>[1]</sup>、语音识别<sup>[2]</sup>、医学图像识别<sup>[3]</sup>等方面取得了良好的应用前景,已经成为模式识别、数据挖掘等领域的重要研究内容。

SVM 是 Vapnik<sup>[4]</sup> 基于统计学习理论提出的最优间隔分类器,其良好的泛化能力、全局最优解和解决非线性问题的能力使其成为应用最为广泛的分类器<sup>[5-6]</sup>。但原始 SVM 并不适用于多分类问题,Weston<sup>[7]</sup> 最先改进 SVM 使其成功应用在多分类问题中。最小二乘支持向量机(LSSVM)<sup>[8]</sup> 是 SVM 的简化和改进形式,避免了原始 SVM 求解二次规划的问题,提高了训练速度,Suykens<sup>[9]</sup> 将 LSSVM 扩展到多分类问题中。

ELM 是 Huang<sup>[10]</sup> 提出的一种快速算法,它将隐层节点权值随机确定,只需计算输出层权值,从而大大加快了网络的训练速度,并且 ELM 本质上可直接应用于多分类问题,因而取得了广泛应用<sup>[11-13]</sup>。核极限学习机(KELM)是 Huang<sup>[14]</sup> 为解决 ELM 结果具有随机性而提出了基于核函数的改进算法,在遥感图像分类<sup>[15]</sup>、基因检测<sup>[16]</sup> 等方面取得了良好的分类效果。

由文献[17]可知:ELM 和 SVM 均是基于单隐层前馈神经网络框架,而 LSSVM 是 ELM 简化的框架结构;ELM 通过最小平方优化法可扩展为 SVM 网络,且具有更好的泛化性能。ELMs 和 SVMs 均可实现相当的分类型准确率,但在不同的问题条件下也表现出不同的特性<sup>[18-19]</sup>。研究发现:经典的 ELM 算法与 SVM 算法相比具有训练速度快、适应性强等优势;SVM 算法需要花费相对更多的训练时间;LSSVM 的训练算法比 SVM 的训练算法简单,训练精度与 SVM 相当或超过 SVM,但对新数据的应用却要对整个训练集进行处理。

本文通过在 UCI 数据集上的对比验证,详细探讨了 ELMs 和 SVMs 在多分类问题上的分类准确率、对类别的敏感程度和算法运行时间等性能指标,并给出相应分析和结论。实验表明,随着分类数目的增加,ELMs 的泛化能力相较于 SVM 提高得更多。

## 1 ELMs 和 SVMs 简介

### 1.1 支持向量机(SVM)

SVM 是指通过支持向量运算的方式进行分类的分类器,是一种基于统计学习理论的有监督机器学习算法<sup>[20]</sup>,通常用来解决线性分类问题和非线性分类问题。其处理分类问题的典型模式是寻找使各类别分类间隔最大的最优分类超平面,给定一组数据作为对应的类别标签,则由文献[4]可得 SVM 求解如下优化方程:

$$\min_{w, \xi_i} \frac{1}{2} \|w\|^2 + C \cdot \sum_{i=1}^N \xi_i \quad (1)$$

$$\text{s. t. } \xi_i \geq 0, y_i [w^T \Phi(x_i) + b] \geq 1 - \xi_i \quad \forall i$$

式中: $\Phi$  为特征映射函数, $w$  为超平面参数, $\xi_i$  为分类软间隔, $C$  为离群点惩罚因子。

使用 KKT<sup>[21]</sup> 条件,则式(1)通过拉格朗日乘数法转化为原形式的对偶问题来进行求解:

$$L(w, b, \xi_i; \alpha_i, \beta_i) = \frac{1}{2} \|w\|^2 + C \cdot \sum_{i=1}^N \xi_i - \sum_{i=1}^N \alpha_i (y_i [w^T \Phi(x_i) + b] + \xi_i - 1) - \sum_{i=1}^N \beta_i \xi_i \quad (2)$$

式中: $\alpha_i, \beta_i \geq 0$  为拉格朗日乘数子,通过对  $w, b, \xi_i$  分别求偏导可得:

$$\begin{cases} \frac{\partial L(w, b, \xi_i; \alpha_i, \beta_i)}{\partial w} = 0 \rightarrow w = \sum_{i=1}^N \alpha_i y_i \Phi(x_i) \\ \frac{\partial L(w, b, \xi_i; \alpha_i, \beta_i)}{\partial b} = 0 \rightarrow \sum_{i=1}^N \alpha_i y_i = 0 \\ \frac{\partial L(w, b, \xi_i; \alpha_i, \beta_i)}{\partial \xi_i} = 0 \rightarrow 0 \leq \xi_i \leq C \end{cases} \quad (3)$$

则式(2)可转化为:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N y_i y_j \Phi(x_i)^T \Phi(x_j) \\ \text{s. t.} \quad & \sum_{i=1}^N \alpha_i y_i = 0 \quad 0 \leq \alpha_i \leq C \quad \forall i \end{aligned} \quad (4)$$

最终的最优分类超平面的解为:

$$f(x) = \text{sgn} \left( \sum_{i=1}^m \alpha_i y_i \kappa(x_i, x) + b \right) \quad (5)$$

在使用 SVM 对线性分类问题求解时,分类效果对惩罚因子的选择依赖较大,而在使用 SVM 对非线性分类问题求解时,核函数的形式和其参数对高维映射的效果影响较大,因此 SVM 在处理分类问题时还存在一定的局限性。

### 1.2 最小二乘支持向量机(LSSVM)

LSSVM 与 SVM 思路基本一致,是 SVM 的一种演变。LSSVM 是由原来的不等式约束演变成了等式约束,主要用来解决等式约束下的优化问题,其次 LSSVM 与 SVM 的最显著区别在于 LSSVM 在求解时使用最小二乘损失函数:

$$\min_{w, \xi_i} \frac{1}{2} \|w\|^2 + \frac{C}{2} \cdot \sum_{i=1}^N \xi_i \quad (6)$$

$$\text{s. t. } y_i [w^T \Phi(x_i) + b] = 1 - \xi_i \quad \forall i$$

同 SVM,式(6)可转化为拉格朗日函数来进行求解:

$$L(w, b, \xi_i; \alpha_i) = \frac{1}{2} \|w\|^2 + \frac{C}{2} \cdot \sum_{i=1}^N \xi_i - \sum_{i=1}^N \alpha_i (y_i [w^T \Phi(x_i) + b] + \xi_i - 1) \quad (7)$$

由式(7)分别对  $w, b, \xi_i, \alpha_i$  求偏导得:

$$\begin{cases} \frac{\partial L(w, b, \xi_i; \alpha_i)}{\partial w} = 0 \rightarrow w = \sum_{i=1}^N \alpha_i y_i \Phi(x_i) \\ \frac{\partial L(w, b, \xi_i; \alpha_i)}{\partial b} = 0 \rightarrow \sum_{i=1}^N \alpha_i y_i = 0 \\ \frac{\partial L(w, b, \xi_i; \alpha_i)}{\partial \xi_i} = 0 \rightarrow \alpha_i = C \xi_i \\ \frac{\partial L(w, b, \xi_i; \alpha_i)}{\partial \alpha_i} = 0 \rightarrow y_i [w^T \Phi(x_i) + b] + \xi_i - 1 = 0 \end{cases} \quad (8)$$

将式(8)中前三个等式代入第四个等式,可得:

$$\begin{bmatrix} 0 & -Y^T \\ Y & ZZ^T + \frac{I}{C} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{1} \end{bmatrix} \quad (9)$$

式中:  $Y = [y_1, y_2, \dots, y_N]^T, \alpha = [\alpha_1, \alpha_2, \dots, \alpha_N]^T, Z = [y_1 \Phi(x_1), y_1 \Phi(x_2), \dots, y_N \Phi(x_N)]^T$ .

令  $\Omega = ZZ^T$ , 根据 Mercer 条件, 有:

$$\Omega_{i,j} = y_i y_j \kappa(x_i, x_j) \quad (10)$$

将式(10)代入式(9), 求解线性方程, 得到的结果与式(5)相同。

与 SVM 相比, LSSVM 算法求解速度更快, 但其预测精度比 SVM 稍差。

### 1.3 极限学习机(ELM)

ELM 是一种基于单隐含层的人工神经网络模型求解算法<sup>[10]</sup>, 它的优势是仅需设置隐藏层节点数, 使用最小二乘法求解隐含层到输出层的权值即可, 不需要进行循环迭代。与 BP 神经网络等相比, ELM 具有快速学习、高准确度、泛化能力以及尽可能地减少人工干预等特点。ELM 的目标是最小化训练误差的同时最小化输出权值的范数来求解单隐层前馈神经网络<sup>[10]</sup>。

有  $L$  个隐层节点、 $N$  个数据点的 ELM 的隐层节点输出矩阵  $H$  可定义为:

$$H = \begin{bmatrix} h(w_1^T x_1 + b_1) & h(w_2^T x_1 + b_2) & \dots & h(w_L^T x_1 + b_L) \\ h(w_1^T x_2 + b_1) & h(w_2^T x_2 + b_2) & \dots & h(w_L^T x_2 + b_L) \\ \vdots & \vdots & & \vdots \\ h(w_1^T x_N + b_1) & h(w_2^T x_N + b_2) & \dots & h(w_L^T x_N + b_N) \end{bmatrix} \quad (11)$$

式中:  $h(\cdot)$  为隐层单元激活函数,  $W = [w_1, w_2, \dots, w_L]$  为隐层单元随机确定权值,  $B = [b_1, b_2, \dots, b_L]$  为隐层和输出层之间偏置。

则 ELM 可被定义为如下形式:

$$\min_{\beta} \frac{1}{2} \|\beta\|^2 + \frac{C}{2} \cdot \sum_{i=1}^N \|\xi_i\|^2 \quad (12)$$

$$\text{s. t. } h(x_i)\beta = y_i^T - \xi_i^T \Leftrightarrow H\beta = Y^T - \xi^T$$

式中:  $\beta$  为隐层和输出层之间权值,  $Y = [y_1, y_2, \dots, y_N] \in \{-1, 1\}$  为样本标签,  $\xi = [\xi_1, \xi_2, \dots, \xi_N]$  为预测误差矩阵,  $C$  为惩罚因子。

由式(12)可得  $\beta$  为:

$$\beta = \begin{cases} H^+ Y = \left( H^T H + \frac{I_{L \times L}}{C} \right)^{-1} H^T Y & N \geq L \\ H^+ Y = H^T \left( H H^T + \frac{I_{N \times N}}{C} \right)^{-1} Y & N < L \end{cases} \quad (13)$$

式中:  $H^+$  为  $H$  的 Moore-Penrose 广义逆矩阵。

最终 ELM 的解为:

$$y = h(x)\beta = \begin{cases} h(x) \left( H^T H + \frac{I_{L \times L}}{C} \right)^{-1} H^T Y & N \geq L \\ h(x) H^T \left( H H^T + \frac{I_{N \times N}}{C} \right)^{-1} Y & N < L \end{cases} \quad (14)$$

### 1.4 核极限学习机

为进一步提高 ELM 的稳定性和泛化能力<sup>[14]</sup>, 将核函数的思想引入 ELM, 从而构成核极限学习机(KELM)。与 ELM 相比, KELM 用核映射的方式取代随机映射, 有效解决了“维数灾难”和随机设置隐藏层参数带来的稳定性差的问题, 从而降低计算复杂度。KELM 多用于特征学习以及多分类问题当中。KELM 可表示为:

$$y = h(x) H^T \left( H H^T + \frac{I_{N \times N}}{C} \right)^{-1} Y = \begin{bmatrix} \kappa(x_1, x) \\ \kappa(x_2, x) \\ \vdots \\ \kappa(x_N, x) \end{bmatrix} \left( H H^T + \frac{I_{N \times N}}{C} \right)^{-1} Y \quad (15)$$

## 2 实验与结果分析

### 2.1 实验数据

实验数据来自 UCI 机器学习数据集<sup>[22-23]</sup>, 为比较 ELMs 和 SVMs 分类器在多分类问题上的性能表现, 随机选取了不同类目数的数据集。表 1 列出了实验所用数据集的基本信息, 包括数据集名称、类别数目、总样本个数和属性个数。

表 1 实验数据集的基本信息

数据集	类别数目	总样本个数	属性个数
Iris	3	150	4
Satimage	6	2 400	36
Segment	7	2 310	18

续表 1

数据集	类别数目	总样本个数	属性个数
Amazon	10	800	4 096
Vowel	11	528	10
Libras	15	360	90
Letter	26	13 000	16

为了消除各维数据间数量级差别,避免因为输入输出数据数量级差别较大而引起的训练误差,本文对所有的输入数据均进行了如下的归一化处理,将数据归一化到[0,1]区间:

$$x_k = (x_k - x_{\min}) / (x_{\max} - x_{\min}) \quad (16)$$

式中: $x_{\min}$ 为数据序列中的最小值, $x_{\max}$ 为数据序列中的最大值。

## 2.2 实验环境简介

本文所使用 SVM 程序来自 libsvm-3.12, LSSVM 程序为 LS-SVMLab-1.7, ELMs 程序来自极限学习机官方网站,程序运行环境为:MATLAB R2010b, Windows 7 系统。

## 2.3 实验结果分析

### 2.3.1 单一数据集 (Amazon) 结果分析

为了对比 ELMs 和 SVMs 在多分类问题上的泛化性能,首先选取 Amazon 数据集进行实验分析。对 Amazon 数据集采取逐类增加的方式,类别数目从 3 类依次增加到 10 类,分别记录 ELMs 和 SVMs 的分类准确率。数据集采取独立随机划分的方式,每次选取 70% 的数据作为训练集,余下 30% 作为测试集,并进行十次独立划分,分别计算 ELMs 相对于 SVMs 在不同类别数目下的平均分类准确率,实验结果如表 2 所示,ELMs 相较于 SVMs 分类准确率增长情况如表 3 所示。

表 2 ELMs 和 SVMs 在 Amazon 数据集上随着类别数目增多平均分类准确率 %

类数	SVM	LSSVM	ELM	KELM
3 类	98.89	98.61	<b>99.17</b>	<b>99.17</b>
4 类	96.46	97.29	98.12	<b>98.33</b>
5 类	94.50	96.67	96.50	<b>96.83</b>
6 类	93.89	95.69	96.67	<b>96.67</b>
7 类	93.09	<b>96.55</b>	95.83	<b>96.55</b>
8 类	92.40	95.52	95.94	<b>96.04</b>
9 类	89.20	92.28	92.44	<b>93.67</b>
10 类	88.47	91.94	93.06	<b>93.47</b>

表 3 在 Amazon 数据集上随着类别数目增多 ELMs 相较于 SVMs 分类准确率增长情况 %

类数	ELM Vs SVM	KELM Vs SVM	ELM Vs LSSVM	KELM Vs LSSVM
3 类	0.28 ↑	0.28 ↑	0.56 ↑	0.56 ↑
4 类	1.72 ↑	1.94 ↑	0.85 ↑	1.07 ↑
5 类	2.11 ↑	2.46 ↑	-0.17 ↓	0.17 ↑
6 类	2.95 ↑	2.95 ↑	1.01 ↑	1.01 ↑
7 类	2.94 ↑	3.70 ↑	-0.73 ↓	0 →
8 类	3.83 ↑	3.94 ↑	0.43 ↑	0.54 ↑
9 类	3.63 ↑	4.01 ↑	0.16 ↑	1.50 ↑
10 类	5.18 ↑	5.65 ↑	1.20 ↑	1.66 ↑

为清晰展示实验结果,根据表 3 的结果绘制了 ELMs 相对于 SVMs 在各类别泛化能力增长情况图,结果如图 1 所示。

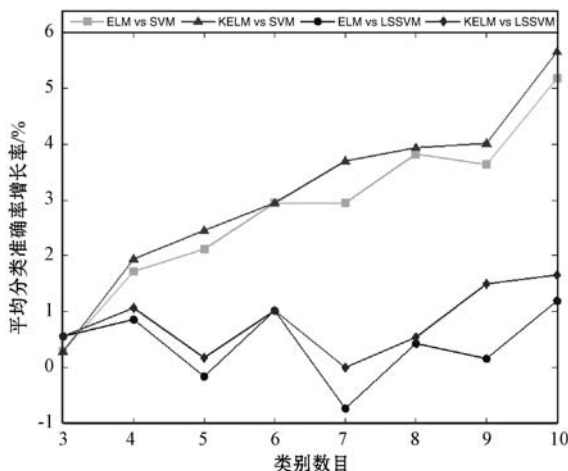


图 1 在同一数据集 Amazon 下不同类别划分 ELMs 相较于 SVMs 泛化性能比较

由表 3 及图 1 可得如下结论:

(1) 对于 Amazon 数据集,ELMs 分类器在各类别上的分类准确率均优于 SVMs 分类器;

(2) ELMs 相较于 SVM 在多分类问题上随着类别数目的增加泛化能力也越来越好,并且 KELM 泛化能力要稍优于 ELM;

(3) ELMs 相对于 LSSVM 并没有表现出(2)中的特性,而是随着类别数目的增加,ELMs 相对于 LSSVM 的泛化能力呈现出明显的波动性。

分别统计 ELMs 和 SVMs 随类别增加分类准确率的下降情况,结果如表 4 所示。

表 4 ELMs 和 SVMs 随类别增加分类准确率的下降情况

类数	SVM	LSSVM	ELM	KELM
3 vs 3	0	0	0	0
4 vs 3	-2.46	-1.34	-1.05	<b>-0.84</b>

续表 4

类数	SVM	LSSVM	ELM	KELM
5 vs 4	-2.03	<b>-0.64</b>	-1.65	-1.53
6 vs 5	-0.64	-1.01	<b>0.17</b>	-0.17
7 vs 6	-0.84	<b>0.89</b>	-0.86	-0.12
8 vs 7	-0.75	-1.06	<b>0.11</b>	-0.52
9 vs 8	-3.46	-3.38	-3.65	<b>-2.47</b>
10vs9	-0.81	-0.36	<b>0.67</b>	-0.21
平均值	-1.38	-0.86	-0.78	<b>-0.73</b>

表 4 可直观表示成图 2 形式。

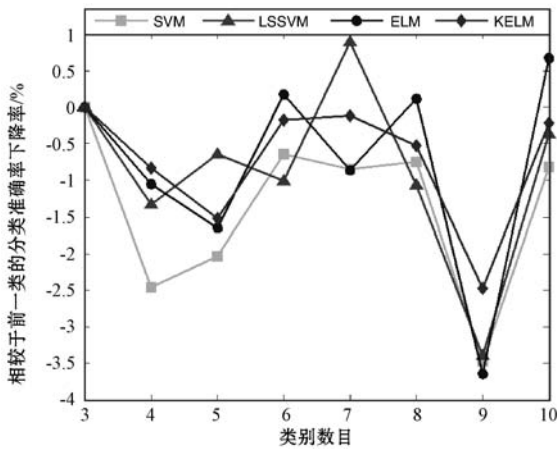


图 2 ELMs 和 SVMs 随类别增加分类准确率的下降情况

由表 4 和图 2 可知:随类别数目的增加,ELMs 分类准确率的下降速率要明显比 SVMs 缓慢,说明 ELMs 对类别数目变化不敏感,更适用于多分类问题。ELMs 相较于 SVMs 在同一数据集上分类稳定性更好,对数据集的宽容度更高。这样的性能表现与其理论密切相关:SVMs 依赖于高维空间映射的准确性,在类别数目较低时更容易对应映射空间找到最优解,而随类别增加高维映射难度增加,分类准确率降低。ELMs 不需要复杂映射,仅需找到对应隐层节点对应权值即可取得最优解,从而减少了其对类别数目增长的所带来的性能损失。

为探究 ELMs 和 SVMs 算法的训练和测试速度,分别统计了各算法在各类别下的平均运行时间,结果如表 5 所示。

表 5 ELMs 和 SVMs 在各类别下平均运行时间 s

类别	SVM	LSSVM	ELM	KELM
#3	2.8	1.78	21.98	<b>0.13</b>
#4	5.59	3.64	23.14	<b>0.16</b>
#5	9.43	6.60	21.50	0.25
#6	13.75	8.95	21.96	<b>0.28</b>
#7	18.83	10.64	22.44	<b>0.38</b>

续表 5

类别	SVM	LSSVM	ELM	KELM
#8	25.51	13.21	22.78	<b>0.47</b>
#9	31.97	17.68	23.36	<b>0.53</b>
#10	40.20	20.46	23.68	<b>0.64</b>

由表 5 可知:

(1) ELMs 和 SVMs 随着类别数目的增加(即样本数目增加),运行时间均有所增加,SVMs 算法运行时间增长幅度明显大于 ELMs。

(2) SVMs 对一个分类问题需要进行复杂运算从而得到最优高维映射的所有解,而 ELMs 仅需计算较少参数即可得到较好的分类准确率,有效降低了运算开销。KELM 在各类别下的运行时间均最短,所需运算负荷最小,适用于快速分类。

### 2.3.2 多数据集结果分析

为进一步探究 ELMs 和 SVMs 在多分类问题上的性能差异,选取 Iris 等 7 个多分类数据集进行实验验证,结果如表 6、表 7 所示。

表 6 ELMs 和 SVMs 在不同数据集上随着类别数目增多平均分类准确率 %

数据集	SVM	LSSVM	ELM	KELM
Iris (#3)	96.27	96.27	96.53	<b>97.07</b>
Sat (#6)	86.39	89.58	89.17	<b>90.28</b>
Seg (#7)	92.87	96.31	96.19	<b>96.79</b>
Ama (#10)	88.47	91.94	93.06	<b>93.47</b>
Vow (#11)	86.62	96.49	96.75	<b>97.53</b>
Lib (#15)	72.33	79.89	<b>83.78</b>	83.67
Lett (#26)	74.28	86.15	86.34	<b>88.01</b>

表 7 在不同数据集下 ELMs 相较于 SVMs 分类准确率增长情况 %

数据集	ELM Vs SVM	KELM Vs SVM	ELM Vs LSSVM	KELM Vs LSSVM
Iris (#3)	0.28 ↑	0.83 ↑	0.28 ↑	0.83 ↑
Sat (#6)	3.22 ↑	4.50 ↑	-0.01 ↓	0.01 ↑
Seg (#7)	3.57 ↑	4.22 ↑	-0.11 ↓	0.59 ↑
Ama (#10)	5.18 ↑	5.65 ↑	1.20 ↑	1.66 ↑
Vow (#11)	11.69 ↑	12.59 ↑	0.27 ↑	1.07 ↑
Lib (#15)	15.82 ↑	15.66 ↑	4.87 ↑	4.72 ↑
Lett (#26)	16.24 ↑	18.48 ↑	0.01 ↑	2.16 ↑

为清晰展示实验结果,根据表 7 的结果分别绘制了 ELMs 和 SVMs 在各类别的分类准确率图和 ELMs 相对于 SVMs 在各类别泛化能力增长情况图,结果如

图 3 所示。

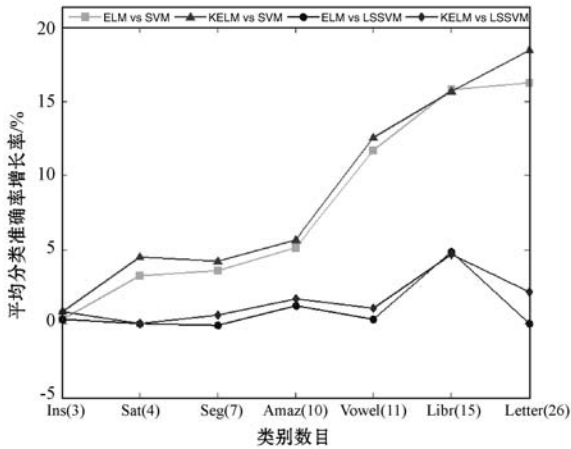


图 3 在不同数据集下 ELMs 相较于 SVMs 泛化性能比较  
由表 7 及图 3 可得如下结论:

(1) 将图 3 和图 1 进行对比,可看出两图呈现出一致的 trend,说明 ELMs 相较于 SVMs 在多分类问题上性能一致性更好;

(2) ELMs 在所有数据集上均取得优于 SVMs 的分类准确率;

(3) 不同数据集下,随着数据集的类别数目的增长,ELMs 相较于 SVM 也同样获得了如单一数据集时更好的泛化能力,但是对于 LSSVM 上述结论并不成立,ELMs 相较于 LSSVM 的泛化能力依然呈现出波动性,相关问题的原因可做进一步研究。

### 3 结 语

本文详细比较了 ELMs 和 SVMs 在多分类问题上泛化性能的差异,并且得出如下结论:(1) ELMs 相较于 SVM 在多分类问题上有更高的分类准确率,而且随着分类数目的增加,ELMs 的泛化能力相较于 SVM 提高更多,但是 ELMs 对于 LSSVM 并没有得到上述结论;(2) ELMs 相较于 SVMs 对数据的类别数目不敏感,分类准确率随类别数目增加下降不明显;(3) ELMs 相较于 SVMs 在多分类问题上所需计算代价更小,且拥有更快的学习和训练速度。

### 参 考 文 献

[1] 杜芳华,冀俊忠,赵学武,等. 基于特征映射的半监督文本分类算法[J]. 北京工业大学学报,2016,42(2):230-235.  
[2] 宋青松,田正鑫,孙文磊,等. 用于孤立数字语音识别的一种组合降维方法[J]. 西安交通大学学报,2016,50(6):42-46.  
[3] 荣晶施,潘海为,高琳琳,等. 基于对称性理论的医学图像多阶段分类算法[J]. 计算机学报,2015,38(9):1809-1823.

[4] Cortes C, Vapnik V. Support-Vector Networks[J]. Machine Learning,1995,20(3):273-297.  
[5] 郭明玮,赵宇宙,项俊平. 基于支持向量机的目标检测算法综述[J]. 控制与决策,2014,29(2):193-200.  
[6] Cavallaro G, Riedel M, Richerzhagen M, et al. On Understanding Big Data Impacts in Remotely Sensed Image Classification Using Support Vector Machine Methods[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2015, 8(10):4634-4646.  
[7] Weston J, Watkins C. Multi-class Support Vector Machines[M]//Support Vector Machines for Pattern Classification. Springer London, 2005:83-128.  
[8] Suykens J A K, Vandewalle J. Least Squares Support Vector Machine Classifiers[J]. Neural Processing Letters, 1999, 9(3):293-300.  
[9] Suykens J A K, Vandewalle J. Multiclass Least Squares Support Vector Machines[C]//Proceedings of the International Joint Conference on Neural Networks (IJCNN'99). IEEE, 1999: 900-903.  
[10] Huang G B, Zhu Q Y, Siew C K. Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks[C]//Proceedings of the 2004 IEEE International Joint Conference on Neural Networks. Budapest, Hungary: IEEE, 2004.985-990.  
[11] Huang G B, Zhu Q Y, Siew C K. Extreme Learning Machine: Theory and Applications[J]. Neurocomputing, 2006, 70(1-3):489-501.  
[12] 徐嘉明,张卫强,杨登舟,等. 基于流形正则化极限学习机的语种识别系统[J]. 自动化学报,2015,41(9):1680-1685.  
[13] Zhao Y, Wang G, Yin Y, et al. Improving ELM-based Microarray Data Classification by Diversified Sequence Features Selection[J]. Neural Computing and Applications, 2016, 27(1):155-166.  
[14] Huang G B, Zhou H M, Ding X J, et al. Extreme Learning Machine for Regression and Multiclass Classification[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 2012, 42(2): 513-529.  
[15] Mahesh Pal, Aaron E. Maxwell & Timothy A. Warner. Kernel-based extreme learning machine for remote-sensing image classification[J]Remote Sensing Letters,2013,4(9): 853-862.  
[16] 杜帮俊,陆慧娟,严珂,等. PSO 和 Cholesky 分解的 KELM 的基因表达数据分类[J]. 中国计量学院学报,2016,27(2):204-209.  
[17] Huang G B, Ding X, Zhou H. Optimization Method Based Extreme Learning Machine for Classification[J]. Neurocomputing, 2010, 74(1-3):155-163.

36-40.

- [3] 李敏,王松艳,张迎春,等. 改进的强跟踪平方根 UKF 在卫星导航中应用[J]. 系统工程与电子技术,2015,37(8):1858-1865.
- [4] 张园,郭晨,李树军,等. 机动目标跟踪的 S 修正无迹卡尔曼滤波算法[J]. 大连海事大学学报,2015,41(2):84-86.
- [5] 黄平,詹洋燕,程广舟. 基于量测一步预测信息的自调整 UKF[J]. 系统工程与电子技术,2016,38(6):1395-1398.
- [6] Sarkka S, Viikari V V, Huusko M, et al. Phase-based UHF RFID tracking with nonlinear Kalman filtering and smoothing [J]. Sensors Journal, 2012, 12(5):904-910.
- [7] de Marina H G, Pereda F J, Giron-Sierra J M, et al. UAV attitude estimation using unscented Kalman filter and TRIAD [J]. IEEE Transactions on Industrial Electronics, 2012, 59(11):4465-4474.
- [8] Kerr T H. Streamlining measurement iteration for EKF target tracking[J]. IEEE Trans. on Aerospace and Electronic Systems, 1991, 27(2):408-421.
- [9] Wan E A, Merwe R V D. The unscented Kalman filter for nonlinear estimation[C]//Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373). IEEE, 2000: 153-158.
- [10] Julier S, Uhlmann J, Durrant-Whyte H F. A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators[J]. IEEE Transactions on Automatic Control, 2000, 45(3):477-482.
- [11] 徐洋,徐松涛,罗文涛,等. 基于虚拟检测函数下的 IMM-UKF 机动目标跟踪[J]. 火力与指挥控制,2015,40(1):37-41.
- [12] Han P, Weilin S, Shi Q Y. Novel nonlinear Kalman filtering method[J]. Chinese Journal of Scientific Instrument, 2015, 36(3):632-638.
- [13] Chang L B, Hu B Q, Li A, et al. Strapdown inertial navigation system alignment based on marginalised unscented Kalman filter [J]. IET Science, Measurement & Technology, 2013, 7(2):128-138.

(上接第 267 页)

- [18] Chorowski J, Wang J, Zurada J M. Review and Performance Comparison of SVM-and ELM-based Classifiers[J]. Neurocomputing, 2014, 128:507-516.
- [19] Zhong H, Miao C, Shen Z, et al. Comparing the Learning Effectiveness of BP, ELM, I-ELM, and SVM for Corporate Credit Ratings[J]. Neurocomputing, 2014, 128:285-295.
- [20] 刘一然,骆力明. 基于 SVM 的学科试题自动分类研究[J]. 计算机应用与软件,2019,36(1):197-203.

- [21] Nocedal J, Wright S J. Numerical Optimization [M]. Springer, 1999.
- [22] <http://archive.ics.uci.edu/ml/>.
- [23] Donahue J, Jia Y, Vinyals O, et al. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition [EB]. eprint arXiv:1310.1531, 2013.

(上接第 273 页)

- [7] Chen X. Decentralized Computation Offloading Game for Mobile Cloud Computing [J]. Parallel & Distributed Systems IEEE Transactions on, 2014, 26(4):974-983.
- [8] Chen X, Jiao L, Li W, et al. Efficient Multi-User Computation Offloading for Mobile-Edge Cloud Computing [J]. IEEE/ACM Transactions on Networking, 2015, 24(5):2795-2808.
- [9] Kosta S, Aucinas A, Hui P, et al. ThinkAir: Dynamic Resource Allocation and Parallel Execution in Cloud for Mobile Code Offloading [C]//2012 Proceedings IEEE INFOCOM. IEEE, 2012:945-953.
- [10] Mahmoodi S E, Subbalakshmi K P, Sagar V. Cloud Offloading for Multi-Radio Enabled Mobile Devices [C]//IEEE International Conference on Communications. IEEE, 2015:5473-5478.
- [11] Chen M H, Dong M, Liang B. Joint Offloading Decision and Resource Allocation for Mobile Cloud with Computing Access Point [C]//IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2016:3516-3520.
- [12] Chen M H, Liang B, Dong M. A Semidefinite Relaxation Approach to Mobile Cloud Offloading with Computing Access Point [C]//2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). IEEE, 2015:186-190.
- [13] Sabella D, Vaillant A, Kuure P, et al. Mobile-Edge Computing Architecture: The Role of MEC in the Internet of Things [J]. IEEE Consumer Electronics Magazine, 2016, 5(4):84-91.
- [14] Jararweh Y, Tawalbeh L, Ababneh F, et al. Scalable Cloudlet-based Mobile Computing Model [J]. Procedia Computer Science, 2014, 34(2):434-441.
- [15] Chen M H, Dong M, Liang B. Joint Offloading Decision and Resource Allocation for Mobile Cloud with Computing Access Point [C]//IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2016:3516-3520.
- [16] Necoara I, Suykens J A K. Interior-Point Lagrangian Decomposition Method for Separable Convex Optimization [J]. Journal of Optimization Theory & Applications, 2009, 143(3):567.