

多任务LS-SVM在时间序列预测中的应用

贾松达¹, 庞宇松^{1,2}, 阎高伟¹

JIA Songda¹, PANG Yusong^{1,2}, YAN Gaowei¹

1. 太原理工大学 信息工程学院, 太原 030024

2. 荷兰代尔夫特理工大学 机械海运与材料工程学院, 荷兰 2628CD

1. College of Information Engineering, Taiyuan University of Technology, Taiyuan 030024, China

2. College of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Delft 2628CD, Holland

JIA Songda, PANG Yusong, YAN Gaowei. Multi-task LS-SVM for application of time series prediction. Computer Engineering and Applications, 2018, 54(3): 233-237.

Abstract: Considering the problems of insufficient information mining and low prediction accuracy in single task time series, a time series prediction method based on Multi Task LS-SVM (MTLS-SVM) is proposed. Multiple time series tasks are simultaneously studied so that task can be pinned down in the training process to induce inductive bias, which improves prediction accuracy. First of all, the several learning tasks are constructed by using the close correlation between adjacent time points. Then the MTLS-SVM model is trained for prediction by the corresponding data sets of each task. This method is applied to several time series data sets. Compared with the single task LS-SVM method, the experimental results show that the proposed method has high prediction accuracy and verify the feasibility and effectiveness.

Key words: time series prediction; multi-task learning; Least Squares Support Vector Machine (LS-SVM); relativity

摘 要: 针对单任务时间序列中存在的信息挖掘不充分、预测精度低等问题, 提出了一种基于多任务最小二乘支持向量机(MTLS-SVM)的时间序列预测方法。该方法将多个时间序列任务同时进行学习, 使得在训练过程中任务之间能够相互牵制起到归纳偏置作用, 最终有效提高模型的预测精度。首先, 利用相邻时间点之间的密切相关性, 构造多个相邻时间点的学习任务, 然后将每个任务对应的数据集同时训练MTLS-SVM模型并将其用于预测。将该方法用于几个时间序列数据集并与单任务LS-SVM方法相比, 实验结果表明该方法具有较高的预测精度, 验证了方法的可行性和有效性。

关键词: 时间序列预测; 多任务学习; 最小二乘支持向量机; 相关性

文献标志码: A **中图分类号:** TP391 **doi:** 10.3778/j.issn.1002-8331.1608-0423

1 引言

时间序列是一种重要的数据对象, 是某个物理量按照时间顺序采样得到的一系列值, 在金融、气象、经济等领域普遍存在^[1]。对时间序列数据进行分析处理, 挖掘其中蕴涵的有效信息, 对于揭示事物发展规律、发现事物之间的相互关系等有重要的实际意义。而时间序列预测正是相应的分析工具, 它可以根据采集到的历史数据建立模型捕捉数据的内在变化规律, 从而用于预测未

来某个时间点或是某段时间内可能产生的结果^[2]。

时间序列预测多采用传统的单任务机器学习方法, 即训练集中的所有样本都是针对单个学习任务的数据, 所有有效信息仅仅蕴含在这些训练数据中。文献[3-4]依据人工神经网络(ANN)灵活的非线性捕捉能力和很强的学习能力, 在输入输出之间建立非线性映射关系, 从原始的序列数据中提取特征并构建模型进行单任务时间序列预测。文献[5]将极限学习机(ELM)应用到时

基金项目: 国家自然科学基金(No.61450011); 山西省自然科学基金(No.2015011052)。

作者简介: 贾松达(1990—), 男, 硕士研究生, 研究领域为智能信息处理; 庞宇松(1970—), 男, 博士, 副教授, 研究领域为智能信息处理; 阎高伟(1970—), 男, 博士, 教授, 研究领域为智能信息处理和多传感器信息融合, E-mail: yangaowei@tyut.edu.cn。

收稿日期: 2016-08-23 **修回日期:** 2017-01-04 **文章编号:** 1002-8331(2018)03-0233-05

CNKI网络优先出版: 2017-03-16, <http://kns.cnki.net/kcms/detail/11.2127.TP.20170316.1432.002.html>

间序列预测中,通过构造 ELM 集合模型解决了时间序列的平稳性与适应性问题。文献[6]则将支持向量机(SVM)用于时间序列预测,它采用风险最小化原则,使得模型具有较好的泛化能力。最小二乘支持向量机(LS-SVM)^[7]是在二次损失函数下 SVM 的一种形式,只对线性方程进行求解并且求解非常迅速,因此,在时间序列预测中应用非常广泛^[8-10]。

然而,这些单任务学习算法在实际应用中经常遇到训练样本不充分的情况,无法对模型进行充分训练,常常会产生欠拟合现象。此外,对任务间的相关性信息没有做进一步的挖掘,从输入样本中所获取的信息有限,最终影响了模型的泛化性能。多任务学习通过利用相关任务的训练样本中所包含的信息进行归纳偏置,对多个具有共享表达的任务进行并行学习,每个任务的学习可以帮助其他任务学习得更好^[11]。因此,多任务学习的研究受到广泛的关注,文献[12]在单任务 LS-SVM 的基础上,提出了一种多输出的最小二乘支持向量回归算法。文献[13]提出了一种核多任务学习算法并应用于电力负荷预测,得到了较好的预测效果。文献[14]将多任务学习与卷积神经网络相结合用于人脸识别,提高了学习效率。

针对以往单任务预测模型中的序列信息挖掘不足、序列间的相关性考虑欠佳等因素导致的预测精度低问题,本文提出了一种基于多任务 LS-SVM(MTLS-SVM)的时间序列预测方法,利用相邻时间点之间的相关性,构造多个相邻时间点的学习任务,并基于多个任务的数据集训练 MTLS-SVM 模型,以提高其泛化能力,并将其用于预测。实验结果表明,与单任务 LS-SVM 方法相比,该方法具有较高的预测精度。

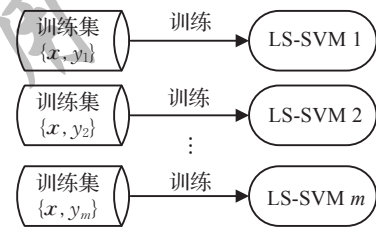
2 基于 MTLS-SVM 的学习过程

多任务学习可以解决多元多维和训练样本不足的问题,是近十几年来机器学习的热点方法。时间序列中相邻时间点蕴含的多种信息可看成是多任务学习中的不同任务,多种任务共享某些共同的数据或模型结构,即任务间存在着一定的相关性,多任务学习方法正是通过共享任务间的相关性来提高机器学习能力,从而缓解小样本过学习问题。单任务 LS-SVM 学习过程与 MTLS-SVM 学习过程的区别如图 1 所示。

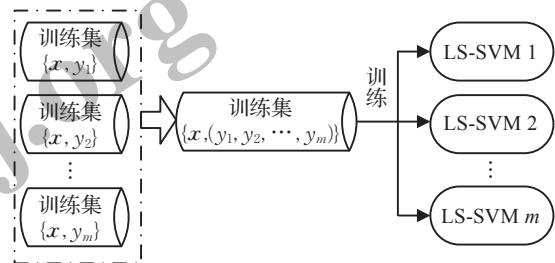
采用单任务 LS-SVM 解决预测问题时,假设训练数据集为 $\{(x_i, y_i)\}_{i=1}^k$,则需要优化式(1)的目标函数,求解权值向量 w 和偏置 b :

$$\begin{aligned} \min J(w, \xi) &= \frac{1}{2} w^T w + \gamma \frac{1}{2} \xi^T \xi \\ \text{s.t. } y &= \varphi^T w + b \mathbf{1}_k + \xi \end{aligned} \quad (1)$$

式中, $\varphi = (\varphi(x_1), \varphi(x_2), \dots, \varphi(x_k))$ 为原始空间向高维特征



(a)单任务LS-SVM学习过程



(b)MTLS-SVM学习过程

图1 单任务LS-SVM与MTLS-SVM的区别

空间的映射, $\xi = (\xi_1, \xi_2, \dots, \xi_n)^T$ 为松弛变量, γ 为正则项。

通过拉格朗日法求解目标函数,最终得到 LS-SVM 的决策函数为:

$$y(x) = \sum_{i=1}^k \alpha_i k(x, x_i) + b \quad (2)$$

式中, α_i 为拉格朗日乘子, $k(x, x_i) = \varphi(x)^T \varphi(x_i)$ 为核函数。若采用 RBF 核函数 $k(x, x_i) = e^{-\frac{\|x - x_i\|^2}{\sigma^2}}$, 则在训练 LS-SVM 的过程中需要优化 α_i 和 σ 。

采用 MTLS-SVM^[15] 解决预测问题时,假设有 m 个学习任务,每个任务有 k_i 个训练样本 $\{x_{i,j}, y_{i,j}\}_{j=1}^{k_i}$, 因此,共有 $k = \sum_{i=1}^m k_i$ 个训练数据。令 $w_i = w_0 + v_i$, 其中, w_0 包含公共信息, v_i 包含特定信息。MTLS-SVM 将通过最小化下面的目标函数,同时找到 $w_0, \{v_i\}_{i=1}^m$ 和 $b = (b_1, b_2, \dots, b_m)^T$:

$$\begin{aligned} \min J(w_0, \{v_i\}_{i=1}^m, \{\xi_i\}_{i=1}^m) &= \\ & \frac{1}{2} w_0^T w_0 + \frac{1}{2} \frac{\lambda}{m} \sum_{i=1}^m v_i^T v_i + \gamma \frac{1}{2} \sum_{i=1}^m \xi_i^T \xi_i \\ \text{s.t. } y_i &= \varphi_i^T (w_0 + v_i) + b_i \mathbf{1}_{n_i} + \xi_i \end{aligned} \quad (3)$$

其中, $\xi_i = (\xi_{i,1}, \xi_{i,2}, \dots, \xi_{i,k_i})^T$ 为松弛变量, $\varphi_i = (y_{i,1} \varphi(x_{i,1}), y_{i,2} \varphi(x_{i,2}), \dots, y_{i,n_i} \varphi(x_{i,k_i}))$ 为原始空间向高维特征空间的映射, λ, γ 是两个正则项参数。

通过拉格朗日法求解上述目标函数,最终得到 MTLS-SVM 的决策函数为:

$$y_i(x) = \sum_{j=1}^m \sum_{l=1}^{k_j} \alpha_{i,j,l} k(x_{i,j,l}, x) + \frac{m}{\lambda} \sum_{j=1}^m \alpha_{i,j} k(x_i, x) + b_i \quad (4)$$

式中, $\alpha_{i,j}$ 为第 i 个任务中第 j 个样本的拉格朗日乘子, $k(x_{i,j}, x)$ 为核函数。

与单任务 LS-SVM 类似,核函数若采用 RBF 核,则

在训练过程中,需要优化的参数有正则化参数 λ , γ 核参数 σ 。相比于单任务LS-SVM,MTLS-SVM中增加了正则化参数 λ ,从式(3)中可以看出,该正则项作用于每个任务,根据 w_0 中包含的公共信息以及 v_i 中包含的特定信息可知,参数 λ 起到了任务之间归纳偏置的作用,从而提高了模型的泛化性能。

3 基于MTLS-SVM的预测模型

本文提出的基于MTLS-SVM的时间序列预测模型包括数据预处理、任务构造和模型训练三个部分,如图2所示。

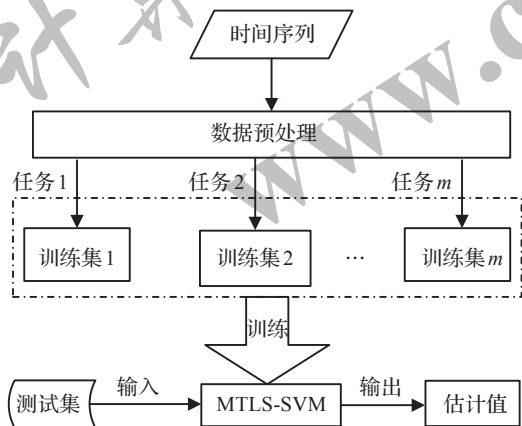


图2 MTLS-SVM时间序列预测模型

3.1 数据预处理

设时间序列 $\{y_1, y_2, \cdots, y_n, y_{n+1}, \cdots, y_t\}$ 为等间隔采样所得,其中 t 表示当前时刻,时间序列预测就是根据历史采样值估计 $t+h$ 时刻的值,即 \hat{y}_{t+h} ,如下式所示:

$$\hat{y}_{t+h} = f(y_t, y_{t-1}, \cdots, y_{t-n+1}) \tag{5}$$

式中, h 为预测步长,若 $h=1$,则为单步预测,若 $h>1$,则为多步预测^[16]。 f 为预测模型, n 为时间延时。将采集到的观测值 $\{y_1, y_2, \cdots, y_n, y_{n+1}, \cdots, y_t\}$ 归一化并构造数据集样本,如表1所示。

表1 构造的数据集样本

输出	输入
y_{n+h}	$y_n, y_{n-1}, \cdots, y_2, y_1$
y_{n+h+1}	$y_{n+1}, y_n, \cdots, y_3, y_2$
\cdots	\cdots
y_{t-1}	$y_{t-h-1}, y_{t-h-2}, \cdots, y_{t-h-n+1}, y_{t-h-n}$
y_t	$y_{t-h}, y_{t-h-1}, \cdots, y_{t-h-n+2}, y_{t-h-n+1}$

3.2 任务构造

将多任务学习方法用于时间序列预测的本质在于相邻时间点的密切相关性,传统的单任务学习方法使用前 n 个时刻的观测值预测 $t+h(h\geq 1)$ 时刻的值,模型输出为单值。而使用多任务学习方法可以将 $t+1, t+2, \cdots,$

$t+m$ 一起作为预测输出,以起到归纳偏置的作用,改进预测精度。时间序列中构造的单任务与多任务如图3所示。

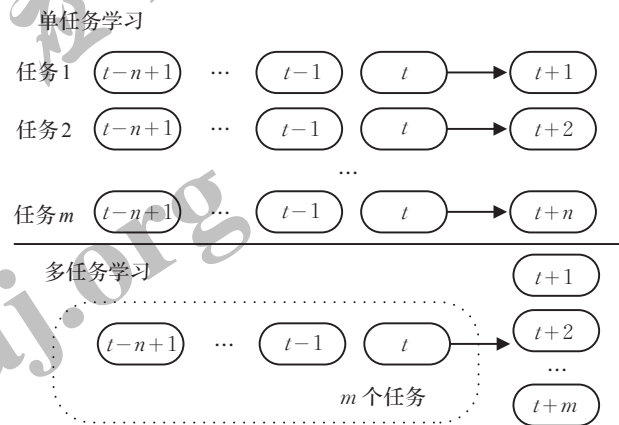


图3 时间序列中的单任务与多任务

时间序列预测策略有多种,采用图3中所示的直接策略进行单任务预测,针对 m 个任务,将会构造出 m 个不同的预测模型,分别输出相应的估计值。而采用多任务学习方法只需构造一个多输出预测模型,得到 m 个任务的输出值。根据时间序列预测目标,选取适当的 m 个任务,并构造每个任务相应的训练集。

3.3 模型训练

将 m 个学习任务相对应的训练集同时训练MTLS-SVM模型。由于这些任务同时对相邻的时刻值进行预测,因此,它们之间是平行关系而且有紧密的相关性。在训练的过程中,任务之间能够相互牵制起到归纳偏置的作用,从而提高模型的泛化性能。基于MTLS-SVM的时间序列预测步骤如下:

输入:时间序列 $y=\{y_1, y_2, \cdots, y_t\}$

输出:未来时刻的估计值 \hat{y}_{t+h}

步骤1 数据预处理,将原始时间序列 y 归一化后,根据预测需要确定不同的 h ,并按照表1列出的规则构造数据集样本。

步骤2 根据目标估计值 \hat{y}_{t+h} 构造 m 个学习任务 $t+1, t+2, \cdots, t+m(m\leq h)$,并针对每个任务对应的输出 \hat{y}_{t+h} ,分别从数据集中选择样本组成 m 个训练集。

步骤3 将 m 个任务的训练集同时训练MTLS-SVM模型,得到最终的预测模型。

步骤4 将 m 个任务相应的测试样本输入训练好的MTLS-SVM模型,可同时得到 m 个估计值,最后从中选择需要的估计值 \hat{y}_{t+h} 。

4 实验结果及分析

本文实验选取的三个数据集来自阿尔托大学理学院的机器学习应用研究组¹,该研究组提供的时间序列数据集均采自实际环境,反应真实的物理现象。这三个

¹ Applications of machine learning group. <http://research.ics.aalto.fi/eiml/datasets.shtml>.

数据集具体的描述为:Sunspot数据集包含了从1700年到1987年每年的太阳黑子数量,数据集规模为288;ESTSP2007数据集取自2007年时间序列预测竞赛欧洲研讨会,数据集规模为875;Darwin数据集为1882年至1998年达尔文海平面的月气压值,数据集规模为1300。具体的数据集划分如表2所示。实验分别从1-step、3-step和5-step不同的步长比较了单任务LS-SVM和MTLS-SVM的预测精度。为了高效训练模型,单任务LS-SVM模型和MTLS-SVM模型均采用交叉验证法^[17]训练,并使用网格搜索法对参数 λ, γ 和 σ 寻优。误差指标采用均方根误差(RMSE)与平均绝对误差(MAE),如式(6)和(7)所示,式中, y_i 表示真实值, \hat{y}_i 表示预测值, N_t 表示预测点数。

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N_t} (y_i - \hat{y}_i)^2}{N_t}}$$

(6)

$$MAE = \frac{1}{N_t} \sum_{i=1}^{N_t} |y_i - \hat{y}_i|$$

(7)

表2 实验数据集划分

时间序列	训练集大小	测试集大小
Sunspot	221	67
ESTSP2007	800	75
Darwin	1 200	100

时间序列预测可以分为单步和多步预测两种情况,在使用MTLS-SVM模型进行单步预测时,采用直接策略可将相邻的3个时刻作为平行预测任务,构建各自的训练集。三个数据集分别在单任务LS-SVM和MTLS-SVM上的1-step的预测结果如图4所示。三个数据集上的1-step预测误差如表3所示,其中粗体表示误差最小值,最后一列表示预测精度的提高程度。从表中可以看出,三个数据集均在MTLS-SVM模型中获得了较小的预测误差。

表3 1-step预测误差对比结果

数据集	误差	LS-SVM	MTLS-SVM	精度提高/%
Sunspot	RMSE	26.170	23.700	9
	MAE	18.820	17.380	8
ESTSP2007	RMSE	0.368	0.242	34
	MAE	0.284	0.197	31
Darwin	RMSE	1.085	0.869	19
	MAE	0.875	0.873	0.2

表4 3-step预测误差对比结果

数据集	误差	LS-SVM	MTLS-SVM	精度提高/%
Sunspot	RMSE	44.760	35.140	21
	MAE	31.060	24.010	23
ESTSP2007	RMSE	0.670	0.609	9
	MAE	0.521	0.449	14
Darwin	RMSE	1.361	1.250	8
	MAE	1.082	0.997	8

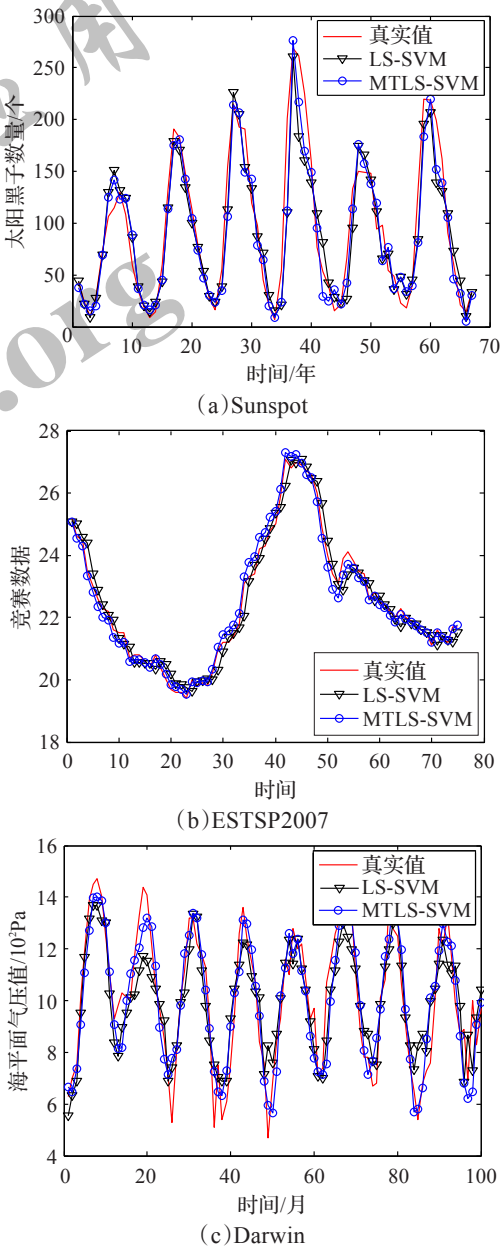


图4 单任务LS-SVM和MTLS-SVM方法1-step预测结果

时间序列的多步预测是一项有挑战性的任务,这是由于随着步长的增大,时间间隔增大,序列之间的关联性降低,对未来时刻取值的决策能力把握不足,使得预测精度降低。为了验证MTLS-SVM在多步预测中的有效性,实验取3-step和5-step两个不同的步长分别作对比,实验结果如表4和表5所示。从表中可以看出,在多步预测中,MTLS-SVM的预测精度仍高于单任务LS-

表5 5-step预测误差对比结果

数据集	误差	LS-SVM	MTLS-SVM	精度提高/%
Sunspot	RMSE	46.460	38.770	17
	MAE	32.310	25.450	21
ESTSP2007	RMSE	0.975	0.652	33
	MAE	0.793	0.409	48
Darwin	RMSE	1.382	1.353	2
	MAE	1.129	1.076	5

SVM, 预测误差均有降低, 这是因为在多任务学习过程中, 多个相关任务同时进行训练, 任务之间相互起到了归纳偏置的作用, 从而提高了每个任务的预测精度。

综上所述, 与单任务 LS-SVM 方法相比, MTL-SVM 在单步预测和多步预测中均取得了较高的预测精度, 验证了该方法的有效性。此外, 在对于类似 Sunspot 数据集类型的小样本集, 缺乏足够的先验知识。因此, 传统的单任务学习方法会遇到泛化能力不强的问题, 而基于多任务的学习方法在不增加输入信息的情况下, 同时关注多个任务的输出, 能够从输入端得到更多的信息, 训练出来的模型也具备了获取这些信息的能力, 而这些信息又能够很好地辅助人们所关注的任务, 进而提高了模型的泛化性能。

从上述的实验结果中可以看出, 传统的单任务时间序列预测方式可能忽略了隐藏在任务之间潜在的、丰富的、有用的训练信息, 而多任务学习可以同时完成对多个训练任务的学习, 它充分地利用了包含在任务间的训练信息, 以改进学习器的泛化性能, 使得 MTL-SVM 模型具有较高的预测精度。

5 结束语

本文提出了一种基于 MTL-SVM 的时间序列预测方法, 利用相邻时间点之间的相关性, 构造多个相邻时间点的学习任务, 并基于多个任务的数据集训练 LS-SVM 模型。该方法利用多个学习任务之间的归纳偏置作用, 提高了预测模型的泛化性能, 通过在三个数据集上的实验结果表明, 与传统的单任务 LS-SVM 模型相比, 该方法获得了较高的预测精度。

将多任务学习与其他主流智能模型结合并用于时间序列预测中, 是下一步的研究方向, 同时, 如何快速有效地优化模型参数, 进一步提高模型的泛化性能同样值得研究。

参考文献:

- [1] Gudmundsson S, Runarsson T P, Sigurdsson S. Support vector machines and dynamic time warping for time series[C]//2008 IEEE International Joint Conference on Neural Networks, 2008: 2772-2776.
- [2] Peng H W, Wu S F, Wei C C, et al. Time series forecasting with a neuro-fuzzy modeling scheme[J]. Applied Soft Computing, 2015, 32: 481-493.
- [3] Yolcu U, Egrioglu E, Aladag C H. A new linear & nonlinear artificial neural network model for time series forecasting[J]. Decision Support Systems, 2013, 54(3): 1340-1347.
- [4] Khandelwal I, Adhikari R, Verma G. Time series forecasting using hybrid ARIMA and ANN models based on DWT decomposition[J]. Procedia Computer Science, 2015, 48: 173-179.
- [5] Van Heeswijk M, Miche Y, Lindh-Knuutila T, et al. Adaptive ensemble models of extreme learning machines for time series prediction[C]//International Conference on Artificial Neural Networks. Berlin Heidelberg: Springer, 2009: 305-314.
- [6] Gui Bin, Wei Xianghe, Shen Qiong, et al. Financial time series forecasting using support vector machine[C]//2014 Tenth International Conference on Computational Intelligence and Security, 2014: 39-43.
- [7] Sapankevych N I, Sankar R. Time series prediction using support vector machines: A survey[J]. IEEE Computational Intelligence Magazine, 2009, 4(2): 24-38.
- [8] Kaytez F, Taplamacioglu M C, Cam E, et al. Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines[J]. International Journal of Electrical Power & Energy Systems, 2015, 67: 431-438.
- [9] Van Gestel T, Suykens J A K, Baestaens D E, et al. Financial time series prediction using least squares support vector machines within the evidence framework[J]. IEEE Transactions on Neural Networks, 2001, 12(4): 809-821.
- [10] Ye Meiyang, Wang Xiaodong. Chaotic time series prediction using least squares support vector machines[J]. Chinese Physics, 2004, 13(4): 454.
- [11] Caruana R. Multitask learning[M]//Learning to learn. US: Springer, 1998: 95-133.
- [12] Xu Shuo, An Xin, Qiao Xiaodong, et al. Multi-output least-squares support vector regression machines[J]. Pattern Recognition Letters, 2013, 34(9): 1078-1084.
- [13] Fiot J B, Dinuzzo F. Electricity demand forecasting by multi-task learning[J]. IEEE Transactions on Smart Grid, 2015.
- [14] 邵蔚元, 郭跃飞. 多任务学习及卷积神经网络在人脸识别中的应用[J]. 计算机工程与应用, 2016, 52(13): 32-37.
- [15] Xu Shuo, An Xin, Qiao Xiaodong, et al. Multi-task least-squares support vector machines[J]. Multimedia Tools and Applications, 2014, 71(2): 699-715.
- [16] Wu S F, Lee S J. Employing local modeling in machine learning based methods for time-series prediction[J]. Expert Systems with Applications, 2015, 42(1): 341-354.
- [17] Arlot S, Celisse A. A survey of cross-validation procedures for model selection[J]. Statistics Surveys, 2010, 4: 40-79.