

基于粗糙集的混合支持向量机长期电力负荷预测研究

李伟, 闫宁, 张振刚

(华北电力大学经济管理系, 河北 保定 071003)

摘要: 影响中长期负荷因素众多, 而且单一核函数支持向量机泛化或学习能力较弱, 预测精度受限。提出一种结合粗糙集和支持向量机智能算法的负荷预测模型, 通过属性约简算法筛选出影响长期电力负荷的核心影响因素, 剔除冗余信息, 选定全社会用电量、人均产值、产值单耗为输入变量, 构建基于多项式核函数、径向基核函数的混合核函数支持向量机预测模型, 有效提高函数的泛化及学习能力。算例结果表明, 所提出的模型预测平均误差仅为 0.59%, 预测精度有了很大提高且适用于长期负荷预测。

关键词: 长期负荷预测; 粗糙集; 属性约简; 支持向量机; 混合核函数

Study on long-term load forecasting of MIX-SVM based on rough set theory

LI Wei, YAN Ning, ZHANG Zhen-gang

(Dept of Business Administration, North China Electric Power University, Baoding 071003, China)

Abstract: Many factors affect long-term load and the generalization and learning capability of single kernel function support vector machine are weak, which makes the prediction accuracy limited. A short-term load forecasting model based on rough sets and support vector machines is proposed to solve this problem. Through the rough set attribute reduction algorithm to find the core of the impact factor of load and remove redundant information, this paper selects the electricity consumption, per capita GDP, consumption per unit output value as input vector, then builds a mixed kernel function support vector machine prediction model based on polynomial kernel function and radial basis kernel function, which can effectively improve the function of generalization and learning ability. The simulation results show that the average error is only 0.59%, and compared with the traditional prediction model, prediction efficiency and accuracy with the new model are greatly improved and more applicable to long-term load forecasting.

This work is supported by National Natural Science Foundation of China(No.70671039).

Key words: long-term load forecasting; rough sets; attribute reduction; support vector machines; mixed kernel function

中图分类号: TM715 文献标识码: A 文章编号: 1674-3415(2010)13-0031-04

0 引言

中长期电力负荷预测对电力系统的生产经营、规划建设等均有重要意义, 其预测结果是进行发电项目、电网改造与建设项目计划的依据。1963年由贝尔实验室的Vapnik等提出的支持向量机(Support Vector Machine, SVM)的机器学习算法能非常成功地处理回归问题(时间序列分析)等诸多问题, 并可推广于预测和综合评价等领域。文献[1-6]分别提出将支持向量机用于电力负荷预测, 证明基于SRM(Structure Risk Minimization, SRM)的预测精度更高, 更适用于电力负荷的预测。SVM方法的不足之处在于当处理的数据量过大时, 冗余信息的

存在使得SVM的训练时间过长、速度变慢。1982年波兰科学家Pawlak提出的粗糙集理论(Rough Sets, RS)能较好地处理不完整知识及冗余信息。文献[7]做了此方面的尝试, 但只是把一般的SVM应用于预测。

基于以上分析, 提出了基于粗糙集的混合核函数支持向量机长期负荷预测新方法。本文使用粗糙集理论中的属性约简算法对与负荷相关的各种历史数据进行约简, 剔除与决策信息不相关的属性, 从而减少了SVM训练数据并大大提高了SVM的预测速度和预测性能。

1 粗糙集属性约简

属性约简是粗糙集理论中的一个重要的研究课

基金项目: 国家自然科学基金项目(70671039)

题，有很多学者做了这方面的工作。一般说来，属性并不是同等重要的，还存在冗余，这不利于做出正确而简洁的决策。属性约简要求在保持知识库的分类和决策能力不变的条件下，删除不相关或不重要的属性。一般而言，较优的属性约简有如下指标：约简后属性个数较少；约简后规则数目较少；最终规范化规则数目较少等。一般说来，属性并不是同等重要的，还存在冗余，这不利于做出正确而简洁的决策。属性约简要求在保持知识库的分类和决策能力不变的条件下，删除不相关的属性。粗糙集属性约简的具体步骤为：

- (1) 由论域中的条件属性和决策属性组成二维视图表即决策表。
- (2) 采用粗糙集离散化算法将决策表中的条件属性和决策属性变量进行离散化。
- (3) 对离散化的新决策表进行属性约简并得到约简的决策规则。

2 支持向量机 (SVM) 预测方法

对于训练样本集 $\{x_i, y_i\}$, $x_i \in R^n$ 为输入变量的值, $y_i \in R$ 为相应的输出值, 为训练样本个数, 支持向量机回归的基本思想是寻找一个从输入空间到输出空间的一个非线性映射 φ , 通过这个非线性映射 φ , 将数据 x 映射到多维特征空间 F , 并在特征空间中用下述线性函数进行线性回归。即

$$f(x) = [w \bullet \varphi(x)] + b \quad \varphi: R^n \rightarrow F, w \in F$$

其中: w 为权值矢量; b 为偏差是常值。影响 w 的有经验风险的总和 $Remp$ 则有

$$R(w) = \frac{1}{2} \|w\|^2 + Remp = \frac{1}{2} \|w\|^2 + \sum_{i=1}^L e(f(x_i) - y_i)$$

其中, L 表示样本的数目, 损失函数通常采用 ε 不敏感区函数。

根据统计学习理论的结构风险最小化准则, 支持向量机回归方法通过极小化目标函数来确定回归函数式:

$$\min \frac{1}{2} \|w\|^2 + \sum_{i=1}^L e(f(x_i) - y_i)$$

由于特征空间的维数很高 (甚至无穷) 且目标函数不可微, 直接求解式几乎是不可行的。支持向量机回归方法的特殊效果在于, 通过引入点积核函数 $k(x_i, y_i)$ 和利用 Wolfe 对偶技巧避开了这些问题, 将上述问题转化为下述有效求解式的对偶问题, 可以表示为:

$$\begin{aligned} & \min \frac{1}{2} \|w\|^2 + \sum_{i=1}^L e(f(x_i) - y_i) \\ & \text{s.t.} \begin{cases} y_i - (w \bullet \varphi(x_i)) - b \leq e + \xi_i \\ (w \bullet \varphi(x_i)) + b - y_i \leq e + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned}$$

建立 Langrange 方程:

$$\begin{aligned} L(w, \xi_i, \xi_i^*) &= \frac{1}{2} \|w\|^2 + C \sum_{i=1}^L (\xi_i - \xi_i^*) - \\ & \sum_{i=1}^L a_i (\varepsilon + \xi_i - y_i + (w \bullet \varphi(x_i)) + b) - \\ & \sum_{i=1}^L a_i^* (e + \xi_i^* + y_i - (w \bullet \varphi(x_i)) - b) - \\ & \sum_{i=1}^L (\lambda_i \xi_i + \lambda_i^* \xi_i^*) \end{aligned}$$

要使上式取得最小值, 对于参数 w, b, ξ, ξ^* 的偏导都应等于零。可以得到:

$$\begin{aligned} & \min \frac{1}{2} \sum_{i,j=1}^L (a_i - a_i^*)(a_j - a_j^*)(\varphi(x_i) \bullet \varphi(x_j)) + \\ & \sum_{i=1}^L a_i (e - y_i) + \sum_{j=1}^L a_j^* (e + y_j) \\ & \text{s.t.} \begin{cases} \sum_{i=1}^L (a_i - a_i^*) = 0 \\ a_i, a_i^* \in [0, C] \end{cases} \end{aligned}$$

由此, 支持向量机的函数回归问题就可以归结为二次规划问题。求解该二次规划问题, 可以得到用数据点表示的 w 。

$$w = \sum_{i=1}^L (a_i - a_i^*) \varphi(x_i) + b$$

$$\begin{aligned} f(x) &= \sum_{i=1}^L (a_i - a_i^*) (\varphi(x_i) \bullet \varphi(x)) + b = \\ & \sum_{i=1}^L (a_i - a_i^*) k(x_i, x) + b \end{aligned}$$

3 基于粗糙集理论的MIX-SVM的预测方法

首先应用粗糙集理论的等价关系对影响负荷的可能的因素进行属性约简分析, 去除冗余属性得到影响负荷的决定性因素, 然后对这些因素构造支持向量机模型进行负荷预测。由于局部性核函数学习能力较强, 泛化能力较弱, 而局部性核函数泛化能力较强, 学习能力较弱, 因此, 我们考虑把两类核

函数混合起来即采用多项式核函数与RBF核函数的混合核函数。

多项式核函数 (POLY) :

$$K_{\text{poly}}(x, x_i) = [(x, x_i) + 1]^q$$

径向基核函数 (RBF) :

$$K_{\text{rbf}}(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right)$$

构造的混合核函数 (MIX) 为:

$$K_{\text{mix}} = \lambda K_{\text{poly}} + (1 - \lambda) K_{\text{rbf}}$$

其中 q 、 σ 、 λ 为常数。

4 算例分析

采用某地区 1998~2005 年的经济及负荷数据为样本数据, 对该地区 2006~2008 年的负荷进行预测。

$R_1 \sim R_6$ 是条件属性 S , 分别表示地区人口、国内生产总值 (GDP)、全社会用电量、人均产值、人均耗电量及产值单耗, 最大负荷为决策属性 D 。原始数据 (表1) 采用等距离散化方法进行离散化, 对以上数据进行处理得到约简结果如表2所示。

表 1 某地区 1998-2008 年的经济及负荷数据

Tab.1 Economy and load of one area from 1998 to 2008

年份	地区人口/万人	GDP/亿元	全社会用电量/ (万 kWh)	人均产值/ (元·人 ⁻¹)	人均耗电量/ (kWh·人 ⁻¹)	产值单耗/ (kWh·元 ⁻¹)	最大负荷 /万 kW
1998	426.0	269.6	861 806	6 328.64	2 023.02	0.32	1 087 000
1999	427.6	285.8	879 225	6 683.82	2 056.19	0.31	1 121 000
2000	427.8	330.8	834 181	7 736.20	1 950.84	0.25	1 218 000
2001	430.0	352.1	866 735	8 188.37	2 015.66	0.21	1 030 000
2002	431.7	401.3	826 162	9 295.81	1 913.74	0.19	1 047 000
2003	431.4	446.4	842 086	10 347.00	1 951.98	0.16	1 018 000
2004	431.0	515.1	840 141	11 951.20	1 949.28	0.16	1 157 000
2005	431.4	600.1	979 256	13 904.00	2 268.90	0.16	1 283 000
2006	432.0	612.3	982 353	14 562.20	2 342.20	0.15	1 321 000
2007	432.2	680.2	1 023 115	16 102.00	2 564.10	0.15	1 524 000
2008	432.3	702.3	1 200 354	18 312.10	2 701.32	0.14	1 602 000

表 2 约简结果

Tab.2 Reduced unitized results

区间 编号	1	2	3	4
约简 集合	R_3, R_4, R_6	R_1, R_5, R_6	R_1, R_3, R_6	R_4, R_5, R_6

从实际经济情况分析, 全社会用电量是负荷按时间积分的结果, 人均产值是国民生产总值与人口的比值, 反映了这二者的共同作用关系, 产值单耗是全社会用电量与 GDP 的比值, 是配电网负荷预测所需的最重要的数据, 与最终的预测值关系最密切。区间标号为 1 的约简属性集合最能反映年最大负荷的变化, 是期望得到的支持向量机的输入向量, 分别为全社会用电量、人均产值、产值单耗。

分别使用 GM(1, 1)、多元回归、基于粗糙集核函数的 SVM ($\lambda=0.6$, $C=30$, $q=3$, $\sigma=0.2$) 进行负荷预测, 结果如表3。

表 3 基于不同核函数模型的负荷预测结果比较

Tab.3 Comparison of forecasting results of power load

年份	10 ⁴ kWh			
	实际值	GM (1, 1)	多元回归	混合核函数 SVM
2006	1 321 000	1 493 100	1 348 348	1 319 303
2007	1 524 000	1 664 900	1 464 070	1 470 741
2008	1 602 000	1 856 400	1 599 129	1 631 630

本文取平均百分比误差 MPE、均方误差 MSE 作为各种预测方法的效果的判断, 如表4所示。

表4 预测结果误差比较

Tab.4 Comparison of prediction error

误差对比	GM (1, 1)	多元回归	MIX-SVM
MPE	-12.7	0.68	0.59
MSE	3 806 352	144 927	123 910

从表4和图1可以看出本文构建的基于粗糙集的

混合核函数支持向量机模型预测误差要比灰色及多元回归模型小得多, 平均百分比误差为0.59%, 均方误差为123 910, 误差曲线也更小, 说明预测效果明显优于传统的灰色和多元回归模型。

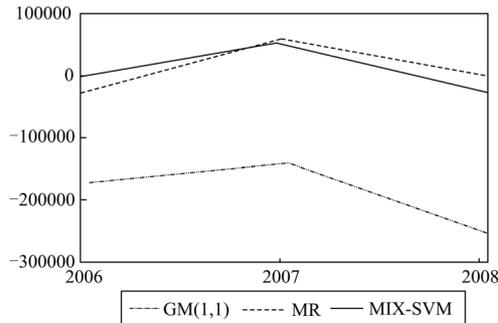


图1 绝对误差对比分析图

Fig.1 Comparison of the forecasting absolute error

5 结论

(1) 本文将粗糙集与支持向量机两种智能算法结合起来用于长期负荷预测, 避免了常规支持向量机算法按经验选取输入参数的盲目性, 因此有效地提高了预测速度和精度。进一步说明了在中长期负荷预测中经济因素是影响负荷的决定因素, 如何提取其中的关键因素决定了预测结果的好坏。

(2) 在 SVM 具体应用过程中, 其核函数的选择对预测效果都有很大的影响, 本文引入混合核函数很好地提高了预测精度。

参考文献

- [1] 刘宝英, 杨仁刚. 基于主成分分析的最小二乘支持向量机短期负荷预测模型[J]. 电力自动化设备, 2008, 28(11): 13-17.
LIU Bao-ying, YANG Ren-gang. Short-term load forecasting model based on LS-SVM with PCA[J]. Electric Power Automation Equipment, 2008, 28(11): 13-17.
- [2] 肖先勇, 葛嘉, 何德胜. 基于支持向量机的中长期电力负荷组合预测[J]. 电力系统及其自动化学报, 2008, 20(1): 84-88.
XIAO Xian-yong, GE Jia, HE De-sheng. Combination method of mid-long term load forecasting based on support vector machine[J]. Proceedings of the CSU-EPSA, 2008, 20(1): 84-88.
- [3] 张林, 刘先珊, 阴和俊. 基于时间序列的支持向量机

在负荷预测中的应用[J]. 电网技术, 2004, 28(19): 38-41.

ZHANG Lin, LIU Xian-shan, YIN He-jun. Application of support vector machines based on time sequence in power system load forecasting[J]. Power System Technology, 2004, 28(19): 38-41.

- [4] 赵登福, 王蒙, 张讲社, 等. 基于支持向量机方法的短期负荷预测[J]. 中国电机工程学报, 2002, 22(4): 26-30.
ZHAO Deng-fu, WANG Meng, ZHANG Jiang-she, et al. Short-term load forecasting based on SVM[J]. Proceedings of the CSEE, 2002, 22(4): 26-30.
- [5] 靳忠伟, 陈康民, 闫伟, 等. 基于支持向量机的中长期电力负荷预测研究与应用[J]. 上海理工大学学报, 2008, 30(2): 129-132.
JIN Zhong-wei, CHEN Kang-min, YAN Wei, et al. Study and application of support vector machine to forecast mid-long-term electric power load[J]. J University of Shanghai for Science and Technology, 2008, 30(2): 129-132.
- [6] 李元诚, 方廷健, 于尔铿. 短期负荷预测的支持向量机方法研究[J]. 中国电机工程学报, 2003, 23(6): 55-59.
LI Yuan-cheng, FANG Ting-jian, YU Er-keng. Study of support vector machines for short-term load forecasting[J]. Proceedings of the CSEE, 2003, 23(6): 55-59.
- [7] 张庆宝, 程浩忠, 刘青山, 等. 基于粗糙集属性约简算法和支持向量机的短期负荷预测[J]. 电网技术, 2006, 30(8): 56-59.
ZHANG Qing-bao, CHENG Hao-zhong, LIU Qing-shan, et al. Short-term load forecasting based on rough set attribute reduction algorithm and support vector machine[J]. Power System Technology, 2006, 30(8): 56-59.

收稿日期: 2009-08-27; 修回日期: 2009-10-27

作者简介:

李伟(1968-), 男, 教授, 硕士研究生导师, 研究方向为电力市场及经济分析;

闫宁(1984-), 男, 硕士研究生, 研究方向为电力市场技术经济管理; E-mail: yanning_003@163.com

张振刚(1977-), 男, 硕士研究生, 研究方向为电力市场技术经济管理。