

文章编号: 1000-5013(2007)03-0252-04

集成 RS 和 SVR 的电力系统短期负荷预测方法

方瑞明

(华侨大学 信息科学与工程学院, 福建 泉州 362021)

摘要: 基于粗糙集(RS)理论和支持向量回归(SVR)方法, 提出一种电力系统短期负荷预测方法. 采用粗糙集理论对影响负荷预测的各因素进行约简, 将约简后得到的最小条件属性集, 以此确定输入样本的维数并构造训练样本, 作为支持向量回归机的输入进行训练预测. 在此基础上, 利用已知历史负荷数据构造训练样本群, 作为 SVR 的输入进行训练, 采用训练完毕后的 SVR 模型进行负荷预测. 实验结果表明, 与神经网络方法和标准 SVR 方法相比, 集成粗糙集和支持向量回归的负荷预测方法, 可以在缩短训练时间的前提下获得较高的预测精度.

关键词: 电力系统; 训练样本; 短期负荷预测; 粗糙集; 支持向量回归

中图分类号: TM 714; TM 743

文献标识码: A

短期负荷预测是电力系统负荷预测的重要组成部分, 是电力系统调度管理部门制订开停机计划及在线安全分析的基础, 对电力系统的可靠和经济运行具有重要意义^[1]. 现有的负荷预测方法, 大体分为基于统计模型的方法^[2-5]和基于人工智能的方法. 文[6]指出, 基于神经网络的混合算法在短期负荷预测中具有较强的优越性. 然而, 神经网络方法本质上是一种基于经验风险最小化的方法, 而不是使期望风险最小化, 它的泛化能力受到制约. 支持向量机(Support Vector Machine, SVM)是不同于神经网络的一种机器学习方法^[7], 主要应用在模式识别和函数逼近两个方面, 当其在处理函数逼近和回归估计问题的时候, 通常称其为支持向量回归方法(Support Vector Regression, SVR). 与神经网络方法相比, SVR 方法在处理短期负荷预测上有着显著的优越性^[8-9]. 粗糙集(Rough Set, RS)是一种新型的数据处理分析方法, 其最大的特点是对数据进行信息约简, 在不影响信息完备性的前提下去掉数据中的冗余信息, 简化输入信息的数据空间的维数^[10]. 因此, 本文在粗糙集及支持向量机回归的基础上, 提出了一种融合两种方法特点的短期负荷预测方法.

1 RS 约简与 SVR 回归

1.1 粗糙集理论与区分矩阵的知识约简

粗糙集理论是一种处理模糊性和不精确性问题的新型数学工具, 能够在保留关键信息的前提下对知识进行处理, 并求得知识的最小表达. 它把对象世界抽象为一个信息系统, $S = (U, A, V, f)$. 其中, U 为对象的有限集合, 称为论域; A 为属性集合, 又可以划分为条件属性集 C 与决策属性集 D 两个不相交的集合, $A = C \cup D$; V 为属性 A 的值域; f 是一个指定 U 中每个对象属性值的信息函数, $f: U \times A \rightarrow V$.

由这种“属性-值”关系构成了一张二维表, 称为信息表(也称决策表). 决策表中的一个属性对应一个等价关系, 在决策表中并非所有的条件属性都是必要的, 有些是多余的, 去除这些多余的属性不影响原有的表达效果. 采用粗糙集描述知识, 通过去除冗余的条件属性求取约简, 可以实现知识的简化, 将一个复杂的决策表约简为不含多余属性, 并保证分类正确的最小条件属性决策表. 本文采用基于区分矩阵的知识约简方法, 其基本步骤有 9 个^[11].

收稿日期: 2006-12-18

作者简介: 方瑞明(1972-), 男, 副教授, 主要从事电机与人工智能的研究. E-mail: fangrm@hotmail.com.

基金项目: 国家自然科学基金资助项目(50477010)

- (1) 根据各个实例特征值, 求出决策区别矩阵 M , M 的元素表示为 M_{ij} , 称为项, 它是 S 中第 i 个实例和第 j 个实例有差别的所有属性的集合.
 - (2) 初始化 $A = (a^1, a^2, \dots, a^n)$, 其中 a^1, a^2, \dots, a^n 为各个条件属性集合.
 - (3) 由 M 求出条件属性关于决策属性的核 C_0 .
 - (4) $R \leftarrow C_0$.
 - (5) $Q = \{M_{ij} : M_{ij} \text{ IR } \neq \emptyset, i, j = 1, 2, \dots, n\}$, $M = M - Q$, $B = B - R$.
 - (6) While ($M \neq \emptyset$) do begin.
 - (7) 对所有 $a_k \in B$, 计算在 M 中的 $p(a_k)$, 并且令 $p(a_q) = \max\{p(a_k)\}$, $p(a_k)$ 是 M 属性 a_k 的属性频率函数, 它定义为属性 a_k 在 M 中出现的次数.
 - (8) $R \leftarrow RY\{a_q\}$.
 - (9) $Q = \{M_{ij} : M_{ij} \text{ IR } \neq \emptyset, i, j = 1, 2, \dots, n\}$, $M = M - Q$, $B = B - R$ end {while}.
- 通过上述 9 步骤的计算, 求出的 R 就是条件属性关于决策属性的最小简化.

1.2 SVR 的函数回归

假设 $\{x_i, y_i\}_{i=1}^n$ 为一训练数据集, 其中 $x_i \in R^k$ 表示样本的输入空间, $y_i \in R$ 表示相应的目标值. 回归问题主要是确定一个可以较精确地逼近未来目标值的一个函数. 通常 SVR 估计函数可以表示为

$$f(x) = w \cdot \Phi(x) + b. \quad (1)$$

式中, $\Phi(x)$ 是输入空间到高维空间的非线性映射. SVR 就是将实际问题通过非线性映射转换到高维特征空间, 在高维特征空间中构造线性回归函数来实现原空间中的非线性回归函数. 系数 w 和 b 可由最小化回归风险来估计, 回归风险表示为

$$R_{\text{SVR}}(C) = C \frac{1}{n} \sum_{i=1}^n L_{\varepsilon}(y_i, w\Phi(x_i) + b) + \frac{1}{2} \|w\|^2. \quad (2)$$

式(2)中, C 是一常数, $L_{\varepsilon}(\cdot)$ 为损失函数, $L_{\varepsilon}(d, y) = \begin{cases} 0, & |d - y| \leq \varepsilon \\ |d - y| - \varepsilon, & |d - y| > \varepsilon \end{cases}$. 为了寻找系数 w 和 b , 引入松弛变量 ξ 和 ξ_i , 则式(2)最小式为

$$\min R_{\text{SVR}}(w, \xi^*) = C \sum_{i=1}^n (\xi_i + \xi^*) + \frac{1}{2} \|w\|^2. \quad (3)$$

$$\left. \begin{aligned} w\Phi(x_i) + b_i - y_i &\leq \varepsilon + \xi_i, & \xi_i &\geq 0, \\ y_i - w\Phi(x_i) - b_i &\leq \varepsilon + \xi_i, & \xi_i &\geq 0. \end{aligned} \right\} \quad (4)$$

依靠引入拉格朗日乘子, 可以得到用数据点表示的向量为

$$w = \sum_{i=1}^n (a_i - a_i^*) \Phi(x_i). \quad (5)$$

把式(5)代入式(1), 则通常的估计函数可以表示为

$$f(x) = \sum_{i=1}^n (a_i - a_i^*) (\Phi(x_i) \cdot \Phi(x)) + b = \sum_{i=1}^n (a_i - a_i^*) (k(x_i, x)) + b. \quad (6)$$

在式(5)中, 点积被核函数 $k(x_i, x)$ 代替, 核函数能够在无需知道映射 Φ 的情况下, 利用低维空间数据输入在高维特征空间进行点积计算, 任何函数只要满足 Mercer 条件都可以作为核函数. 本文主要采用径向基核函数(RBF)进行回归计算, 即

$$k(x_i, x) = \exp\left\{-\frac{\|x - x_i\|^2}{\sigma^2}\right\}. \quad (7)$$

2 RS 约简的 SVR 负荷预测方法

RS 约简的 SVR 负荷预测模型, 主要是利用 RS, 对可能影响负荷预测(即决策属性)的众多因素(即条件属性)进行知识约简. 在不影响信息完备性的前提下, 求得最小条件属性集, 以此确定输入样本的维数, 在此基础上, 利用已知历史负荷数据构造训练样本群, 作为 SVR 的输入进行训练, 采用训练完毕后的 SVR 模型进行负荷预测. 日最大负荷预测是短期负荷预测的重要组成部分之一, 它对于发电计划制定、输电安全分析, 以及安排设备检修等都具有重要的指导意义. 以日最大负荷预测为例, 其 SVR 负荷

预测方法有如下 5 个步骤流程.

- (1) 建立初始决策表. 将各个可能的影响负荷预测因素作为条件属性, 将待预测日最大负荷视为决策属性, 构造初始决策表. 本文条件属性选取预测日之前 13 d 的日最大负荷、日最高温度、日平均温度、日最低温度、日平均湿度, 以及预测日的日最高温度、日最低温度、日最低温度、日平均湿度及预测日类型共 60 个相关因素作为条件属性.
- (2) 连续变量离散化. 考虑到负荷和气象因素均为连续变量, 通过最小信息熵离散化算法对属性值进行离散化, 从而形成新的离散化决策信息表^[12].
- (3) 条件属性约简. 按照本文所述方法, 对决策信息表进行约简, 获得最小条件属性集.
- (4) 根据历史数据, 以最小条件属性集中的变量为依据, 构造输入样本群.
- (5) SVR 进行训练. 在 SVR 模型训练时, 需要合理确定惩罚因子 C 和 RBF 函数核宽度控制常数 σ . 本文通过交叉验证法确定 C, σ 分别为 76, 0.0485.
- 基于上述步骤最终确定了 SVR 模型后, 即可以进行日最大负荷预测.

3 数值实验结果与分析

数值实验的负荷数据来源于某省电力公司的历史数据. 在短期负荷预测的分析和建模中, 以日最大负荷预测为目标进行讨论. 图 1 为 2004 年度该省 1 月份和 7 月份每日最大负荷(F_{\max})的实际数据和预测结果. 由图 1 可见, 预测结果与每日实际最大负荷变化曲线吻合较好. 作为对比, 基于相同数据, 采用文[13]中的神经网络法进行了预测. 与此同时, 还采用了一个不带粗糙集预处理的支持向量回归法进行了预测, 预测结果比较如表 1 所示.

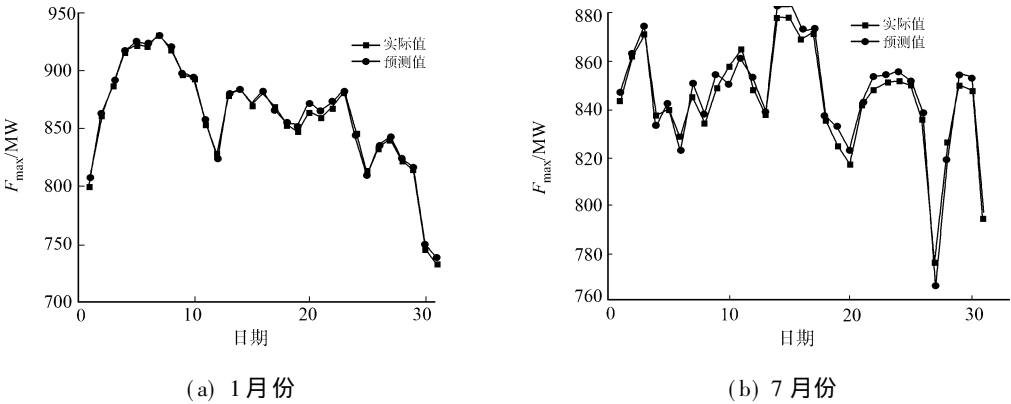


图 1 每日最大负荷预测值和实际值

Fig. 1 The actual value and predicted value of everyday's peek load

表 1 从预测精度和训练时间两方面对 3 种方法的预测性能进行了比较, 其中预测精度采用平均相对误差(η)作为分析指标. 其定义如下

$$\eta = \frac{\left[\sum_{i=1}^n \frac{(|P_{Ai} - P_{Fi}| \cdot 100)}{P_{Ai}} \right]}{n}$$

(8)

式中, P_{Ai} 为实际负荷值, P_{Fi} 为预测负荷值, n 预测次数. 由表 1 中可以看出, 支持向量回归预测方法的预测精度和训练时间均要优于神经网络方法. 与标准 SVR 预测法相比, 集成粗糙集和支持向量回归的负荷预测方法不仅可以进一步提高预测精度, 而且能够显著缩短训练时间, 说明本文所提出的方法在短期负荷在线预测方面具有良好的应用前景. 上述数值仿真实验均在中央处理器(CPU)主频 2.5 GHz, 1GB 内存的同一台式计算机上完成.

表 1 预测性能比较

Tab. 1 Comparison of forecasting performances

预测方法	1 月份		7 月份	
	$\eta / (\%)$	t / s	$\eta / (\%)$	t / s
本文方法	0.35	7.6	0.38	7.6
神经网络法 ^[13]	1.86	172.5	1.67	164.6
标准 SVR 法	1.61	72.3	1.53	74.5

4 结束语

本文提出了一种基于粗糙集和支持向量回归的电力系统短期负荷预测方法. 数值实验结果表明, 该方法能够在综合考虑与负荷预测相关的各种因素的前提下, 在较短的训练时间内达到较高的预测精度, 是一种有效的负荷预测方法.

参考文献:

- [1] 康重庆, 白利超, 夏清, 等. 电力市场中发电商的风险决策[J]. 中国电机工程学报, 2004, 24(8): 1-6.
- [2] HAIDA T, MUTO S. Regression based peak load forecasting using a transformation technique[J]. IEEE Transactions on Power Systems, 1994, 9(4): 1788-1794.
- [3] PAPALEXOPOULOS A D, HESTERBERG T C. A regression based approach to short term load forecasting[J]. IEEE Transactions on Power Systems, 1990, 5(4): 1535-1550.
- [4] 赵宏伟, 任震, 黄雯莹. 基于周期自回归模型的短期负荷预测[J]. 中国电机工程学报, 1997, 17(5): 348-351.
- [5] INFELD D G, HILL D C. Optimal smoothing for trend removal in short term electricity demand forecasting[J]. IEEE Transactions on Power Systems 1998, 13(3): 1115-1120.
- [6] SPYROST, ELPIDA T. Computational intelligence techniques for short term electric load forecasting[J]. Journal of Intelligent and Robotic Systems, 2001, 31: 7-68.
- [7] VAPNIK V N. The nature of statistical learning theory[M]. 2nd ed. New York: Springer, 2000.
- [8] MOHAMED M. Support vector machines for short term electrical load forecasting[J]. International Journal of Energy Research, 2002, 26: 335-345.
- [9] 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(5): 55-59.
- [10] 王国胤. Rough 集理论与知识获取[M]. 西安: 西安交通大学出版社, 2001.
- [11] 王珏, 王任, 苗夺谦, 等. 基于 Rough Set 理论的数据浓缩[J]. 计算机学报, 1998, 21(5): 393-400.
- [12] 赵荣泳, 张浩, 李翠玲. 粗糙集连续属性离散化模型研究与应用要点分析[J]. 计算机工程与应用, 2005, 41(8): 40-42.
- [13] 谢宏, 程浩忠, 张国立, 等. 基于粗糙集理论建立短期电力负荷神经网络预测模型[J]. 中国电机工程学报, 2003, 23(11): 1-4.

Short-Term Load Forecasting of Power System with Rough Set and Support Vector Regression

FANG Ruiming

(College of Information Science and Engineering, Huaqiao University, Quanzhou 362021, China)

Abstract: A novel method to short term load forecasting of power system based on rough set theory (RST) and support vector regression (SVR) is presented. Firstly, the factors that affect the load forecasting are reduced using RST method. Then a SVR module is trained with the historical load data whose dimension is constructed according to the minimum attributes set acquired by RST. Finally the trained SVR module is used to forecast the future short term load. The experimental results show that, when compared against both neural network method and standard SVR method, the proposed method can forecast more accurate results while shortening the training time.

Keywords: power system; trained sample; short term load forecasting; rough set; support vector regression

(责任编辑: 黄仲一)