

最小二乘支持向量机短期负荷预测研究

侯贺飞, 刘俊勇

(四川大学电气信息学院, 四川 成都 610065)

摘要:电力系统短期负荷预测是一项非常重要的工作, 准确的短期负荷预测对于电力系统经济、安全、可靠的运行具有特别重要的意义。随着电力系统的日趋复杂化, 特别是电力市场的逐步深入, 短期负荷预测被赋予了更高的要求。提出了基于负荷日周期性进行前后向外推的数据预处理新方法, 为短期负荷预测模型利用这些历史数据奠定了基础。最小二乘支持向量机是新一代机器学习方法, 将其应用于电力系统短期负荷预测, 在充分利用日周期性和同时刻负荷相近性的基础上, 提出了基于最小二乘支持向量机回归算法 (LSSVR) 的短期负荷预测点模型。该模型通过采用不同天同时刻的负荷样本训练 LSSVR 来获取负荷的最优线性回归函数, 实现了在最小化负荷样本点误差的同时, 缩小模型泛化误差的上界, 获取了较好的负荷预测性能。

关键词:电力系统; 短期负荷预测; 最小二乘支持向量机

Abstract: Short-term load forecasting is a very important task of power system. Accurate short-term load forecasting is meaningful for the economical, safe and credible operation of power system. With the development of power system, especially the development of power market, the forecasting method with high accuracy must be researched. A new method of pre-disposing history data is proposed based on the daily periodicity of load, which lays the foundation for using history data by the model. Least square support vector regression (LSSVR) algorithm is a new generation of machine learning algorithms. So based on daily periodicity and the same moment similarity of load, a LSSVR-based short-term load forecasting model is put forward. The optimal load linear regression function is obtained by use of the samples in the same time but in the different days to train LSSVR. While the minimum load sample error is achieved, the model generalization error on the sector is reduced, so a better load forecasting performance is obtained.

Key words: power system; short-term load forecasting; least square support vector machine

中图分类号: TM714 **文献标志码:** A **文章编号:** 1003-6954(2009)增-0010-05

0 引言

电力系统短期负荷预测是一项非常重要的工作^[1], 准确的短期负荷预测对于电力系统经济、安全、可靠的运行具有特别重要的意义。随着电力系统的日趋复杂化, 特别是电力市场的逐步深入, 短期负荷预测被赋予了更高的要求^[2]。

支持向量机^[3]是一种基于结构风险最小化原则和小样本学习理论的实用化机器学习方法, 它克服了神经网络方法存在一些缺陷^[4, 6~8]。同时, 最小二乘支持向量机回归模型将标准支持向量机模型中的损失函数设定成误差平方和, 并把不等式约束改成等式约束, 这样既减少了标准支持向量机模型中的待定参

数, 又将求解二次规划的问题转化成线性 KKT 方程组的求解, 极大地降低了求解的复杂性, 提高了支持向量机的实用性。

通过对电力负荷特性的分析, 认识到不同天同一时刻的电力负荷具有大致相近的从影响因素到负荷值的函数映射关系, 进而根据负荷数据预处理应充分利用待处理数据点前后向历史数据的要求, 提出了基于负荷日周期性进行前后向外推的数据预处理新方法。同时应用以结构风险最小化原则为理论基础的最小二乘支持向量机, 较好地解决了小样本、非线性、高维数、局部极小点等实际问题, 所以将其应用于电力系统短期负荷预测, 在充分利用日周期性和同时刻负荷相近性的基础上, 提出了基于最小二乘支持向量机^[9]回归算法 (LSSVR) 的短期负荷预测点模型, 获取了较好的负荷预测性能。

基金项目:国家重点研究发展计划项目 (973 项目) (2004CB217905)

1 负荷数据预处理方法

负荷预测模型的性能在很大程度上取决于负荷历史资料的质量,但是采集到的初始负荷数据难免存在错误和缺漏^[1],在总结和归纳多种负荷数据分析及预处理方法的基础上^[5],使用“偏离率”统计学方法进行负荷“异常数据”的查找,并首创性提出了一种基于日周期前后向外推的负荷数据预处理的新方法,用于该文的数据预处理。

对异常数据的处理包括两个方面:①缺失负荷数据的补遗;②非缺失类的异常数据预处理。

1.1 缺失负荷数据的补遗

在众多的短期负荷一步预测法中,考虑到基于负荷日周期多点外推法是依据负荷日周期性、同类型日负荷相似性进行外推预测,能综合利用日负荷纵向和横向信息,并具有简单、快速的优点,而且它克服了基于日周期单点外推的预测精度对单点负荷值依赖过大、易受该点负荷采样记录值影响的缺陷,因此采用该方法进行单点缺失数据预处理,其建模过程如下。

设修补点为: $L(d, t)$, 该时刻前 k 点负荷值为: $L(d, t-k)$, 其中 $k=1, 2, \dots, m$, 用 d_t 表示修补点所在日的同类型日, 那么, 如果假定同类型日的同时刻的负荷增量相近, 则可以用历史上的同类型的日负荷增量 ΔL 来代替修补点所在日同时刻的负荷增量, 则有:

$$L(d, t) = L(d, t-k) + \Delta L = L(d, t-k) + [L(d_t, t) - L(d_t, t-k)] \quad (1)$$

取 $k=1, 2, \dots, m$, 将式 (1) 写成 m 个式子, 然后各式相加并整理有

$$L(d, t) = \frac{1}{m} \sum_{k=1}^m [L(d, t-k) - L(d_t, t-k)] + L(d_t, t) \quad (2)$$

如果在历史数据库中取 s 个同类型日, 将式 (2) 写成 s 个式子, 合并整理可得

$$L(d, t) = \frac{1}{s} \sum_{i=1}^s \left\{ \frac{1}{m} \sum_{k=1}^m [L(d, t-k) - L(d_i, t-k)] \right\} + L(d_i, t) \quad (3)$$

为了更充分地利用已有历史数据库的信息, 并进一步提高修补点的修补精度和稳定性, 考虑利用修补点后 k 点负荷值 $L(d, t+k)$ 来进行向前外推, 类似于上述推导过程, 最终可得到

$$L(d, t) = \frac{1}{s} \sum_{i=1}^s \left\{ \frac{1}{m} \sum_{k=1}^m [L(d, t+k) - L(d_i, t+k)] \right\} + L(d_i, t) \quad (4)$$

然后利用式 (3) 和 (4) 的结果, 计算平均值作为最终的修补结果。也就是所提出的基于日周期多点前后向外推法。这样的好处在于: 计算修补点的负荷值不仅利用了该时刻前 m 个时段来外推, 而且也利用了该时刻后 m 个时段的负荷值来外推, 因此综合地利用了已有历史数据的信息, 从而更充分地应用了同类型日在修补时刻的前后各时段的负荷变化趋势。

1.2 非缺失类的异常数据预处理

对于非缺失类的异常数据处理, 首先需要按一定原则将异常数据从负荷数据库中辩识出来, 然后再进行预处理。因此利用这种负荷特性进行异常数据的辩识, 其思路是: 对不同天同时刻的负荷值求出其期望值, 并计算不同天同时刻的历史负荷值相对于该期望值的变化量, 再对该变化量选取一个合适的阈值, 当变化量超过这个阈值时, 就认为其对应的负荷值是异常值, 最后进行预处理。其建模过程如下。

设 $L(d, t)$ 表示二维矩阵历史负荷数据, 所以每个时刻历史负荷数据可用式 (5) 和 (6) 求出其均值 $E(t)$ 和标准差 $\delta(t)$ 。

$$E(t) = \frac{1}{m} \sum_{d=1}^m L(d, t) \quad (5)$$

$$\delta(t) = \sqrt{V(t)} = \sqrt{\frac{1}{m} \sum_{d=1}^m (L(d, t) - E(t))^2} \quad (6)$$

定义第 d 天 t 时刻的负荷偏离率为 $\omega(d, t)$ 如下。

$$\omega(d, t) = \frac{|L(d, t) - E(t)|}{\delta(t)} \quad (7)$$

由概率论原理, 可以设定一个偏离率上限值 ω_0 , 当某天某时刻的负荷值满足式 (8), 则认为该点负荷值是异常的, 否则认为该点正常。

$$\omega(d, t) \geq \omega_0 \quad (8)$$

在异常数据辩识后, 仍按“缺失负荷数据的补遗”方法进行负荷数据预处理。

2 LSSVR 的短期负荷预测点模型

考虑到最小二乘向量机^[9]既具有支持向量机的结构风险最小化、非线性拟合能力好、泛化能力强等特性, 又具有求解简单、计算快速、待定参数少等优势, 这里将其应用于电力系统短期负荷预测^[10], 在建

模过程中利用短期负荷的日周期性和不同天相同时刻负荷的相近性,通过采用不同天同时段的历史负荷值来组建样本训练集,构建了基于最小二乘支持向量机回归算法的短期负荷预测点模型,并通过算例展示了该模型的预测过程。

2.1 最小二乘支持向量机回归 (LSSVR) 模型

支持向量机能拟合复杂的非线性模型并具有良好的泛化能力,但是在建模过程中涉及参数较多并需求解二次规划问题,这就大大地限制了它在实践中的运用。从结构风险最小化原则出发将函数回归中的最小二乘问题转化为支持向量机形式的问题加以解决,提出了最小二乘支持向量机的回归估计 (LSSVR) 数学模型,它将标准支持向量机模型中的损失函数设定成误差平方和并把不等式约束改成等式约束。这样,既减少了标准支持向量机模型中的待定参数,又将求解二次规划的问题转化成线性 KKT 方程组的求解,极大地降低了求解的复杂性,提高了支持向量机的实用性。

对于样本训练集 $\{(x_i, y_i) | i=1, 2, \dots, n\}$, 其中, $x_i \in R^n$, $y_i \in R$ 分别为输入和输出目标值,用非线性映射 $\Psi(\cdot)$ 将样本输入从原空间映射到高维特征空间,可以构造出如下最优线性回归函数。

$$f(x) = \omega^T \varphi(x) + b \quad (9)$$

根据结构风险最小化准则,求解上述回归问题的最小二乘支持向量机模型如下。

$$\min \frac{1}{2} \|\omega\|^2 + \frac{1}{2} \gamma \sum_{i=1}^{n-1} \xi_i \quad (10)$$

$$s.t. y_i - \omega^T \varphi(x_i) - b = e_i, (i=1, 2, \dots, n)$$

引入拉格朗日函数将上述约束优化问题转变为无约束优化问题,在对偶空间得到下式。

$$L(\omega, b, \xi, \alpha) = \frac{1}{2} \|\omega\|^2 + \frac{1}{2} \gamma \sum_{i=1}^{n-1} \xi_i^2 + \sum_{i=1}^n \alpha_i [\omega^T \varphi(x_i) + b + e_i - y_i] \quad (11)$$

根据 KKT (Karush-Kuhn-Tucker) 优化条件可得如下等式约束条件。

$$\begin{cases} \frac{\partial L}{\partial \omega} = 0 \rightarrow \omega = \sum_{i=1}^n \alpha_i \varphi(x_i) \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^n \alpha_i = 0, i=1, \dots, n \\ \frac{\partial L}{\partial \xi_i} = 0 \rightarrow \alpha_i = \gamma e_i, i=1, \dots, n \\ \frac{\partial L}{\partial \alpha_i} = 0 \rightarrow \omega^T \varphi(x_i) + b + e_i - y_i = 0, i=1, \dots, n \end{cases} \quad (12)$$

对上式消去变量 e_i 和 w 后,可以得到如下线性方程组。

$$\begin{bmatrix} 0 & 1_n^T \\ 1_n & ZZ^T + I/\gamma \end{bmatrix}_{(n+1) \times (n+1)} \begin{bmatrix} b \\ d \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (13)$$

根据 Mercer 条件,可定义如下核函数

$$K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j) \quad (14)$$

则式 (13) 变为

$$\begin{bmatrix} 0 & 1 & \dots & 1 \\ 1 & K(x_1, x_1) + \frac{1}{\gamma} & \dots & K(x_1, x_n) \\ \dots & \dots & \dots & \dots \\ 1 & K(x_p, x_1) & \dots & K(x_p, x_n) + \frac{1}{\gamma} \end{bmatrix} \begin{bmatrix} b \\ \alpha_1 \\ \dots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} 0 \\ y_1 \\ \dots \\ y_n \end{bmatrix} \quad (15)$$

求解上式得到 α 和 b 后,可得到 LSSVR 的最优线性回归函数。

$$f(x) = \sum_{i=1}^n \alpha_i K(x_i, x) + b \quad (16)$$

2.2 基于 LSSVR 短期负荷预测的点模型

由负荷特性分析结果得知“不同天同一时刻负荷近似服从相近或同一的影响因素与负荷之间的函数映射关系;而不同时刻的负荷与影响因素之间的非线性映射函数关系是不同的。”该节从这个特性出发,构建 LSSVR 短期负荷预测的点模型。为此,对预测日各预测时刻分别构建不同的 LSSVR 回归预测函数式。在这个过程中需要确定样本的构建、训练样本集的组成、核函数及模型各项参数等等,用以最终可获得各时刻的负荷回归函数。

(1) 负荷历史数据的选取

在建模过程中,取用预测日前 2 个月左右的负荷数据来建立预测模型。

(2) 样本构建

样本构建的主要任务是确定输入向量,也就是确定负荷的主要影响因素。这里以预测时刻负荷 $L(d, t)$ 为输出目标值,考虑到相近日和同类型日负荷变化的相关性,选择如下形式的输入向量。

$$[L(d-7, t-2), L(d-7, t-1), L(d-7, t), L(d-2, t-2), L(d-2, t-1), L(d-2, t), L(d-1, t-2), L(d-1, t-1), L(d-1, t)]$$

进行输入-输出样本对的构建。

(3) 样本训练集的确定

根据前面负荷特性分析的结论,可知不同天同一时刻负荷具有相近的输入-输出函数映射关系。所以本节用不同天同一时刻的目标值来构建参与 LSS-

VR 训练的样本集。

(4) 核函数的选取

LSSVR 由训练样本集和核函数完全刻画, 选取不同形式的核函数就可以生成不同的 LSSVR 回归模型, 但是必须要满足 Mercer 定理的函数才能作为核函数。

考虑到高斯函数具备表示形式简单且解析性好、径向对称、光滑性好等优点。此外, 选取高斯径向基核函数 $k(x, y) = \exp(-\|x - y\|^2 / 2\sigma^2)$ 还可以使回归模型的每个支持向量 (即每个样本数据点) 处产生一个以其为中心的局部高斯函数, 使用结构风险最小化原则, 能找出全局的基函数宽度 δ 因此选择它作为 LSSVR 的核函数。

(5) 参数的选取

为了构建基于 LSSVR 的短期负荷预测点模型, 还需要确定核参数 δ 和平衡参数 γ 。这里采用网络搜索法来获得一组较优的模型参数, 在搜索的过程中运用交叉验证法来评估模型性能。

2.3 基于 LSSVR 点模型的负荷预测步骤

用 LSSVR 点模型方法进行电力系统短期负荷预测的步骤如下。

(1) 对历史数据进行预处理, 并截取适宜长度的历史负荷数据作为建模的基础负荷数据;

(2) 对建模的基础负荷数据进行归一化, 这里的归一化公式如下:

$$L = (L - L_{\min}) / (L_{\max} - L_{\min}) \quad (17)$$

(3) 按该节所述方法组建输入-输出样本对, 同时构建预测输入样本;

(4) 根据不同的预测时刻确定各自的训练样本集, 并用训练样本集, 建立如式 (15) 形式的各时刻线性方程组;

(5) 求取各时刻的 α_i 和 b

(6) 将各时刻的 α_i 和 b 代入式 (16), 形成各时刻的最优线性回归函数,

(7) 利用预测输入样本和形成的最优线性回归函数, 对未来某时刻的负荷进行预测。

3 算例分析

利用本章所提的 LSSVR 点模型对甘肃电网的负荷数据进行仿真预测。根据所获得的历史数据情况, 截取 2005 年 1 月 1 日到 2 月 23 日作为建模基础数

据, 对 2 月 24 日 24 个整点负荷进行预测, 其预测结果及误差情况见表 1。

表 1 基于 LSSVR 点模型的 24 小时整点负荷预测结果

时间 (2005. 2. 24)	实际值 /MW	LSSVR 点模型预测值		
		预测值 /MW	绝对误差 /MW	相对误差 /%
0:00	4 330	4 332.8	-2.810 0	-0.064 9
1:00	4 270	4 268.2	1.782 2	0.041 7
2:00	4 240	4 244.5	-4.493 0	-0.106 0
3:00	4 220	4 237.9	-17.901 7	-0.424 2
4:00	4 360	4 349.5	10.487 2	0.240 5
5:00	4 430	4 420.8	9.153 9	0.206 6
6:00	455 0	454 2.6	7.389 9	0.162 4
7:00	4 540	4 531.3	8.742 7	0.192 6
8:00	471 0	4 720.1	-10.059 3	-0.213 6
9:00	4 770	4 800.4	-30.359 1	-0.636 5
10:00	4 780	4 807.5	-27.549 5	-0.576 3
11:00	4 690	4 696.4	-6.408 7	-0.136 6
12:00	4 570	4 590.7	-20.721 4	-0.453 4
13:00	4 530	4 561.2	-31.231 4	-0.689 4
14:00	4 510	4 548.4	-38.366 3	-0.850 7
15:00	4 500	4 538.5	-38.478 0	-0.855 1
16:00	4 660	4 684.2	-24.217 1	-0.519 7
17:00	4 770	4 791.1	-21.058 3	-0.441 5
18:00	5 030	5 059.1	-29.007 0	-0.576 7
19:00	5 200	5 207.1	-7.298 8	-0.140 4
20:00	5 170	5 191.1	-21.484 8	-0.415 6
21:00	5 060	5 057.4	2.611 1	0.051 6
22:00	4 840	4 815.8	24.167 7	0.499 3
23:00	4 480	4 475.3	4.657 2	0.104 0
平均误差 (绝对值)			16.684 8	0.358 3
最大误差 (绝对值)			38.478 0	0.689 4

2 月 24 日整点负荷预测曲线和实际负荷曲线如图 1 所示。运用 LSSVR 点模型方法预测该省 2005 年 2 月 14 日到 2 月 20 日连续一周的负荷, 其一周的负荷整点预测曲线如图 2 所示。

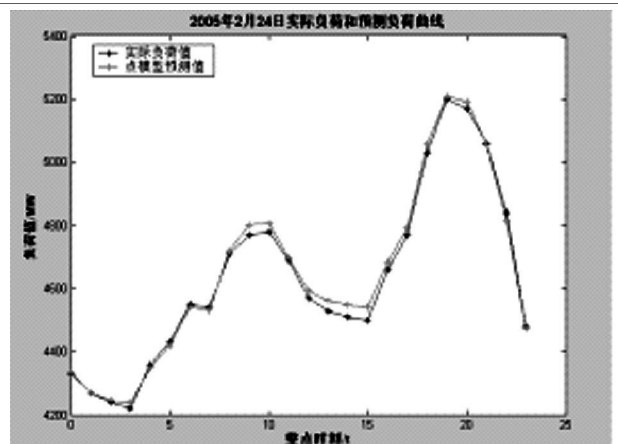


图 1 2 月 24 日实际负荷和预测负荷曲线

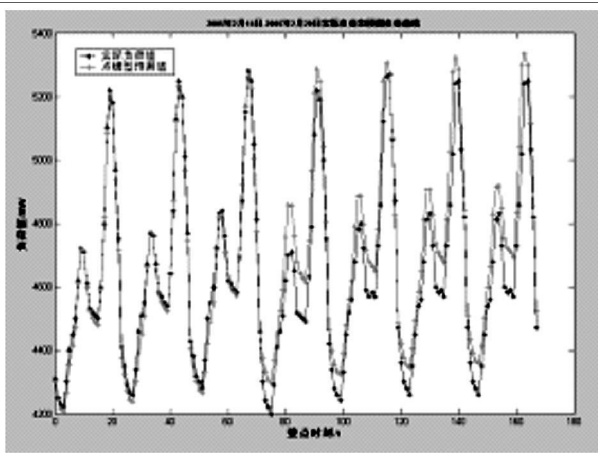


图 2 连续一周的实际负荷和预测负荷曲线

从以上预测结果和统计表可以看出,采用 LSSVR 点模型所得到的预测曲线基本上能与实际负荷曲线相符合。但是仍存在许多不足,例如该模型在负荷的峰、谷段预测效果不太理想;而且周末休息日与工作日相比,其预测效果偏差。因此有必要改进预测模型,进一步提高预测精度。

4 结 论

前面在系统地探讨了国内外多种短期负荷预测技术的基础上,围绕如何提高预测精度、降低预测风险这两大主题,充分利用 LSSVR 优越的非线性学习及推广性能提出了基于 LSSVR 的短期电力负荷预测模型,并通过仿真证明了其可行性和有效性,在电力系统短期负荷预测研究方面取得了较好的成果。在此,将主要的研究成果总结如下。

(1)提出了不同天同一时刻的电力负荷具有近似相同的数学分布特征的假设,并认为不同天同一时刻的电力负荷具有大致相近的从影响因素到负荷值的函数映射关系。

(2)提出了使用统计学中的“偏离率”方法进行负荷“不良数据”查找,进而考虑到负荷数据预处理的特点,提出了基于负荷日周期性进行前后向外推的数据预处理新方法,并在数据预处理中取得了较好的效果。

(3)将最小二乘支持向量机回归算法用于短期负荷预测,基于不同天同时刻的负荷具有相近的函数

映射关系的假设,构建了短期负荷预测的 LSSVR 点模型,该模型具有较强非线性拟合能力和较好的推广能力,在甘肃电网的短期负荷预测仿真中取得了一定效果。

参考文献

- [1] 牛冬晓,曹树华,赵磊,等. 电力负荷预测技术及其应用 [M]. 北京:中国电力出版社, 1998.
- [2] 金卓睿. 浅谈如何搞好电力市场中的负荷预测 [J]. 四川电力技术, 2000, 23(6): 1-5.
- [3] Moulin L S da Silva A P A, El-Sharkawi M A, et al. Support vector machines for transient stability analysis of large-scale power systems [J]. IEEE Transactions on Power Systems 2004, 19(2): 818-825.
- [4] 李云天,孙艳鹤,范广良. 基于弹性系数的电力系统短期负荷预测方法 [J]. 吉林电力, 2003, 3: 26-28.
- [5] 瓦普尼克著,张学工译. 统计学习理论 [M]. 北京:电子工业出版社, 2004.
- [6] Vapnik V. An overview of statistical Learning Theory [J]. IEEE Transactions on Neural Networks 1999, 10(5): 988-999.
- [7] Mukherjee S, Osuna E, Girosi F. Nonlinear prediction of chaotic time series using support vector machines [C]. Proceedings of the 1997 IEEE Neural Networks for Signal Processing Amelia Island FL 1997: 511-520.
- [8] Suykens J A K, Gestel T V, Brabanter J De et al. Least Squares Support Vector Machines [M]. Singapore: World Scientific 2002.
- [9] 章健. 电力系统负荷模型与辨识 [M]. 北京:中国电力出版社, 2007.
- [10] 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.
- [11] 陈亚红,穆钢,段方丽. 短期电力负荷预报中几种异常数据的处理 [J]. 东北电力学院学报, 2002, 22(2): 1-5.

作者简介:

侯贺飞 (1986-), 女, 硕士研究生, 研究方向为电力市场及需求特性。

刘俊勇 (1963-), 男, 教授, 博士生导师, 长期从事电力系统稳定分析及电力市场等方面的研究。

(收稿日期: 2009-11-04)