

基于相关向量机的电力负荷中期预测

刘遵雄¹, 张德运¹, 孙钦东¹, 徐 征²

(1. 西安交通大学电子与信息工程学院, 710049, 西安; 2. 华东交通大学电子与电气工程学院, 330013, 南昌)

摘要: 针对电力负荷中期预测比较困难并且存在较大误差的问题,提出了一种基于相关向量机的中期预测方法. 结合 EUNITE 网络提供的实际数据,研究了日最大负荷前后期关系、日最大负荷与节假日的关系和当日与对应星期数的相关性,并建立了相应的电力负荷中期预测模型. 该模型是将与某天相关的 n 个前期信息作为该天的日最大负荷的输入量,而日最大负荷与节假日、当日(星期数)的关系信息用两个二元值表示. 在模型训练前,将输入量的前 7 个属性值和预测目标值进行归一化处理. 采用不同训练样本集的仿真实验结果表明,相关向量机方法比支持向量机方法具有更多的优点,当高斯核函数的宽度值取为 210 时,相关向量机方法具有较为理想的预测效果.

关键词: 电力负荷; 中期预测; 相关向量机; 模型实验

中图分类号: TM 715; TP 18 **文献标识码:** A **文章编号:** 0253 - 987X(2004)10 - 1005 - 04

Mid-Term Electric Load Prediction Based on the Relevant Vector Machine

Liu Zunxiong¹, Zhang Deyun¹, Sun Qindong¹, Xu Zheng²

(11School of Electronics and Information Engineering, Xi an Jiaotong University, Xi an 710049, China; 21School of Electronics and Electric Engineering, East China Jiaotong University, Nanchang 330013, China)

Abstract: The mid-term electric load prediction is an existing difficult problem whose predicted solution often has a larger error. An RVM (relevant vector machine)-based mid-term prediction method is proposed for solving the problem. With the practical data provided by EUNITE-network, the relations of before-after duration of daily maximum load, the relation between daily maximum load and holidays, and existing relation between the day's maximum load and the corresponding weeks are investigated and the model of the mid-term electric load prediction is constructed. In the model, the preceding information related to certain day is regarded as daily maximum load of that day, and the relation information between the daily maximum load and holidays, number of weeks (the day) is represented as two bi-values. Before training the model, the preceding 7 attribute values of input variables and the predictive goal value are normalized. Using different train sample sets, the simulation experiment result demonstrates that the RVM method has more advantages than support vector machine method. When the Gaussian kernel function width is 210, RVM method possesses more perfect prediction performance.

Key words: electricity load; mid-term forecasting; relevant vector machine; model experiment

电力负荷预测是电力规划和系统运行中非常重要的一项工作,其工作量大,而且需要反复进行. 它通常依据一定的信息,如天气及前期负荷数据来建立预测模型. 在预测中,较为广泛使用的方法是神经

网络预测法^[1-3],它们大多是用来解决短期负荷预测问题. 电力负荷中的长期预测需要预测一段时期的负荷情况,由于存在误差传播,这种预测在可靠性上存有问题,并将影响电力系统生产调度计划、电

收稿日期: 2003 - 11 - 05. 作者简介: 刘遵雄(1967~),男,博士生,现任华东交通大学信息工程学院讲师;张德运(联系人),男,教授,博士生导师.

厂机组起停计划的制定和实施。

在解决电力负荷中期预测的问题上,目前已有的方法主要是基于时间序列模型和回归模型,包括神经网络预测法和支持向量机(SVM)法^[4]等。本文结合 EUNITE 网络提供的实际数据^[5],研究了日最大负荷前后期关系、日最大负荷与节假日的关系和当日负荷与对应星期数的相关性,建立了电力负荷中期预测模型,提出了一种基于相关向量机的中期预测方法。

1 相关向量机

SVM 是近来机器学习领域的研究热点之一,它通过最小化推广性错误的上界获得良好的推广性能。相关向量机(RVM)是 Michael E. Tipping 提出的一种与 SVM 相似的稀疏概率模型^[6],它的训练是在贝叶斯框架下进行的,可以用它进行回归估计预测,从而获得预测值的分布。

Tipping 将拉格朗日系数向量 α 作为权参数,建立了相应的显著度框架,由此求得的非零 α_i 所对应的学习样本被称为相关向量(RV)。RV 相当于 SVM 中的支持向量,但对同一个训练样本集而言,RV 的数目比 SVM 的支持向量数目少得多,因此 RVM 比 SVM 所用的测试时间短。由于 RVM 需要不断迭代计算超参数,因此它比 SVM 所用的训练时间长。Bishop 等人曾利用变分法求取 RVM 的显著度,并给出了 RVM 的误差条^[7]。Chen 等人将 RVM 用于信道均衡^[8],其仿真实验表明 RVM 与 SVM 的性能相当,但生成的 RV 占训练样本总数的比例很小,大大提高了信道均衡的速度。

设 RVM 的模型输出为

$$y(x) = \sum_{j=1}^M \alpha_j \phi_j(x) + \epsilon \quad (1)$$

式中: $\{\phi_j\}$ 为非线性基函数; $\{\alpha_j\}$ 为模型的权值。SVM 使用的基函数必须满足 Mercer 核定理,而 RVM 方法对基函数无此限制,所以我们选用以每个训练样本为中心的高斯函数作为基函数。

在定义了模型(1)的基函数之后,可以在贝叶斯框架下用最大似然方法来训练模型权值 $\{\alpha_j\}$,这样可回避过学习问题,提高模型的泛化能力。因此,RVM 为每个权值定义了先验概率分布

$$p(\alpha_j | a_j) = \left(\frac{a_j}{2}\right)^{1/2} \exp\left[-\frac{1}{2} a_j \alpha_j^2\right] \quad (2)$$

式中: a_j 是决定权值 α_j 先验分布的超参数。

给定训练样本集 $\{x_i, t_i\}_{i=1}^N$,假定目标值 t_i 是独

立的,并且数据的噪声服从方差为 σ^2 的高斯分布,则相应的训练样本集的似然函数为

$$p(t | \alpha, \sigma^2) = (2\sigma^2)^{-N/2} \exp\left[-\frac{1}{2\sigma^2} (t - \Phi \alpha)^T (t - \Phi \alpha)\right] \quad (3)$$

式中: $t = (t_1, \dots, t_N)^T$; $\alpha = (\alpha_1, \dots, \alpha_N)^T$; Φ 为矩阵,其行包含所有基函数对输入 x_i 的响应 $(\phi_j)_i = [1, \phi_1(x_i), \dots, \phi_N(x_i)]$ 。

根据先验概率分布和似然分布,再用贝叶斯公式计算权值的后验概率分布,即

$$p(\alpha | t, a, \sigma^2) = \frac{p(t | \alpha, \sigma^2) p(\alpha | a)}{p(t | a, \sigma^2)} \quad (4)$$

而该权值的后验分布属于多变量高斯分布,即

$$p(\alpha | t, a, \sigma^2) = N(\mu, \Sigma) \quad (5)$$

式中: $\Sigma = (\sigma^2 \Phi^T \Phi + A)^{-1}$ 为协方差,其中的 A 是 (a_0, \dots, a_N) 的对角矩阵; $\mu = \sigma^2 \Phi^T t$ 为均值。

训练目标值的似然分布通过对权值变量进行积分,即

$$p(t | a, \sigma^2) = \int p(t | \alpha, \sigma^2) p(\alpha | a) d\alpha \quad (6)$$

实现边缘化,从而求得超参数的边缘似然分布

$$p(t | a, \sigma^2) = N(0, C) \quad (7)$$

其中的协方差 $C = \sigma^2 I + A^{-1} \Phi^T \Phi$ 。

最后,RVM 方法中的模型权值的估计值由后验分布的均值给出,同时它也是权值的最大后验(MAP)估计。权值的 MAP 估计取决于超参数 a 和噪声方差 σ^2 ,它们的估计值 \hat{a} 和 $\hat{\sigma}^2$ 可以通过最大化边缘似然分布得到。后验分布反映出的权值最优值的不确定性,可以表示模型预测的不确定性。若给定输入值 x_3 ,相应输出的概率分布

$$p(t_3 | x_3, \hat{a}, \hat{\sigma}^2) = \int p(t_3 | x_3, \alpha, \hat{\sigma}^2) p(\alpha | t, \hat{a}, \hat{\sigma}^2) d\alpha \quad (8)$$

服从高斯分布的形式,即

$$p(t_3 | x_3, \hat{a}, \hat{\sigma}^2) = N(y_3, \sigma_3^2) \quad (9)$$

其中的预测均值和方差(不确定性)分别为

$$y_3 = \mu^T(x_3) \\ \sigma_3^2 = \hat{\sigma}^2 + \Sigma^T(x_3) \Sigma(x_3)$$

RVM 使用贝叶斯框架下的显著度解决了模型的参数选取问题,具有较好的适用性。将 RVM 用于回归预测,可以较好地得出预测值及其方差范围。

2 电力负荷中期预测模型

采用 EUNITE 网络给出的比赛数据^[5],使用

RVM 方法来预测 1999 年 1 月的日负荷最大值. EU2 NITE 数据是 1997 年至 1998 年期间每 015 h 记录一次的日电力负荷数据和两年节假日的日电力负荷数据分布情况. 通过分析两年的日最大电力负荷数据,参考文献[4]和文献[9]的解决方法,并在考虑了日最大负荷前后期关系,日最大负荷与节假日、当日(星期数)关系的基础上,建立了电力负荷中期预测模型.

假定 y_i 是第 i 天的日最大负荷,一般将与当天相关的 n 个前期信息 $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$ 作为对应 y_i 的输入量.

首先引入时间序列概念,在进行输入量选择时,必须将前期的日最大负荷量作为对应现期日负荷最大值的输入量属性. 具体地说, y_i 是第 i 天的预测目标值,而前 1, 2, ..., n 天的日最大负荷值 $y_{i-1}, y_{i-2}, \dots, y_{i-n}$ 应作为相应输入向量 x 的分量. 例如,已求得了 1999-01-01 的日最大负荷值(取 $n=7$),可将其与 1998-12-26~1998-12-31 每天的日负荷最大值一起作为预测 1999 年 1 月 2 日的日负荷最大值的分量. 具体选取哪几天的日负荷最大值作为预测输入量的分量是一个现实的问题,考虑到所取负荷数据呈周期性变化趋势,可以取 $n=7$. x_i 的前 7 个量表示预测日的前 7 天的日负荷最大值.

日最大负荷与节假日、当日(星期数)的关系信息,也必须作为日负荷最大值对应的输入量的分量. 用两个二元取值量表示当日星期数信息,即当日是周正常工作日,其输入量第 8 属性值取 1,相应的第 9 属性值取 0,否则输入量第 8 属性值取 0,相应的第 9 属性值取 1. 用一个二元取值量表示节假日信息,就是说预测值当日是节假日,其输入量的第 10 属性值取 1,否则取 0.

3 计算结果

在模型训练时,应同时输入各训练样本的输入属性值和对应的目标值. 为了提高模型的计算性,必须对数据进行归一化处理. 本例中需要归一化处理的是类型值为数值型的数据,即为输入量的前 7 个属性值和预测目标值. 由于数据的数值型数值小于 1 000,所以可以采用小数定标规范化处理方法^[10]将这些数值除以 1 000,以达到数据归一化的目的. 实验中,取高斯核函数的宽度值为 210,预测效果较为理想.

使用不同的样本分割方法进行了 3 种实验:选

用两年的全部样本数据训练 RVM 模型,并计算出 1999 年 1 月中每天的日最大负荷,记为 E_1 ;选两年中的 1 月~3 月和 10 月~12 月共 12 个月的数据,来训练 RVM 模型,得到预测值,记为 E_2 ;选两年中的 1 月、2 月、11 月和 12 月共 8 个月的数据,得到预测值,记为 E_3 . 3 组实验得到的 1999 年 1 月的日最大负荷预测值和真实值的图形如图 1 所示.

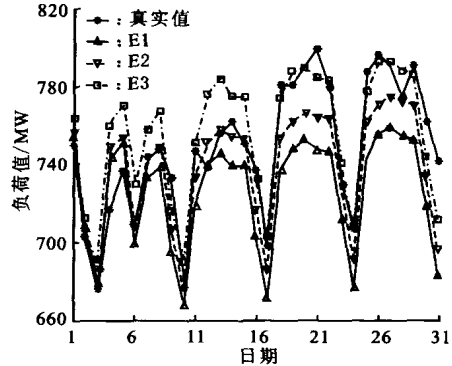


图 1 1999 年 1 月的日负荷真实值及 3 种实验预测值比较

可用相对平均误差来衡量不同方法和实验的预测性能,即

$$e_{MAPE} = 100 \frac{\sum_{i=1}^n | (L_i - \hat{L}_i) / L_i |}{n}, \quad n = 31 \tag{10}$$

式中: L_i 和 \hat{L}_i 分别为预测期的第 i 天日最大电力负荷需求量的预测值和真实值.

日最大负荷值相对于真实值的误差可用最大误差

$$e = \max(L_i - \hat{L}_i), \quad i = 1, 2, \dots, 31 \tag{11}$$

来衡量. 其中的 e 表示 31 天中的日最大负荷预测值相对于真实值的最大偏差. 表 1 列出了 3 种实验的结果.

表 1 3 种实验的 e_{MAPE} 和 e 值

	E_1	E_2	E_3
e_{MAPE}/MW	31367	21014	11947
e/MW	58179	44133	42162

从表 1 可以发现,实验 2 和实验 3 都取得了比较理想的预测结果,它们比文献[4]方案所获得的结果更好. 本模型实验 E_3 的 e_{MAPE} 值小于 2,而文献[4]方案的 e_{MAPE} 值约等于 2. 计算两组实验的日最大负

荷值与真实值最大的误差绝对值分别为 44133 MW 和 42162 MW,而文献[4]方案的该值在 50 MW 以上.

根据表 1 可知,实验 E_3 的预测效果优于实验 E_2 的,这反映出在使用 RVM 进行 1999 年 1 月份的日最大负荷预测时,选择 1997 年和 1998 年数据样本中 1 月和 2 月、11 月和 12 月的数据作为训练样本,训练后获得的相关向量能更为精确地反映需要预测月份的负荷变化规律. 结果是使用 RVM 方法预测 1999 年 1 月的日最大负荷不仅可以获得对应点的中心值,而且能够计算出对应点的方差. 方差范围越小,预测的准确性越好. 3 次实验中 E_3 预测结果的方差偏小,其值小于 19 MW. 图 2 表示 E_3 的预测值和对应的上、下界.

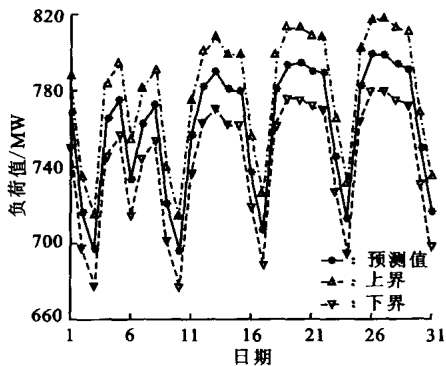


图 2 试验 E_3 的预测值及其上、下界

4 结 论

综合上述,可以认为相关向量机是一种理想的、适用于回归估计预测的方法,用它可取得比 SVM 更好的预测效果. RVM 的优越性总结如下.

(1) 不仅仅输出预测目标量的点估计值,还可以输出预测值的分布.

(2) 使用更少数量的支持向量,从而显著减少输出目标量预测值的计算时间.

(3) RVM 不需要估计过多的参数.

(4) RVM 对是否满足 Mercer 定理的核函数没有

限制,适应性更好.

作为一种新型的稀疏模型学习方法,RVM 在回归预测时不仅输出预测值,而且能输出预测值方差,它非常适用于有实时性要求的工程预测估计.

参考文献:

- [1] Park D C, El - Sharkawi M A, Marks R J, et al. Electric load forecasting using an artificial neural network [J]. IEEE Transactions on Power Systems, 1991, 6(2): 442 - 449.
- [2] Kholanzad A, Afkhami - Rohani R, Lu T L, et al. ANNSILF: a neural network based electric load forecasting system[J]. IEEE Trans on Neural Networks, 1997, 8(4): 835 - 846.
- [3] Hippert H S, Pedreira C E, Souza R C. Neural networks for short - term load forecasting: a review and evaluation [J]. IEEE Transactions on Power Systems, 2001, 16 (1): 44 - 55.
- [4] Chen B J, Lin C J. Load forecasting using support vector machine [EB/OL]. <http://www.csie.ntu.edu.tw/~cjlin/pa2/pers.html>, 2003 - 07 - 20.
- [5] Otto P. World - wide competition within the EUNITE network [EB/OL]. <http://neuron.tuke.sk/competition/>, 2003 - 05 - 13.
- [6] Tipping M E. Sparse Bayesian learning and the relevance vector machine [J]. Journal of Machine Learning Research, 2001, 1(3): 211 - 244.
- [7] Bishop C M, Tipping M E. Variational relevance vector machine [A]. The 16th Conf on Uncertainty in Artificial Intelligence, Morgan Kaufmann, USA, 2000.
- [8] Chen S, Gunn S R, Harris C J. The relevance vector machine technique for channel equalization application [J]. IEEE Trans on Neural Networks, 2002, 12(6): 1 529 - 1 532.
- [9] Zivcak D. Electricity load forecasting using ANN [EB/OL]. <http://neuron.tuke.sk/competition/reports/DaliborZivcak.pdf>, 2003 - 06 - 12.
- [10] 韩家炜, Micheline K. 数据挖掘概念与技术[M]. 范明译 1 北京:机械工业出版社, 2001.

(编辑 苗 凌)