

基于核主成分分析-支持向量机模型的电力负荷最大值短期预测方法研究

潘巧波

华东电力科学研究院有限公司, 浙江省 杭州市 310030

Short-term Prediction Method of Power Load Maximum Based on KPCA-SVM Model

PAN Qiaobo

Huadian Electric Power Research Institute Co., Ltd., Hangzhou 310030, Zhejiang Province, China

摘要: 电力负荷最大值是电网企业调度工作的重要组成部分, 其预测结果的准确度将对电能的配送、有效利用率、供电服务的质量以及电力系统的发展产生重要影响。该文以安徽某市 81 天的电力负荷最大值数据为基础, 选取影响当天电力负荷最大值的 10 个因素, 并采用 KPCA 算法将 10 维的影响因素降为 5 维, 其累计贡献率可达 93.70%。以降维后的 5 维数据为输入, 以径向基函数为核函数, 并采用交叉验证选择 SVM 回归的最佳参数, 随机选取 54 组数据训练 SVM 预测模型, 最后进行 27 组数据的拟合预测, 拟合预测的均方误差为 0.0041, 相关系数为 0.9631。研究表明应用 KPCA 结合的 SVM 预测模型对电力负荷最大值具有很好的预测能力。

关键词: 核主成分分析; 支持向量机; 电力负荷; 预测模型

ABSTRACT: The maximum of power load is an important part of power enterprise dispatching work, and the accuracy of its prediction results will have an important impact on the distribution of power, the effective utilization rate of power, the quality of power supply service and the development of power system. Based on the 81-day power load maximum data of a city in Anhui province, this paper selected 10 factors influencing the power load maximum of the day, and used KPCA algorithm to reduce the influencing factors from 10 dimensions to 5 dimensions. The cumulative contribution rate can reach 93.70%. The 5 d data which was reduced dimension as input, the radial basis function as kernel function, we used the cross validation to select the optimum parameters of SVM regression. Randomly we selected 54 sets of data to train SVM prediction model, then we selected 27 groups of data to fit prediction. The mean square error of fitting prediction is 0.0041, and the correlation coefficient is 0.9631. This shows that the KPCA combining the SVM prediction model for power load maximum has good prediction ability.

KEY WORD: KPCA; SVM; Power load; Prediction model

0 引言

电力系统负荷最大值的有效预测直接影响到电力调度是否高效, 电力能否得到合理分配与利用^[1-2]。当电力负荷最大值预测偏高的时候, 会导致当前电力设备闲置, 得不到充分利用, 从而造成超前投资和资源的浪费; 当电力负荷最大值预测偏低的时候, 就会造成电网的电能供应能力不足, 用户正常的电力需求不会得到满足, 并存在整个供电系统崩溃的风险。因此, 电力规划和调度部门应具备电力负荷最大值预测的基本数据信息, 从而制定调度合理、分配高效、节省

能源、持续可靠的科学决策和方案^[3-4]。

针对电力负荷最大值的预测问题, 目前研究主要集中于短期负荷最大值预测和中长期负荷最大值预测方面^[5]。对于短期电力负荷最大值预测, 通过预测未来 7~10 天左右的电力用量, 去让发电企业制定对应的发电策略, 以此让电力企业能够对市场的需求做出及时的反应^[6]。要保障每日电力负荷最大值预测的准确率, 选择适当的预测算法是十分关键的。

传统的预测方法采用固定的数学方程式建立预测模型, 其运算量相对较小, 计算速度可以

保证,但没有自学习、自适应能力^[7]。随着电力系统复杂度的演变,其时变性、非线性和不确定性特点日渐突出,这样就难以建立一个准确的数学模型来进行预测^[8]。为解决该问题,周涛等提出了一种改进的神经网络算法,以预测城市中长期电力需求,进而为城市电力系统的建设提供参考,但预测精度还需改进^[9]。此外,动态神经网络在预测领域使用较多的是 Elman 神经网络^[10]。但是,Elman 神经网络经常陷入局部极值且收敛速度慢,迭代时间长。这些不足在一定程度上影响了模型的预测精度。

为此结合上述的研究基础,笔者提出并建立了基于核主成分分析-支持向量机(Kernel Principal Component Analysis - Support Vector Machine, KPCA - SVM)的电力负荷最大值预测模型,并对此模型进行了详细设计与仿真验证。

1 电力负荷最大值的影响因素

要提高电力负荷最大值预测的准确性,就必须考虑到电力负荷最大值的影响因素。电力负荷最大值短期预测是以每天的电力负荷最大值为预测对象的,电力负荷最大值不同预测周期的影响因素是不一样的,主要有以下几方面^[11]。

(1) 天气因素

在温度、湿度、雨雪、霜雾等诸多表征天气的要素中,温度是相对最为重要的一个,此外天气类型也存在一定影响。

(2) 社会因素

社会因素包括人口数量、经济发展水平以及不同时间段的电价变化。

(3) 特殊事件因素

由于特殊事件,配备具有特殊效用的大型用电设备对电力负荷最大值存在一定影响。自然灾害等特殊情况也会对电力负荷最大值产生影响。

(3) 其它因素

日期类型(工作日、节假日)、季节需求差异、人口规模变化、空气质量状况、实时电价等因素,在某种程度上都对电力负荷最大值具有影响。

文章只分析每天的电力负荷最大值的短期

预测,因此其影响因素主要有温度、前一至两天的电力负荷最大值、当天的天气类型和日期类型等。

2 KPCA 与 SVM 算法描述

2.1 KPCA 算法

KPCA 属于机器学习的范畴,是对主成分分析(Principal Component Analysis, PCA)的非线性扩展^[12]。它是一种基于核的保留全局特征的非线性数据降维方法,它能够挖掘到数据集中蕴含的非线性信息^[13]。

KPCA 算法是通过非线性映射 $\Phi: R^N \rightarrow F$ 将输入空间映射到高维空间 F ,从而达到数据降维的目的^[14],在高维空间 F 中协方差矩阵为:

$$C = \frac{1}{M} \sum_{j=1}^M \varphi(x_j)(x_j)^T \quad (1)$$

分解式(1),可得到对应的特征方程:

$$\lambda v = Cv \quad (2)$$

其中, λ 为特征值 ($\lambda \geq 0$); v 为对应的特征向量。同时式(2)等价于式(3):

$$\lambda[\varphi(x_k), v] = [\varphi(x_k), Cv] \quad (3)$$

且 v 可作如下表示:

$$v = \sum_{i=1}^M \alpha \varphi(x_i) \quad (4)$$

联合式(1)、式(2)、式(3)可得:

$$\lambda \sum_{i=1}^M \alpha [\varphi(x_k), \varphi(x_i)] = \frac{1}{M} \sum_{i=1}^M \alpha [\varphi(x_k), \sum_{j=1}^M \varphi(x_j) \mathbf{I}[\varphi(x_j), \varphi(x_i)]] \quad (5)$$

引用核函数 $k(x_i, x_j) = [\varphi(x_i), \varphi(x_j)]$, 即可通过映射计算投影如下:

$$p(i) = \sum_{j=1}^M \alpha_j(j) \cdot K(x_i, x_j) \quad (6)$$

2.2 SVM 回归算法

在机器学习领域中, SVM 是一种有监督的

学习模型,可用于识别分类、回归预测^[15-16]。它对有限样本信息在模型中的复杂性进行最佳优化,可对小样本、非线性关系的问题进行很好的解决。

设 $\{x_i, y_i\} \in (R^d \times R), i=1,2,\dots,n$ 为输入样本, SVM 的回归算法就是通过非线性函数将其映射到 k 维特征空间,并建立新的线性模型^[17],具体方法如下:

$$f(x, w) = w \cdot \phi(x) + b \quad (7)$$

其中, $f(x, w)$ 为预测函数; w 为权数; $\phi(x)$ 为非线性映射函数集合; b 为阈值。函数计算是否具有-致性,需要引入不敏感损失函数进行验证:

$$L_e(y, f(x, w)) =$$

$$\begin{cases} 0 & |y - f(x, w)| \leq \varepsilon \\ |y - f(x, w)| - \varepsilon & \text{其他} \end{cases} \quad (8)$$

因为需要对误差进行拟合,引入松弛因子 ζ_i, ζ_i^* , 优化公式如下:

$$\min_{w, \zeta_i, \zeta_i^*} J(w, \zeta_i, \zeta_i^*) = \frac{1}{2} w^T w + C \sum_{k=1}^n (\zeta_i + \zeta_i^*) \quad (9)$$

约束条件为:

$$\begin{cases} y_i - w \cdot x - b \leq \varepsilon + \zeta_i \\ w \cdot x + b - y_i \leq \varepsilon + \zeta_i^* \\ \zeta_i \geq 0, \zeta_i^* \geq 0 \end{cases} \quad i=1,2,\dots,n \quad (10)$$

式中,常数 C 控制对误差的补偿程度。其对偶问题可采用相同方法得到:

$$\begin{cases} \sum_{i=1}^n (a_i - a_i^*) = 0 \\ 0 \leq a_i, a_i^* \leq C \end{cases} \quad i=1,2,\dots,n \quad (11)$$

对拉格朗日因子 a_i, a_i^* 最大化目标函数:

$$W(a, a^*) = -\varepsilon \sum_{i=1}^n (a_i^* + a_i) + \sum_{i=1}^n y_i (a_i^* + a_i) -$$

$$-\frac{1}{2} \sum_{i,j=1}^n (a_i^* - a_i)(a_j^* - a_j) \kappa_i \cdot x_j) \quad (12)$$

得到最终回归函数:

$$f(x) = w \cdot \phi(x) + b = \sum_{i=1}^n (a_i^* - a_i) K(x_i \cdot x) + b^* \quad (13)$$

3 KPCA 数据降维

3.1 样本数据的选取及预处理

单日的电力负荷最大值以安徽某市 2018 年 7 月 12 日至 2018 年 9 月 30 日,单日的负荷最大值作为基础数据,期间一共 81 天,所以可得到 81 组原始数据。

预测每天的电力负荷最大值属于电力负荷短期预测,影响因素方面选取了 10 个,分别为:前一天的负荷最大值、当天 8 点的温度、当天的最高温度、当天的最低温度、前一天的平均温度、前一天的最高温度、前一天的最低温度、前两天的负荷最大值、当天的天气类型、当天的日期类型。未考虑季节需求差异、人口规模变化、空气质量状况、实时电价、特殊事件等因素。

电力负荷最大值的数据取自当地的电力公司,气温数据来自当地的气象局。以电力负荷最大值的 10 个影响因素为自变量,以当天的电力负荷最大值为因变量。并对自变量及因变量分别进行归一化预处理,采用如下归一化映射:

$$f: x \rightarrow y = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (14)$$

$x, y \in R^n; x_{\min} = \min(x); x_{\max} = \max(x)$ 。归一化之后原始数据被规整到 $[0,1]$ 范围内,即 $y_i \in [0,1], i=1,2,\dots,n$ 。

3.2 KPCA 自变量降维处理

KPCA 进行自变量的降维,在 MATLAB 上编程实现,下面对其关键步骤部分进行说明。

(1) 数据预处理,进行标准化处理,处理后均值为 0,方差为 1。

(2) 采用高斯径向基核函数,求核矩阵。高斯径向基核函数:

$$K(\|x-x_c\|) = \exp\left(-\frac{\|x-x_c\|^2}{(2*\sigma)^2}\right) \quad (15)$$

其中, x 为空间中任一点, x_c 为核函数中心, σ 为函数的宽度参数, 控制函数的径向作用范围。

(3) 中心化矩阵, 得到一个 81×81 的矩阵。

(4) 特征值分解。求协方差矩阵的特征向量和特征值, 将主对角线上为特征值的对角阵变换成特征值列向量并对特征值按降序排列。

(5) 确定主元贡献率。计算各特征值的贡献率以及累计贡献率, 并记录主元所在特征值向量中的序号, 同时保存主元序号以及主元个数。

(6) 计算主元向量以及主元所对应的特征向量(负荷向量), 并构建主元对角阵。

经过 KPCA 降维处理之后, 电力负荷最大值的影响因素由 10 维降为了 5 维, 其累计贡献率可达到 93.70%。KPCA 前 5 个主成分, 每个的贡献率及累计贡献率如表 1 所示。

表 1 KPCA 对 10 种影响因素的处理结果
Tab.1 KPCA processing results for 10 influencing factors

主成分	单个贡献率	累计贡献率
1	53.74%	53.74%
2	12.72%	66.46%
3	12.21%	78.67%
4	10.55%	89.22%
5	4.48%	93.70%

4 SVM 预测模型

以 KPCA 自变量降维处理后的 5 个主成分为输入, 以当天电力负荷最大值为输出, 采用径向基核函数, 进行 SVM 预测模型的训练与测试, 并从以上 81 组数据中, 随机抽取 54 组作为训练集, 27 组作为测试集。算法流程如图 1 所示。



图 1 SVM 预测模型整体流程

Fig. 1 The overall process of SVM prediction model

4.1 最佳参数的选择

SVM 回归预测分析最重要的参数有 2 个, 分别是惩罚参数 c 与径向基核函数参数中的 g 。 c 表示对误差的宽容度, 用于对分类间隔及错分样本进行平衡。 c 越高, 对出现的误差越不能容忍, 容易产生过拟合; c 越小, 则容易欠拟合^[18]。 c 过大或过小, 都会使模型泛化性能变差, 因此需要找到一个最佳的 c 。

径向基核函数是一个以向量为自变量的函数, 可以基于向量距离运算输出一个标量。 g 是径向基核函数自带的一个参数, 隐含地决定了数据映射到新的特征空间后的分布。其值越大, 支持向量越少, 其值越小, 支持向量越多^[19]。支持向量的个数对训练与预测的速度存在影响。

关于 SVM 回归预测分析最佳 c 、 g 的优化选取, 目前还没有公认统一的最好方法, 比较常用的方法就是让 c 和 g 在一定的范围内取值, 对于取定的 c 和 g , 把训练集作为原始数据集并利用 K-CV (K-fold Cross Validation) 方法在此组 c 和 g 下验证预测准确率, 最终取使得预测准确率最高的那组 c 和 g 作为最佳的参数。

首先进行粗略的寻找, 观察粗略寻找的结果后再进行精细选择。图 2、图 3、图 4、图 5 分别是 c 、 g 粗略寻找结果等高线图, 粗略寻找结果 3D 视图、精细选择结果等高线图、精细选择结果 3D 视图。粗略寻找的 c 为 1.74, g 为 1; 精细选择的 c 为 1, g 为 2.83。

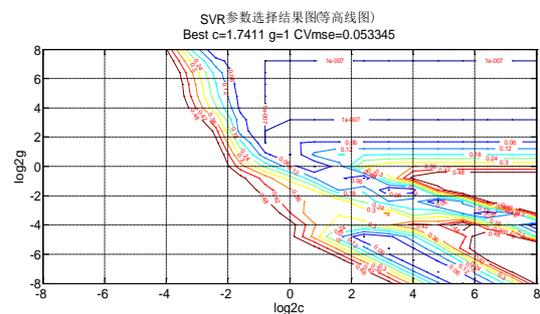


图 2 粗略寻找结果等高线图

Fig. 2 Contour map of the rough sought results

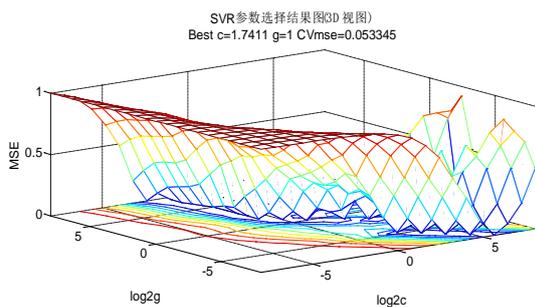


图 3 粗略寻找结果 3D 视图
Fig. 3 3D view of the rough sought results

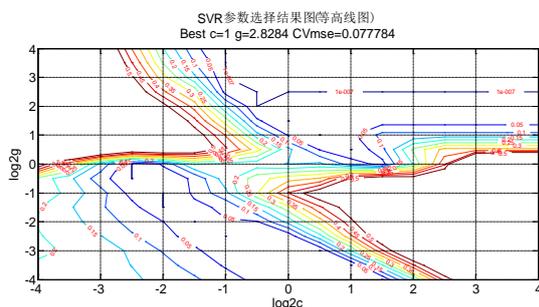


图 4 精细选择结果等高线图
Fig. 4 Contour map of the fine selection results

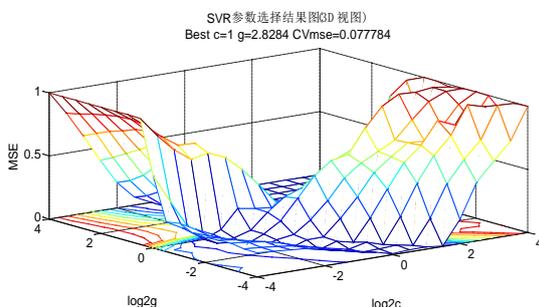


图 5 精细选择结果 3D 视图
Fig. 5 3D view of the fine selection results

4.2 SVM 训练与回归预测

从 81 天的电力负荷最大值数据集中，随机选取 54 天的数据，利用上面得到的最佳参数 c 和 g 对 SVM 模型进行训练，然后再对随机选取的 27 天的数据进行回归预测。

最终预测出 27 天的电力负荷最大值，并与其实际负荷最大值进行比较，仿真结果如图 6 所示。

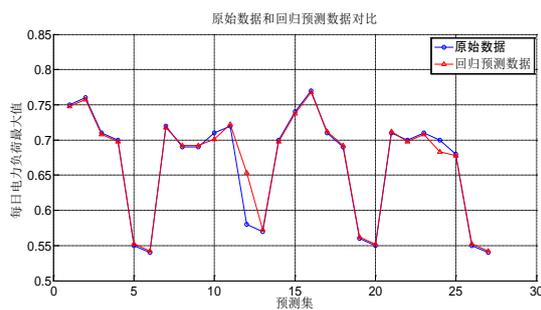


图 6 SVM 回归预测结果图
Fig. 6 Result figure of SVM regression prediction

4.3 预测结果分析

预测误差为预测值与实际值之差，相对预测误差为预测误差与实际值的比值。预测误差图与相对预测误差图分别如图 7、图 8 所示。由图可知预测的第 10、12、24 个的误差相对较大，其中第 12 个的误差最大，其他的预测结果基本准确，总体上具有较高的预测精度。最后，求得预测结果的均方误差 $MSE=0.0041$ ，相关系数 $R=0.9631$ 。

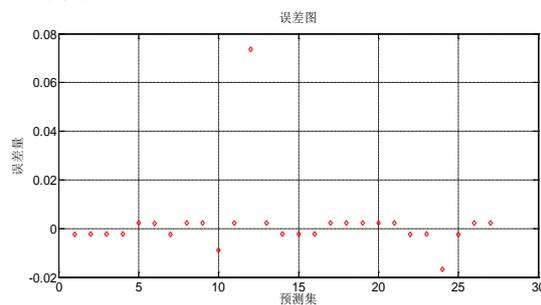


图 7 SVM 预测的误差图
Fig. 7 Error figure of SVM prediction

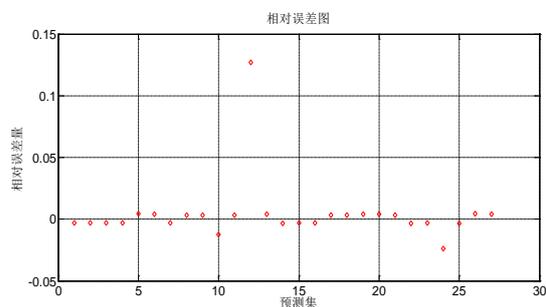


图 8 SVM 预测的相对误差图
Fig. 8 Relative error figure of SVM prediction

5 结论

对电力负荷最大值短期预测的准确度进行有效地提升，先对 10 个最相关的影响因素进行

了 KPCA 降维处理,然后构建了 SVM 预测模型,采用了交叉验证法选择最佳回归参数并利用此最佳参数训练 SVM 预测模型,最后是预测结果的拟合,其均方误差达到了 0.0041,相关系数达到了 0.9631。仿真结果表明,基于 KPCA-SVM 电力负荷最大值预测模型具有很好的预测效果。

参 考 文 献

- [1] 梁青艳,孙彦广. 钢铁企业电力负荷动态预测建模问题的应用研究[J]. 科学技术与工程, 2018, 18(16): 44-54.
- [2] 周博文,陈麒宇,杨东升. 巴西大停电的思考[J]. 发电技术, 2018, 39(2): 97-105.
- [3] 肖白,赵殿平,姜卓,等. 城市配电网供电能力评估综述[J]. 发电技术, 2018, 39(3): 213-219.
- [4] 陈晓静. 电力负荷预测与调度策略研究[J]. 通信电源技术, 2018, 35(8): 45-46.
- [5] 杨博宇,陈仕军. 电力负荷预测研究综述及预测分析[J]. 四川电力技术, 2018, 41(3): 56-60.
- [6] 郭威,巴秀玲. 基于神经网络的电力系统负荷预测问题研究[J]. 自动化与仪器仪表, 2017(10): 192-194.
- [7] 肖白,周潮,穆钢. 空间电力负荷预测方法综述与展望[J]. 中国电机工程学报, 2013, 33(25): 78-79.
- [8] 邹京希,曹敏. 基于总体测辨和人工神经网络的负荷建模及预测方法[J]. 电力系统及其自动化学报, 2018, 30(8): 108-112.
- [9] 周涛. 基于改进神经网络的电力系统中长期负荷预测研究[J]. 电气应用, 2013, 32(4): 26-29.
- [10] 杨楠,叶迪,周峥. 基于 NACEMD-Elman 神经网络的风功率组合预测[J]. 水电能源科学, 2018, 36(9): 209-211.
- [11] 胡朝举,李云霞. 关于电力负荷影响因素及智能预测方法的研究[J]. 通信电源技术, 2018, 35(2): 231-232.
- [12] 许杏花,潘庭龙. 基于 KPCA-RF 的风电场功率预测方法研究[J]. 可再生能源, 2018, 36(9): 1323-1327.
- [13] 刘剑,刘丽华. 基于 KPCA 与 SVM 的混合核交流数据检测[J]. 沈阳建筑大学学报(自然科学版), 2018, 34(5): 921-928.
- [14] 李军,王秋莉. KPCA-KELM 在短期交通流量预测中的应用[J]. 江苏大学学报(自然科学版), 2018, 39(5): 570-575.
- [15] 胡伟,张玮灵,闵勇. 基于支持向量机的电力系统紧急控制实时决策方法[J]. 中国电机工程学报, 2017, 37(16): 4567-4576.
- [16] 简献忠,顾祎婷. 一种基于全过程优化支持向量机的短期电力负荷预测方法[J]. 电力科学与工程, 2018, 34(11): 45-51.
- [17] 鞠建波,胡胜林,祝超. 一种改进的支持向量机回归故障预测方法[J]. 电光与控制, 2018, 25(1): 6-9.
- [18] 吐松江·卡日,高文胜,张紫薇. 基于支持向量机和遗传算法的变压器故障诊断[J]. 清华大学学报(自然科学版), 2018, 58(7): 623-629.
- [19] 许洪华,李勇,施恂山. 基于振动噪声 SC 与 SVM 的变压器局部过热故障诊断[J]. 陕西电力, 2017, 45(6): 11-17.