

基于支持向量机的城市小区燃气管网日负荷预测模型

张建秋

(中石油昆仑燃气有限公司吉林分公司 吉林省吉林市 132013)

摘 要:该文使用支持向量机中的两种核函数,采用grid-search算法、遗传算法、粒子群算法优化参数,建立对吉林市某小区燃气管网日负荷预测的支持向量机模型。将日最高温度、日最低温度、日平均温度、小区人员最高年龄、小区人员最低年龄、小区人员平均年龄作为燃气管网日负荷变化密切相关的主要影响因素,分别作为支持向量机的输入量,将小区人员临时出差、小区临时增加暂住人口等随机因素作为燃气管网日负荷变化密切相关的次要影响因素,将随机因素统一归为支持向量机的一个输入量。采用[0,1]归一化方法,对作为影响因素的输入量数据与日负荷预测输出量数据进行归一化处理。对节假日和工作日的燃气管网日负荷预测采用独立处理方法,避免了相互之间的干扰影响。试验结果表明,采用径向基核函数的支持向量机预测模型对燃气管网日负荷预测拟合程度达到90%以上。

关键词:燃气管网 支持向量机 日负荷预测

中图分类号:TP391.4

文献标识码:A

文章编号:1674-098X(2014)04(c)-0099-04

Daily Load Prediction Model for Community Gas Network Based on Support Vector Machine

Zhang jian qiu

China oil KunLun Gas Co., Ltd. Jilin branch, Jilin City 132013, China

Abstract: This paper constructs daily load prediction model for community gas network based support vector machine in Jilin, which is used of two kernel function and grid-search algorithm, genetic algorithm and particle swarm optimization algorithm. Inputs of support vector machine include daily maximum temperature, daily minimum temperature, daily mean temperature, maximum age of staff, minimum age of staff, average age of staff, which are the main factors affecting the daily load of gas network. Secondary factors include some random factors such as temporary travel of staff, temporary increase of staff and so on. The paper constructs random factors as an input of support vector machine. Output of support vector machine is daily load prediction of community gas network. Normalization method of [0,1] is used to normalize inputs and output of support vector machine. The independent processing method on holidays and days of gas network daily load prediction avoids mutual interference effects. The result shows the fitting degree is more than 90% on prediction models which are constructed of support vector machine based on radial basis function kernel function.

Key words: Gas network Support vector machine Daily load prediction

城市小区燃气管网的日负荷指标是城市燃气的基础工作,对分析燃气管网的年负荷、月负荷具有重要指导意义,并且燃气经营企业对于燃气管网的日负荷及预测情况日益重视^[1]。燃气管网日负荷数据的变化规律十分复杂,主要同天气、气温、终端用户人口数量等因素密切相关,并且国家法定节假日的日负荷变化与工作日亦不相同。传统的燃气管网日负荷预测方法主要有回归综合位移平均模型、多元主要有线性回归法、三角

函数模型、BP神经网络法等^[2],试验结果表明这些预测方法具有一定的精度^[3-5]。但上述方法所需样本数据较多,并且样本数据越多,上述方法得到的预测模型泛化能力越强,可信度越高。

支持向量机(SVM)是机器学习研究重大成果,其具有泛化能力强、全局寻优的特点,尤其具有所需样本数据少的特点^[6]。SVM利用松弛变量和核函数^[7],针对样本数据线性不可分的情况,在高维空间寻找

其最优分类面。

该文将吉林市某小区燃气管网的日负荷作为研究对象,将燃气管网日负荷变化密切相关的主要影响因素——日最高温度、日最低温度、日平均温度、小区人员最高年龄、小区人员最低年龄、小区人员平均年龄,作为SVM的6个输入量,将燃气管网日负荷变化密切相关的次要影响因素——小区人员临时出差、小区临时增加暂住人口等随机因素,统一作为SVM的1个输入量,利用不同核函数、不同参数的支持向量机建立吉林市某小区燃气管网日负荷预测模型,通过该预测模型得到该小区的燃气管网日负荷预测值。

1 支持向量机

1.1 支持向量机原理

设 $f(x)$ 表示线性函数,其表达式为:

$$f(x) = \langle w, x \rangle + b = \sum_{i=1}^n w_i x_i + b \quad (1.1)$$

其中: b 为偏置, w 为可调权值向量,决策规则: $\text{sgn}(0) = 1$, 将 $f(x)$ 称为超平面(Hyper Plane)。由式 $f(x) = 0$ 定义超平面 L , L 将输入空间 X 划分成两部分,如图1所示。

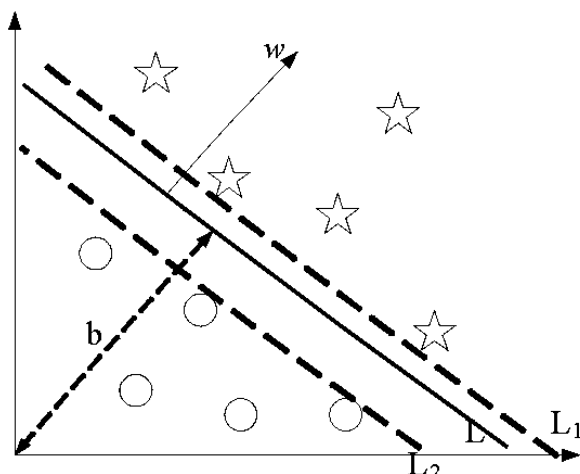


图1 二维训练集的一个分类超平面

显然,有许多分类超平面可以将图1中的“圆圈”和“五角星”两类点正确区分开。

取训练样本 $\{x_i, y_i\}_{i=1}^l$, 其中: x_i 为第 i 个样本, $y_i \in \{-1, 1\}$ 。对于线性可分训练集存在超平面 $L: \langle w, x \rangle + b = 0$, 即式1.2和式1.3。

$$\langle w, x \rangle + b \geq 1, y_i = 1 \quad (1.2)$$

$$\langle w, x \rangle + b \leq -1, y_i = -1 \quad (1.3)$$

定义超平面 $L_1: \langle w, x \rangle + b = 1, y_i = 1$, 设 x_1 为超平面 L_1 上的一点; 超平面 $L_2: \langle w, x \rangle + b = -1, y_i = -1$, x_2 为超平面 L_2 上的一点。 x_1 和 x_2 满足式1.4和式1.5。

$$\langle w, x_1 \rangle + b_1 = 1 \quad (1.4)$$

$$\langle w, x_2 \rangle + b_2 = -1 \quad (1.5)$$

超平面 L_1 与超平面 L_2 之间间隔 dis 为式1.6。

$$dis = \frac{w}{\|w\|} (x_1 - x_2) \quad (1.6)$$

寻找超平面 L_1 和 L_2 间隔最大化的平面——最优超平面, 即为二次规划问题, 如式1.7和式1.8所示。

$$\min_w \frac{\|w\|^2}{2} \quad (1.7)$$

$$s.t. y_i (\langle w, x_i \rangle + b) \geq 1, \forall i = 1, 2, \dots, N \quad (1.8)$$

利用lagrange乘子法解决上述二次规划问题, 建立lagrange函数如式1.9所示。

$$J(w, b, a) = \frac{1}{2} w^T w - \sum_{i=1}^l a_i [y_i (\langle w, x_i \rangle + b) - 1] \quad (1.9)$$

其中 a_i 为lagrange乘子。

将lagrange函数对 w, b 求其最小值, 对 a_i 求其最大值, 解 w_0, b_0 和 a_0 在函数的鞍点上满足式1.10和式1.11。

$$\sum_{i=1}^l a_{0i} y_i = 0, a_{0i} \geq 0, i = 1, 2, \dots, l \quad (1.10)$$

$$w_0 = \sum_{i=1}^l y_i a_{0i} x_i, a_{0i} \geq 0, i = 1, 2, \dots, l \quad (1.11)$$

将式1.10代入lagrange函数式1.9, 利用式1.11可得原优化问题的对偶问题, 可构造出最优超平面, 如式1.12、式1.13和式1.14所示。

$$\max_a - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j a_i a_j \langle x_i, x_j \rangle + \sum_{j=1}^l a_j \quad (1.12)$$

$$s.t. \sum_{i=1}^l y_i a_i = 0 \quad (1.13)$$

$$a_i \geq 0, i = 1, 2, \dots, l \quad (1.14)$$

基于最优超平面的分类规则如式1.15所示。

$$f(x) = \text{sgn} \left(\sum_{i=1}^l y_i a_i^* \langle x, x_i \rangle - b_0 \right) \quad (1.15)$$

1.2 核函数

对于二维空间的线性不可分数据群, 为构造最优超平面, 其函数表达式如式1.16所示。

$$g(x) = c_0 + c_1 x + c_2 x^2 \quad (1.16)$$

构造向量 y 和 a , 如式1.17和式1.18所示。

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix} \quad (1.17)$$

$$a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} \quad (1.18)$$

$g(x)$ 可转化为 $f(y) = \langle a, y \rangle$, 即式1.19。

$$g(x) = f(y) = \langle a, y \rangle = a^T y \quad (1.19)$$

可见, 映射到四维空间后, 原来在二维空间中一个线性不可分的问题变成了线性可分, 构造函数 $K(w, x)$, 其满足式1.20和式1.21。

$$g(x) = K(w, x) + b \quad (1.20)$$

表1 工作日不同参数组合下的预测集相关系数R

	grid-search	GA	PSO
polynomial	80.139%	83.926%	83.893%
RBF	89.865%	90.628%	90.781%

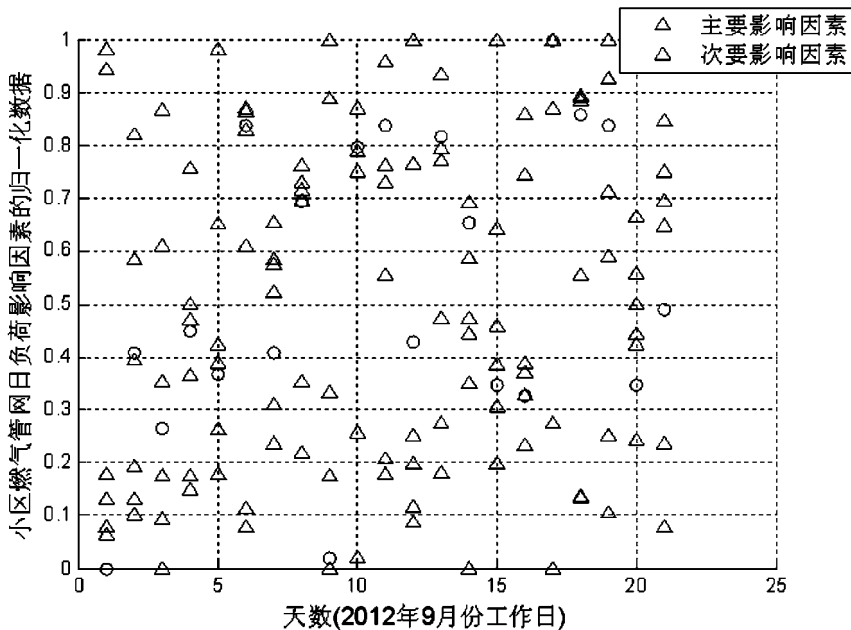


图2 工作日小区燃气管网日负荷影响因素数据的归一化

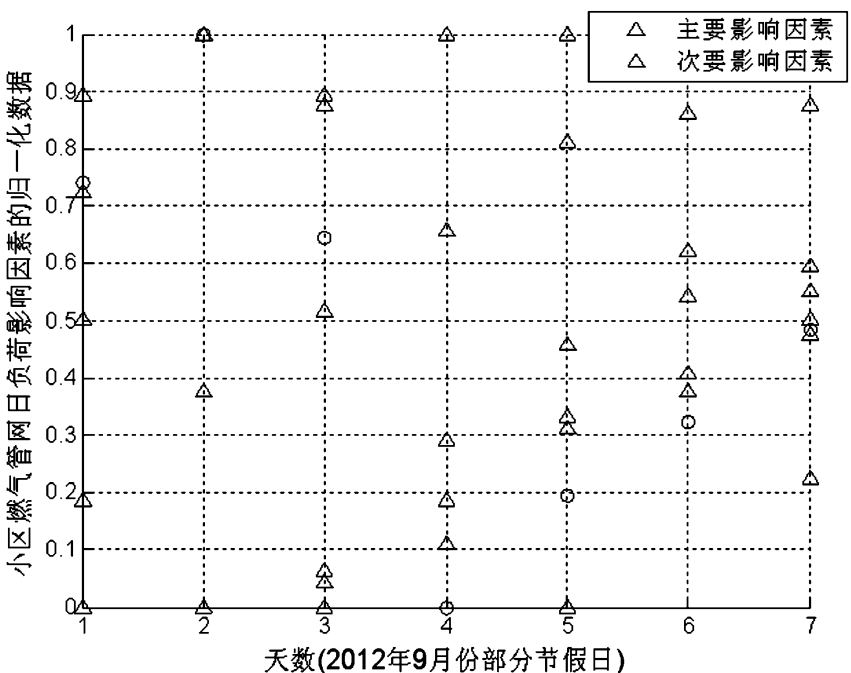


图3 节假日小区燃气管网日负荷影响因素数据的归一化

$$f(x') = \langle w', x' \rangle + b \quad (1.21)$$

其中, w 为低维空间多维常量, x 为低维空间多维变量, w' 为由 w 映射到高维空间的对应常量, x' 为由 x 映射到高维空间的对应向量。

若只需将低维空间的输入代入式1.20, 计算即可得到高维空间的线性分类函数 $f(x')$, 则函数 $K(w, x)$ 称为核函数, 能使得上述式1.20和式1.21中的两个函数计算结

果完全一致。

2 燃气管网日负荷的影响因素及数据处理

本文以吉林市某小区燃气管网日负荷为研究对象, 时间为2012年9月份非节假日和节假日, 非节假日的时间跨度为2012年9月份工作日, 即9月3日至9月7日、9月10日至9月14日、9月17日至9月21日、9月24日至9月

29日, 节假日的时间跨度为2012年9月份中的部分法定节假日, 取9月8日、9日、15日、16日、22日、23日、30日。将日最高温度、日最低温度、日平均温度、小区人员最高年龄、小区人员最低年龄、小区人员平均年龄作为影响小区燃气管网日负荷的主要因素, 分别作为SVM的6个输入量。将小区人员临时出差、小区临时增加暂住人口等随机因素作为影响小区燃气管网日负荷的次要因素, 采用随机变量法统一作为SVM的1个输入量。将小区燃气管网日负荷实际消耗量作为SVM的输出量。由于温度、人员数量、人员年龄等数据单位不同, 会引起数据数量级差异, 本文对日最高温度、日最低温度、日平均温度、小区人员最高年龄、小区人员最低年龄、小区人员平均年龄、随机因素、日负荷实际消耗量的数据进行[0,1]归一化, 使所测数据归为一个数量级。工作日和节假日中小区燃气管网日负荷影响因素的归一化数据, 分别如图2和图3所示。

表2 节假日不同参数组合下的预测集相关系数R

	grid-search	GA	PSO
polynomial	92.003%	91.491%	92.538%
RBF	88.538%	88.839%	88.072%

原始数据和回归预测数据对比

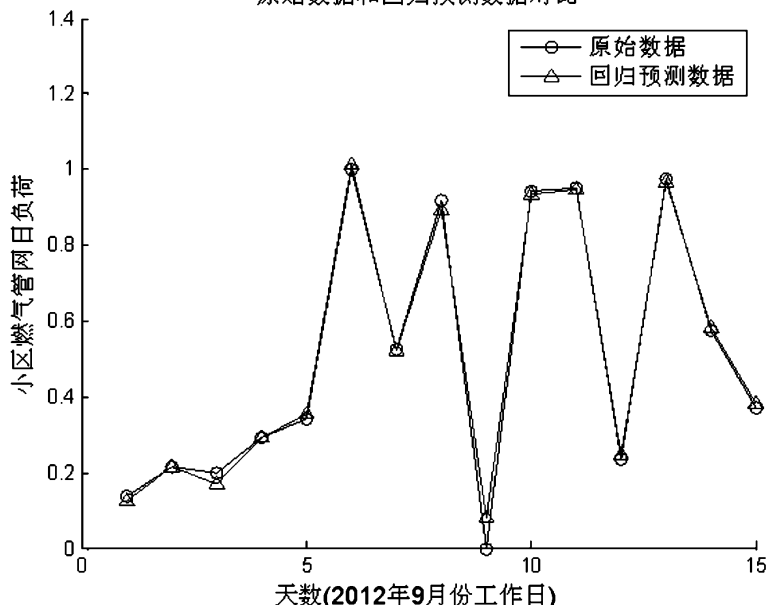


图4 工作日训练集的小区燃气管网日负荷回归预测曲线

原始数据和回归预测数据对比

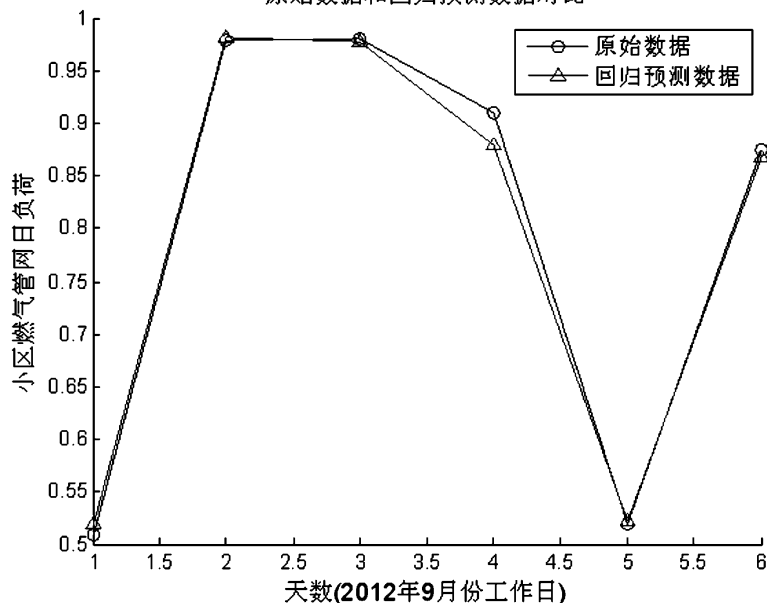


图5 工作日预测集的小区燃气管网日负荷回归预测曲线

3 建立小区燃气管网日负荷的预测模型

3.1 支持向量机建模

本文采用多项式核函数 (polynomial 核函数), 径向基核函数 (radial basis function 核函数), 惩罚参数 c 和 γ 值采用 grid-search 算法、遗传算法 (GA)、粒子群算法 (PSO) 进行参数寻优。对以上组合所建立的不同支持向量模型进行交叉试验, 寻求最优预测模型。利用该模型对工作日和节假日的小区燃气管网日负荷进行预测。

3.2 试验结果分析

寻求工作日小区燃气管网日负荷最优预测模型阶段, 训练集时间跨度为9月3日至9月7日、9月10日至9月14日、9月17日至9月21日, 预测集时间跨度为9月24日至9月29日; 寻求节假日小区燃气管网日负荷最优预测模型阶段, 训练集时间跨度为9月8日、9日、15日、16日、22日, 预测集时间跨度为9月23日、30日。训练集和预测集的数据为日最高温度、日最低温度、日平均温度、小区人员最高年龄、小区人员最低年龄、小区人员平均年龄、随机因素、日负荷实际值。利用支持向量机中 polynomial 核函数和 radial basis function 核函数 (RBF)、优化方法采用 grid-search 算法、遗传算法 (GA)、粒子群算法 (PSO)。对工作日、节假日小区燃气管网日负荷产生的预测效果分别见表1和表2。

由表1可见, 采用 radial basis function 核函数和 PSO 优化算法的模型, 可以实现工作日小区燃气管网日负荷预测集的拟合程度最高, 达到了90.781%。预测结果如图4、图5所示。

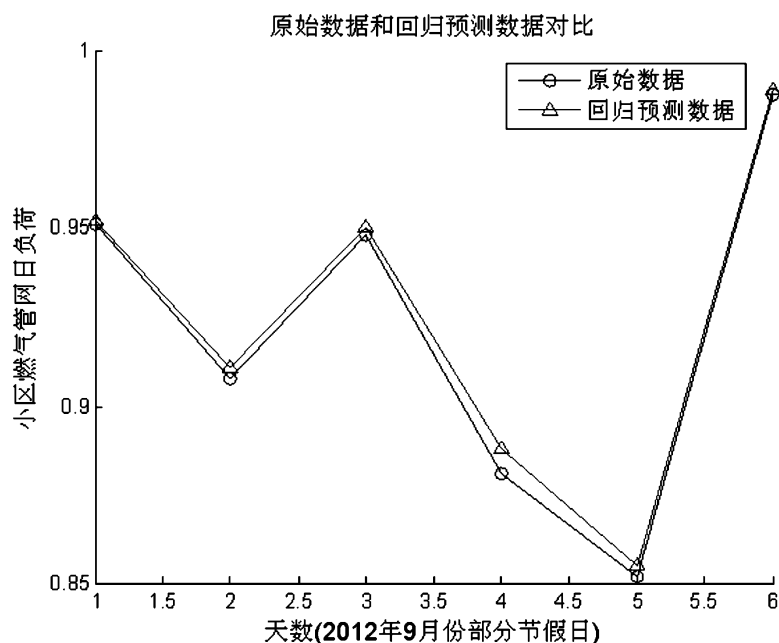


图6 节假日训练集的小区燃气管网日负荷回归预测曲线

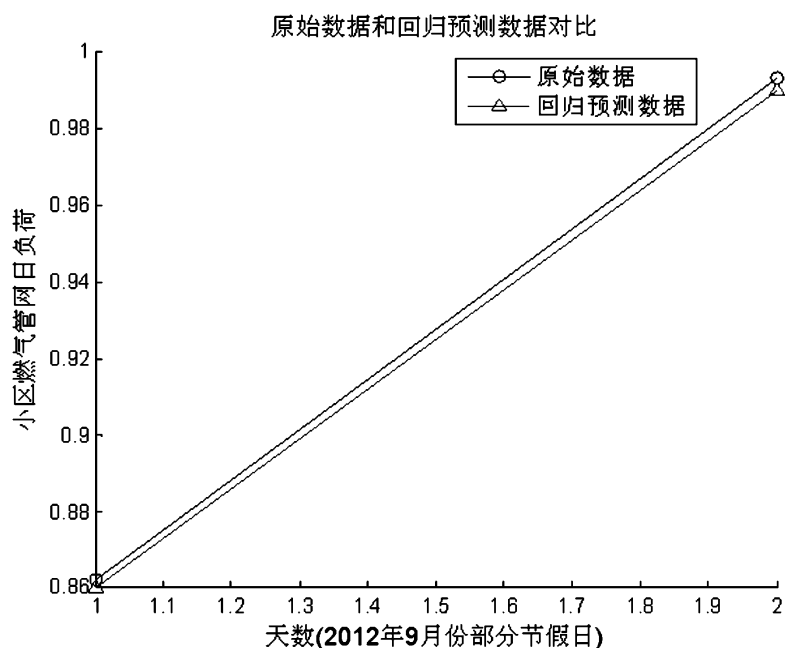


图7 节假日预测集的小区燃气管网日负荷回归预测曲线

由表2可见,采用polynomial核函数和PSO优化算法的模型,可以实现节假日小区燃气管网日负荷预测集的拟合程度最高,达到了92.538%。预测结果如图6、图7所示。

可见,支持向量机采用radial basis function核函数和PSO优化算法时,对工作日小区燃气管网日负荷预测集的拟合程度最高;采用polynomial核函数和PSO优化算法时,对节假日小区燃气管网日负荷预测集的拟合程度最高,均高于90%。

4 结语

该文采用多项式核函数、径向基核函数,利用支持向量机对吉林市某小区工作

日、节假日的燃气管网日负荷建立预测模型,模型参数采用grid-search算法、遗传算法、粒子群算法进行优化,将日最高温度、日最低温度、日平均温度、小区人员最高年龄、小区人员最低年龄、小区人员平均年龄作为影响小区燃气管网日负荷的主要因素,将小区人员临时出差、小区临时增加暂住人口等随机因素作为影响小区燃气管网日负荷的次要因素,主要因素与次要因素统一作为SVM的输入量,日负荷预测值作为SVM的输出量。试验结果表明,采用支持向量机建立的城市小区燃气管网日负荷预测模型,其预测精度达到90%以上,为城市小区燃气管网年负荷、月负荷的进一步分析提供了理论依据、技术支持和试验

方法。

参考文献

- [1] 谭羽非.城市燃气管网日负荷预测的灰色神经网络模型[J].哈尔滨工业大学学报,2003,35(6):679-682.
- [2] 朱刚.城市燃气管网负荷预测的研究[D].天津大学,2009.
- [3] Potocnik, Primoz, Govekar. Short-term natural gas consumption forecasting[J]. Proceeding of the International Conference on Applied Simulation and Modelling,2007(5):353-357.
- [4] 杜元顺.煤气日负荷用的回归分析方法[J].煤气与热力,1982(4):26-28.
- [5] 田一梅,赵元,赵新华.城市煤气负荷预测[J].煤气与热力,1998,18(14):20-23.
- [6] 武海巍,于海业,张蕾.光合有效辐射预测模型的核函数组合优化[J].农业机械学报,2011,42(6):167-173.
- [7] 武海巍.核函数与仿生智能算法在林下参光环境评价系统中的研究[D].吉林大学,2012.