

电网自然区域负荷预测研究

陈志强¹, 牛东晓¹, 李春祥²

(1. 华北电力大学工商管理学院, 河北 保定 071003; 2. 华北电力大学信息与网络管理中心, 河北 保定 071003)

摘要: 农网改造中变电站的规划是一项核心工作, 其中对变电站所带区域的负荷, 称为自然区域负荷, 进行准确的预测又是关键。分析了典型的农村用电量结构, 将之分为四类, 并对其中三类主要用电量进行单独分析和预测。首先利用粗糙集理论对每类用电量的影响因素集合进行约简, 之后利用支持向量回归机对其进行分别预测。最后的预测结果是三个预测结果相加再按比例加上其它用电量。实验结果表明粗糙集结合支持向量回归机的方法比传统方法有比较大的优势, 预测精度比较好。

关键词: 电网自然区域; 负荷预测; 粗糙集; 支持向量回归机; 分类预测

Grid natural area load forecasting research

CHEN Zhi-qiang¹, NIU Dong-xiao¹, LI Chun-xiang²

(1. College of Business Administration, North China Electric Power University, Baoding 071003, China; 2. Information and Network Management Center, North China Electric Power University, Baoding 071003, China)

Abstract: Layout of transformer substations is a central job in rebuilding rural grid, in which accurate load forecasting for the area supplied by transformer substations, called natural area, is a key element. Typical composition of electric power consumed in countryside, which is divided into four sub-classes, is analyzed. Among the four sub-classes three main sub-classes are analyzed and forecasted separately. At first, influence factors set for each sub-class is reduced based upon rough set theory, then forecasted by support vector regression for every sub-class. The final forecasting result is three individual forecasts plus one additional forecast pro rata. Experiment shows that rough set combined with SVM regression is superior to traditional methods and has a good prediction precision.

This project is supported by National Natural Science Foundation of China (No.70671039).

Key words: grid natural area; load forecasting; rough set; SVM regression; classified forecasting

中图分类号: TM715 文献标识码: A 文章编号: 1674-3415(2009)08-0019-06

0 引言

农网改造是我国电力系统建设的一项重要内容。在农村电网规划中, 合理地安排和设置变电站是最重要的工作, 既要做到不过度浪费又能满足今后发展的用电需要。一般地一个变电站所带负荷区域是根据地理位置因地制宜地进行规划, 称为自然区域。一个自然区域可能跨越多个乡镇, 这些乡镇的人口、自然条件、地理环境、产业结构等可能各不相同。要做好农村电网规划, 对自然区域内的用电量和负荷进行准确合理的预测具有十分重要的意义。

空间负荷预测理论^[1]在城市电网规划中得到广泛应用, 这些理论对自然区域负荷预测有一定的借

鉴意义, 但两者也存在比较大的差别。自然区域的负荷密度一般较小且分布十分不均匀, 用电量和负荷受到自然条件影响较大, 按负荷密度来估算区域负荷难以得到满意的结果。由于电网负荷能力和电价等方面的原因, 某个自然区域的历史负荷往往不能准确地反映其真实的电力需要, 因此使用时间序列分析法^[2,3]、趋势外推法等方法估计自然区域的负荷也难以取得较好的效果。

近年来, 粗糙集理论^[4,5]成为数据挖掘研究的新热点, 它能有效地分析和处理不精确、不完整、不一致的数据, 可以发现数据间隐藏的关系, 提取有用信息, 简化信息处理。粗糙集理论的主要思想是在保持分类能力不变的前提下, 通过对知识的约简, 导出概念的分类规则。支持向量机(SVM)是基于Vapnik^[6]等提出的统计学习理论而产生的一种模式识别的新方法, 它基于结构风险最小化原则, 根据

基金项目: 国家自然科学基金项目(70671039)

小样本信息在模型的复杂性和学习能力之间寻求最佳折衷,从而获得最好的泛化能力。支持向量机可以用于分类和回归估计^[7~11],具有以往一些人工智能方法无法比拟的优点^[12, 13]。

农网负荷有着各种复杂的影响因素,将其作为一个整体来研究很难抓住其本质特点。实际上农网用电量的统计口径一般分为九类,对这九类进行单独分析,并分别给出预测结果,最后再将这九类用电量的预测结果进行叠加,得到最后的预测结果。首先利用粗糙集理论对每一类用电量及影响因素进行分析并对影响因素进行约简,然后利用约简后的数据集作为训练样本对SVM回归机进行训练并预测。本文最后对某县的实际用电量进行实验,并给出实验结果。

1 粗糙集理论

在粗糙集中,设一个知识表示系统为 $S = \langle U, A, V, F \rangle$, U 是对象非空有限集合, A 是包含条件属性 C 和决策属性 D 的属性集,其中 $A = C \cup D, C \cap D = \emptyset, a \in A, V \in V_a, V_a$ 是属性 a 的值域, $f: U \times A \rightarrow V$ 是一个信息函数,它为每个对象的每个属性赋予一个信息值。决策表可以看作是定义的一族等价关系,即知识库。用决策表描述论域中的对象,每一行描述一个对象,每一列表征对象的一种属性,属性分为条件属性和结论属性。决策表中不是所有的条件属性都是必要的,一些条件属性是冗余的,当它们被删除时,原来的分类效果并不会改变。

粗糙集利用信息熵概念定义条件属性重要度,并通过属性重要度实现条件属性的约简。在约简前必须对条件属性和决策属性特征化,常见的算法有等距离、等频率、Naive Scaler算法、Semi Naive Scaler算法和离散化法,根据负荷预测的特点本文采用离散化法。为了表述方便,离散化后的知识系统记为 $S = \langle U, C \cup D \rangle$ 。

设 $C_i \in C, U/C_i$ 上的子集组成的 σ 代数上的概率分布为

$$P(X) = \frac{\text{card}(X)}{\text{card}(U)}, X \in U/C_i \quad (1)$$

式中: $\text{card}(X)$ 为集合 X 的元素个数,成为 X 的势。条件属性子集 C_i 的信息熵用如下公式定义:

$$H(C_i) = - \sum_{x \in U/C_i} P(X) \ln(P(X)) \quad (2)$$

条件属性子集 C_i 对于决策属性集 D 的条件信息熵为:

$$H(D|C_i) = - \sum_{x \in U/C_i} P(X) \sum_{y \in U/D} P(Y|X) \ln(P(Y|X)) \quad (3)$$

其中: $P(Y|X) = \frac{\text{card}(Y \cap X)}{\text{card}(X)}$ 。 $H(D|C_i)$ 越大表示决策属性 D 对条件属性子集 C_i 的依赖性越大。设 $C_i = C - \{x_i\}$, 其中 $x_i \in C$, 条件属性值 x_i 对决策属性的影响函数为

$$\lambda(C_i, D) = H(D|C_i) - H(D|C) \quad (4)$$

$\lambda(C_i, D)$ 越大表示 x_i 越重要,如果 $\lambda(C_i, D) = 0$, 则 x_i 是冗余的,应该从条件属性集中去掉,条件属性集就约简为 C_i 。对每个条件属性值进行如上计算,使得每个 x_i 的影响函数都大于0,即可得到最简的条件属性集合。约简后的条件属性集有助于对事物本质有更准确的认识,从而进行更有针对性的研究。

2 支持向量回归机

SVM^[6]回归机处理非线性函数回归问题的主要思想是利用非线性映射 $\Phi(\cdot)$ 将非线性训练数据映射到一个高维空间,从而使非线性函数估计问题转化为高维特征空间中的线性函数估计问题。SVM回归机的问题可表示为:

$$\begin{aligned} \min \{ & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \} (i=1, 2, \dots, l) \\ \text{s.t. } & \begin{cases} y_i - \langle \omega \cdot \Phi(x_i) \rangle - b \leq \varepsilon + \xi_i^* \\ \langle \omega \cdot \Phi(x_i) \rangle + b - y_i \leq \varepsilon + \xi_i \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \quad (5)$$

其中: $C > 0$ 是平衡因子, ε 为损失函数, ξ 和 ξ^* 为松弛变量。根据对偶原理,建立Lagrange方程。

$$\begin{aligned} L(\omega, b, \xi, \alpha) = & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) - \sum_{i=1}^l \alpha_i (\varepsilon + \xi_i + \\ & y_i - \langle \omega \cdot \Phi(x_i) \rangle - b) - \sum_{i=1}^l \alpha_i^* (\varepsilon + \xi_i^* - y_i + \\ & \langle \omega \cdot \Phi(x_i) \rangle + b) - \sum_{i=1}^l (\eta_i \xi_i + \eta_i^* \xi_i^*) \end{aligned} \quad (6)$$

上式对 b, ω, ξ 分别求偏导数并令它们等于零,

可得上述问题的对偶优化问题。即:

$$\begin{aligned} \min \{ & \frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle \Phi(x_i), \Phi(x_j) \rangle + \\ & \sum_{i=1}^l \alpha_i (\varepsilon - y_i) + \sum_{i=1}^l \alpha_i^* (\varepsilon + y_i) \} \\ \text{s.t. } & \begin{cases} \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases} \end{aligned} \quad (7)$$

式(7)是一个二次规划问题, 可得到一组 α_i 和 α_i^* 最优解, 其中 $\alpha_i \neq 0$ 和 $\alpha_i^* \neq 0$ 对应的 x_i 即为支持向量。SVM回归机的回归函数可表示为:

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \langle \Phi(x_i), \Phi(x) \rangle + b = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (8)$$

$K(x_i, x) = \Phi(x_i) \cdot \Phi(x)$ 为核函数。核函数有很多

选择, 本文采用的核函数为高斯函数, 即:

$$K(x_i, x) = \exp\left(-\frac{\|x_i - x\|^2}{2\sigma^2}\right) \quad (9)$$

σ 为核宽度参数。

3 数据分析和预处理

以河北某县的 23 个乡镇用电量为例, 表 1 是 2004 年至 2006 年该县的用电量。

表1 2004年至2006年河北某县23个乡镇用电量 (单位: kWh)

Tab.1 Electric power consumed in 23 towns of some county in Hebei province from 2004 to 2006(Unit: kWh)

年份	农林牧渔	工业	建筑业	交通运输	信息传输	商业	金融 房地产	公共事业	居民生活用 电
2004	178 463 000	162 905 820	1 025 640	1 367 520	0	1 538 460	0	4 102 560	87 561 000
2005	182 644 000	232 616 416	1 270 194	1 661 022	977 072	1 758 730	854 938	5 129 628	92 975 000
2006	244 698 000	170 479 653	896 695	1 076 034	896 695	1 434 712	968 431	3 586 780	137 339 000

由于用电量统计口径的不同, 2004年前后的用电量的统计结果有所差别, 但是其中在农林牧渔、工业和居民生活三类用电量的统计上是一致的。分析表1可以发现, 这三类用电量在其总用电量中占了很大比重, 其中2004年占了98.2%, 2005年97.8%, 2006年98.4%, 因此把研究重点放在这三类用电量上, 而将其它分类用电量合在一起考虑。用 s 表示用电量, $s = s_1 + s_2 + s_3 + s_4$, s_1 = 农林牧渔用电量, s_2 = 工业用电量, s_3 = 居民生活用电量, s_4 = 其它分类用电量。图1是23个乡镇的用电总量和三类主要用电量对比统计图, 分析后者占前者的比例, 其中有22个乡镇的比例高于98%, 这其中有8个乡镇的比例高于99%。从这些数据可以看出, s_4 在整个用电量中所占比例非常小, 可以在其它三类主要用电量预测基础上按比例叠加。

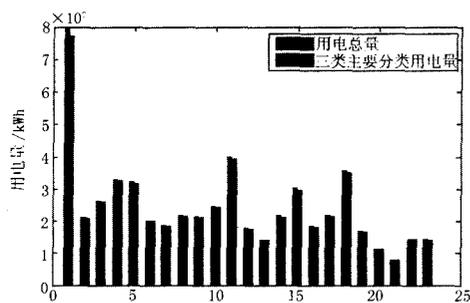


图1 23个乡镇用电总量和三类主要用电量对比图

Fig.1 23-town contrast bar graph for the grass consumption of electric power and three main sub-classes

影响因素用集合 R 表示, $R = \{r_i\}, i = 1, \dots, 6$, 表 2 为各影响因素的含义。结合农村用电负荷的特点, 特别加入了耕地面积和地下水位深度等因素。

表2 各影响因素含义表

Tab.2 Meaning of the influence factors

变量名	影响因素含义
r_1	区域面积
r_2	耕地面积
r_3	人口数
r_4	GDP值
r_5	地下水位深度
r_6	年降雨量

每个影响因素对各用电量分量的影响作用不尽相同, 为了能更准确地把握二者之间的关系, 将每个用电量分量作为独立对象考虑, 分析其与各影响因素之间的关系。最后再将各分量的预测结果进行叠加。下面以农林牧渔用电量 s_1 (以下简称农业用电量) 为例阐述分析方法。

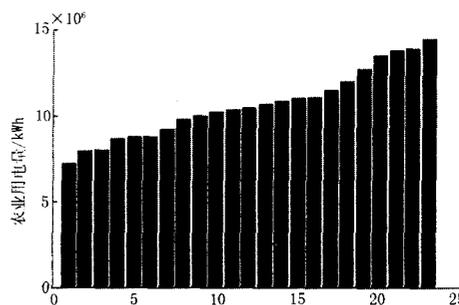


图2 23个乡镇2006年农业用电量

Fig.2 Agricultural power consumption of 23 towns in 2006

首先需要确定农业用电量的分类标准, 图2显示了 23 个乡镇 2006 年农业用电量。用集合

$M = \{m_i | (i=1,2,\dots,23)$ 表示23个乡镇的农业用电量, 将 M 分为三类: M_1 、 M_2 和 M_3

$$M_1 = \{m_i | m_i \geq 11.5 \times 10^6\}$$

$$M_2 = \{m_i | 11.5 \times 10^6 > m_i \geq 9.5 \times 10^6\}$$

$$M_3 = \{m_i | m_i < 9.5 \times 10^6\}$$

第二步采用离散化法将影响因素特征化, 如表3所示。

表3 影响因素离散化表

Tab.3 Discretization table for the influence factors

变量名	离散化值与原值对应关系
r_1 (km ³)	0: (+∞, 60]; 1: (60, 40]; 2: (40, 20]; 3: (20, 0]
r_2 (公顷)	0: (+∞, 3500]; 1: (3500, 2500]; 2: (2500, 1500]; 3: (1500, 0]
r_3 (万人)	0: (+∞, 5]; 1: (5, 4]; 2: (4, 3]; 3: (3, 2]; 4: (2, 0]
r_4 (万元)	0: (+∞, 10]; 1: (10, 5]; 2: (5, 1]; 3: (1, 0]
r_5 (m)	0: (+∞, 20]; 1: (20, 10]; 2: (10, 0]
r_6 (mm)	0: (+∞, 300]; 1: (300, 200]; 2: (200, 100]; 3: (100, 0]

经过离散化的影响因素集合仍用 R 表示。在分析工业用电量 s_2 和居民生活用电量 s_3 时, 影响因素沿用表3的离散化标准, s_2 和 s_3 的分类标准可以根据实际情况进行划分为3~5类。

4 预测模型

4.1 粗糙集约简影响因素

基于上述数据分析, 需要对三个用电量分量分别

进行影响因素约简, 仍以农业用电量为例进行阐述。集合 $M = \{m_i | (i=1,2,\dots,23)$, m_i 为第 i 个乡镇的农业用电量, M 分为三个子集 M_1 、 M_2 和 M_3 , 根据该分类标准对影响因素集合 R 中的元素进行逐个约简, 根据农业用电量的特点, 约简的顺序按照 $\{r_4, r_3, r_1, r_6, r_5, r_2\}$ 进行。按照1节所述的粗糙集条件属性约简办法, 可以得到约简后的影响因素集合 $R_1 = \{r_2, r_5, r_6\}$ 。分析 R_1 可以发现农业用电量的特点, 河北农村主要以农业生产为主, 农业用电主要用于农田抽水灌溉, 因此耕地面积、地下水位深度和降雨量都会对农业用电量产生较大影响。

同样的方法可以得到 s_2 和 s_3 的约简影响因素集合, $R_2 = \{r_1, r_4\}$, $R_3 = \{r_1, r_3, r_4\}$ 。其中居民生活用电量也受到了GDP值的影响, 这主要是因为GDP在一定程度上反映了当地居民的收入水平, 相应地影响到家用电器的保有数量。

4.2 利用SVM回归机进行预测

以农业用电量为例, 图3中三个图分别显示了耕地面积与农业用电量、地下水位与农业用电量、降雨量与农业用电量的对应关系图。从图3中可以看出农业用电量和每个影响因素的关系比较模糊, 很难用简单的模型进行模拟。SVM回归机能够通过非线性变换将问题转换到高维特征空间中, 在高维空间中利用线性方法解决实际中的非线性问题, 并且能够得到有限样本情况下的最优解, 在保证学习精度和模型的泛化能力上取得了较好的平衡。地理位置和人文环境都会对农村用电结构产生比较大的影响, 在进行自然区域预测时应该尽量选择地理位置较近的县市作为参考, 因此可供选择的参考样本有限。在这种情况下, SVM回归机在小样本下的学习能力的优势就会得到体现。

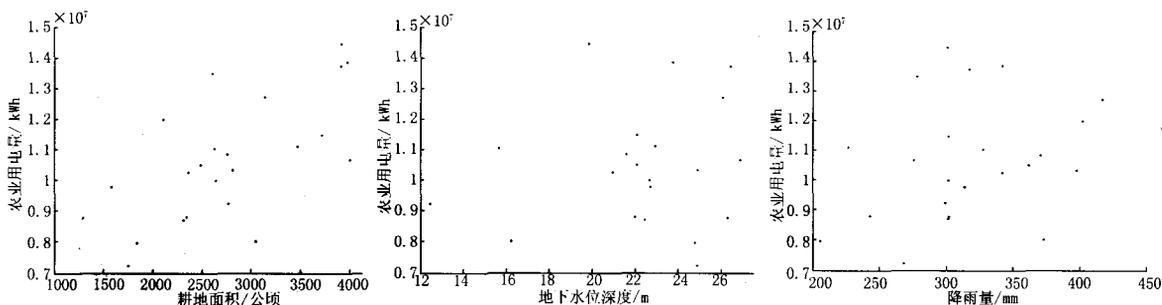


图3 农业用电量和影响因素对应关系散点图

Fig.3 Scattered points figure for the corresponding relationship between agricultural power consumption and influence factors

首先计算农业用电量的预测模型, 根据2节中所述的方法构建SVM回归机, 输入为约简后的影响

因素构成的向量 $X = [r_2, r_3, r_6]$, 输出为农业用电量 s_1 , 核宽度参数 σ 取0.1。利用样本数据训练SVM

回归机, 通过求解式 (8) 的二次规划问题, 得到 SVM 回归机的预测模型。由于三类主要用电量的影响因素各不相同, 因此需要针对每类用电量单独构建预测模型。

5 实验结果

下面对预测结果进行测试, 以河北某县 23 个乡镇的实际数据为实验数据, 选出其中 19 个乡镇作为训练样本, 4 个乡镇作为测试样本。为了验证本文提出的方法的有效性, 将其与多元线性回归方法和神经网络预测模型进行对比试验, 观察其中的差别。

多元线性回归方法是一种被广泛使用的回归模型, 经常被用于多影响因素的回归预测问题。这里仍然按照 3 节的分析结果, 将用电量按照四类进行划分, 而不是把用电量作为一个整体进行预测, 构建模型时影响因素也采用 4.1 节约简后的结果。同样的神经网络预测模型也采用这样的方式进行构建。这里采用最常用的三层前反馈人工神经网络 (BPANN), BPANN 具有较强的学习能力和自适应能力, 能够解决复杂的非线性问题。BPANN 的输入为约简后的影响因素集合, 输出为用电量。

表 4 是 4 个乡镇的 2006 年农业用电量的预测结果及误差分析。

表 4 农业用电量预测结果对比 (单位: kWh)

Tab.4 Forecasting results contrast table for agricultural power consumption (Unit: kWh)

乡镇	实际用电量	线性回归	BPANN	SVM 回归
阳城镇	13 847 264	11 568 971	15 045 897	14 223 789
相对误差		-0.164 5	0.086 6	0.027 2
大庄镇	7 952 432	8 858 738	8 945 213	7 721 438
相对误差		0.114 0	0.124 8	-0.029 0
望亭乡	11 020 608	12 067 894	11 514 797	11 167 435
相对误差		0.095 0	0.044 8	0.013 3
焦庄乡	8 686 637	7 025 136	8 354 786	8 418 792
相对误差		-0.191 3	-0.038 2	-0.030 8

从表 4 中可以看出, 多元线性回归模型预测误差一般在 10% 以上, 存在比较大的误差波动, 这表明农业用电量和影响因素之间的作用关系比较复杂, 用线性关系来模拟存在比较大的误差。BPANN 的预测误差要小于线性回归模型, 但是也存在误差波动比较大的问题, 这主要因为训练样本较少, 造成 BPANN 性能下降。SVM 回归机的预测误差较小, 实验结果中基本都小于 3%, 而且误差波动较小, 具有较好的泛化能力, 这主要得益于 SVM 的结构风险最小化的特性。表 5 和表 6 分别为四个乡镇 2006 年工

业用电量和居民生活用电量的预测结果与误差分析。

表 5 工业用电量预测结果对比 (单位: 千瓦时)

Tab.5 Forecasting results contrast table for industrial power consumption (Unit: kWh)

乡镇	实际用电量	线性回归	BPANN	SVM 回归
阳城镇	3 353 956	3 454 127	3 115 429	3 389 213
相对误差		0.029 9	-0.071 9	0.010 6
大庄镇	6 381 629	6 547 512	6 488 735	6 410 238
相对误差		0.026 0	0.016 8	0.004 5
望亭乡	10 411 146	11 041 235	11 487 132	9 987 689
相对误差		0.060 5	0.103 3	-0.040 7
焦庄乡	2 107 215	2 014 247	2 345 912	2 079 826
相对误差		-0.044 1	0.113 3	-0.013 0

表 6 居民生活用电量预测结果对比 (单位: kWh)

Tab.6 Forecasting results contrast table for residential power consumption (Unit: kWh)

乡镇	实际用电量	线性回归	BPANN	SVM 回归
阳城镇	8 820 466	8 048 790	9 745 123	8 845 621
相对误差		-0.087 5	0.104 8	0.002 9
大庄镇	3 825 918	4 542 378	4 154 326	3 941 236
相对误差		0.187 2	0.085 8	0.030 1
望亭乡	13 531 467	11 945 678	14 234 650	13 174 568
相对误差		-0.117 2	0.052 0	-0.026 4
焦庄乡	3 301 790	3 104 578	3 554 786	3 254 268
相对误差		-0.059 7	0.076 6	-0.014 4

工业用电量的预测结果中, 线性回归模型预测效果要好于 BPANN, 且误差波动也较农业用电量的要小, 说明工业用电量与 GDP 值和区域面积近似属于线性关系。居民生活用电量的预测结果中, 前两个预测模型又出现了比较大的预测误差波动。这两项预测中 SVM 回归机的性能表现还是最好的, 充分体现了 SVM 回归机的优势。

下面将乡镇的用电量作为一个整体来进行预测实验, 并和前面按照分类用电量的预测结果加以比较。其中多元线性回归模型的影响因素只考虑区域面积和人口数, 这主要是考虑到这两个因素在用电量中的重要作用, 同时也能避免维数过多造成的维数灾难。BPANN 和 SVM 回归机将全部影响因素都考虑进去。表 7 是 4 个乡镇用电量预测结果对比。其中 RS-SVM 回归机的预测值为 $s = (s_1 + s_2 + s_3) \times (1 + 2\%)$, $s_i (i = 1, 2, 3)$ 是分类用电量 s_i 的 SVM 回归机的预测值。

表7 乡镇总用电量预测结果对比(单位: kWh)

Tab.7 Forecasting results contrast table for the towns' gross power consumption (Unit: kWh)

乡镇	线性回归	BPANN	SVM回归	RS-SVM回归机
阳城镇	25 097 821	29 451 795	25 134 125	26 987 795
相对误差	-0.041 9	0.124 3	-0.040 5	0.030 2
大庄镇	20 554 873	16 474 321	19 673 256	18 434 370
相对误差	0.111 6	-0.109 1	0.063 9	-0.003 1
望亭乡	39 042 698	32 479 512	33 465 187	35 016 286
相对误差	0.099 7	-0.085 2	-0.057 4	-0.013 7
焦庄乡	15 386 324	14 987 256	15 032 149	14 027 944
相对误差	0.083 2	0.055 1	0.058 2	-0.012 5

(4个乡镇的实际总用电量分别为(单位:kWh):

阳城镇:26 195 982; 大庄镇: 18 491 614; 望亭乡:

35 504 259; 焦庄乡: 14 205 148)

从表7中可以看出, SVM回归机按分类用电量分别预测后得到的预测结果明显要好于将用电量作为整体的预测结果, 这主要得益于粗糙集对各分类用电量影响因素的约简, 避免了无关因素的影响, 而单纯用SVM回归机对总用电量进行预测也难以达到较好的效果。BPANN的预测误差较大, 这主要是因为输入的影响因素比较多而训练数据又比较少, 难以将BPANN训练到一个比较好的性能。线性回归模型的预测误差也比较大, 说明单纯考虑区域面积和人口数的方法存在弊端。

从实验结果分析来看, 粗糙集结合SVM回归机进行分类用电量的预测方法具有比较好的效果, 粗糙集能够在复杂的影响因素集中发现和用电量具有本质联系的影响因素, 排除无关因素的影响, SVM回归机在小样本条件下具有比较好的学习能力和泛化能力, 能够获得比较好的预测精度。

6 结论

本文结合农网改造中的实际问题提出了自然区域负荷预测的概念, 并对其进行了深入的研究。根据典型的农村用电量结构分析, 农业用电量、工业用电量、居民生活用电量这三类用电一般占总量的 98%左右, 因此将这三类用电量作为研究的重点。本文在研究了粗糙集和 SVM 的特点后, 提出了二者相结合的自然区域负荷预测模型, 基本过程概括为“分解-约简-预测-合成”, 首先将用电量分为四类分用电量, 之后利用粗糙集对前三类主要用电量的影响因素集合进行约简, 然后利用 SVM 回归机进行分别预测, 最后将分类用电量预测结果叠加并按 2%比例加上其

它类用电量, 得到最后的预测结果。文中以河北某县的实际数据为例, 对模型进行了详细的测试, 并与线性回归模型和 BPANN 模型进行了比较实验, 结果证明该模型有较好的预测精度和适应能力。

参考文献

- [1] 孙旭, 任震. 空间负荷预测在城市电网规划中的应用[J]. 继电器, 2005, 14 (33): 79-81.
SUN Xu, REN Zhen. Application of Spatial Load Forecasting in Urban Power Network Planning[J]. Relay, 2005, 14 (33): 79-81.
- [2] Box G E P, Jenkins G M. Time Series Analysis, Forecasting and Control[M]. San Francisco:1970.
- [3] 孙德山, 吴今培, 肖健华, 等. SVR 在混沌时间序列预测中的应用[J]. 系统仿真学报, 2004, 16(3): 519-521.
SUN De-shan, WU Jin-pei, XIAO Jian-hua, et al. The Application of SVR to Prediction of Chaotic Time Series[J]. Journal of System Simulation, 2004, 16(3): 519-521.
- [4] 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.
XIE Hong, CHENG Hao-zhong, ZHANG Guo-li, et al. Applying Rough Settheory to Establish Artificial Neural Networks for Short Term Load Forecasting[J]. Proceedings of the CSEE, 2003, 23(11): 1-4.
- [5] 邵能灵, 侯志俭. 小波模糊神经网络在电力系统中短期负荷预测中的应用[J]. 中国电机工程学报, 2004, 24(1): 24-29.
TAI Neng-ling, HOU Zhi-jian. New Short-term Load Forecasting Principle with the Wavelet Transform Fuzzy Neural Network for the Power System[J]. Proceedings of the CSEE, 2004, 24(1): 24-29.
- [6] Vapnik V N. The Nature of Statistical Learning Theory [M]. New York: Springer-Verlag, 1995.
- [7] 牛东晓, 谷志红, 邢 棉, 等. 基于数据挖掘的 SVM 短期负荷预测方法研究[J]. 中国电机工程学报, 2006, 26(18): 6-12.
NIU Dong-xiao, GU Zhi-hong, XING Mian, et al. Study on Forecasting Approach to Short-term Load of SVM Based on Data Mining[J]. Proceedings of the CSEE, 2006, 26(18): 6-12.
- [8] 杨金芳, 翟永杰, 王东风, 等. 基于支持向量回归的时间序列预测[J]. 中国电机工程学报, 2005, 25(17): 110-114.
YANG Jin-fang, ZHAI Yong-jie, WANG Dong-feng, et al. Time Series Prediction Based on Support Vector Regression[J]. Proceedings of the CSEE, 2005, 25(17): 110-114.

(下转第 29 页 continued on page 29)

- [2] 冯志强, 杨宗霄, 程传业, 等. 基于 B/S 模式的火电厂安全评价信息系统[J]. 电网技术, 2007, 31(21): 67-71.
FENG Zhi-qiang, YANG Zong-xiao, CHENG Chuan-ye, et al. Security Assessment Management Information System for Power Plants Based on Browser/Server Mode[J]. Power System Technology, 2007, 31(21): 67-71.
- [3] 明玲玲, 李晓明. 基于 B/S 和 C/S 结合模式的地区模拟电力市场负荷考核系统[J]. 继电器, 2005, 33(4): 49-52
MING Ling-ling, LI Xiao-ming. A New Load Test Simulation System in Power Market of Local Region Based on B/S&C/S Combined Mode[J]. Relay, 2005, 33(4): 49-52.
- [4] 楼平, 张鹰. 基于 B/S 结构的变电站自动化信息采集发布系统[J]. 电网技术, 2006, 30(6): 92-94.
LOU Ping, ZHANG Ying. Automation Information Collection and Release of Transformer Substation Based on B/S Mode[J]. Power System Technology, 2006, 30(6): 92-94.
- [5] 胡俊, 沈利华, 等. 电力市场技术支持系统主站的三层体系方案[J]. 电力系统自动化, 2004, 28(2): 17-20.
HU Jun, SHEN Li-hua, et al. A Three-tier Infrastructure Solution Based on Socket API and ITS Application to Electricity Market[J]. Automation of Electric Power Systems, 2003, 13(13): 89-91.
- [6] 卫伟, 陈慧娟. 微软 .NET 平台初探[J]. 微机发展, 2003, 13(13): 89-91.
WEI Wei, CHEN Hui-juan. The Primarily Study of Microsoft .NET Application Architecture[J]. Microcomputer Development, 2003, 13(13): 89-91.
- [7] 马永光, 陆振国, 林永君. 基于 Internet 的电厂远程实时监控[J]. 电网技术, 2007, 31(2): 80-83.
MA Yong-guang, LU Zhen-guo, LIN Yong-jun. An Internet Based Remote Real-time Monitoring and Control System for Power Plant[J]. Power System Technology, 2007, 31(2): 80-83.
- [8] 陈亚红, 马丽, 等. 两种短期负荷预测精度考核标准的比较[J]. 电力系统自动化, 2003, 27(17): 73-77.
CHEN Ya-hong, MA Li, et al. Comparison Studies on Two Types of Accuracy Criteria for Short-term Load Forecast [J]. Automation of Electric Power Systems, 2003, 27(17): 73-77.

收稿日期: 2008-06-08; 修回日期: 2008-07-23

作者简介:

张红旭(1984-), 男, 硕士研究生, 主要从事电力市场软件开发、负荷预测及电力系统稳定性方向的研究;
E-mail: zhhexu@163.com

邱武(1972-), 男, 工程师、高级技师, 主要从事电力系统自动化方向研究;

姚建刚(1952-), 男, 教授, 博士生导师, 主要从事负荷预测、电力市场、配电系统自动化、高压绝缘等方面的研究与教学工作。

(上接第 24 页 continued from page 24)

- [9] 李元诚, 方廷健, 于尔铿, 等. 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(6): 55-59.
LI Yuan-cheng, FANG Ting-jian, YU Er-keng, et al. Study of Support Vector Machines for Short-term Load Forecasting[J]. Proceedings of the CSEE, 2003, 23(6): 55-59.
- [10] QI Miao, WANG Shi-fu. Nonlinear Model Predictive Control Based on Support Vector Regression[A]. In: Proceedings of the First International Conference on Machine Learning and Cybernetics[C]. Beijing: 2002.
- [11] Theodore B T, Huseyin I. Support Vector Machine for Regression and Application to Financial Forecasting[J]. Proceeding of the IEEE-INNSENNS International Joint Conference on Neural Network, 2000, (6): 348-353.
- [12] 谢开贵, 李春燕, 周家启, 等. 基于神经网络负荷组合预测模型研究[J]. 中国电机工程学报, 2002, 22(7): 85-89.
XIE Kai-gui, LI Chun-yan, ZHOU Jia-qi, et al. Research of the Combination Forecasting Model for Load Based on Artificial Neural Network[J]. Proceedings of the CSEE, 2002, 22(7): 85-89.
- [13] 王成山, 肖峻, 罗凤章. 多层分区空间负荷预测结果综合调整的区间方法[J]. 电力系统自动化, 2004, 28(12): 12-17.
WANG Cheng-shan, XIAO Jun, LUO Feng-zhang. Interval-based Multilayer Decomposed Calibration Method for Spatial Load Forecasting[J]. Automation of Electric Power Systems, 2004, 28(12): 12-17.

收稿日期: 2009-03-01

作者简介:

陈志强(1965-), 男, 博士研究生, 从事电力系统规划、可靠性及配电网设备优化检修以及需求管理研究。E-mail: czq@he.sgcc.com.cn