

基于扩展粗糙集的短期电力负荷预测模型

刘学琴¹, 吴耀华², 崔宝华¹

(1.保定电力职业技术学院电气工程系, 河北 保定 071051; 2.陕西理工学院电气工程系, 陕西 汉中 723003)

摘要: 影响短期电力负荷预测的因素众多, 如何有效地判断和选择这些相关因素是改善电力负荷预测的关键, 通过引入数据挖掘中粗糙集约简算法来解决这一难题。针对常规粗糙集算法计算量大, 且不具备容错性和泛化能力, 在属性约简过程中设置了分类可信度 β , 因而对数据具有了一定的容错性和泛化能力, 增强了抗噪声能力。经过对实际数据的计算分析, 证实了本文提出的方法在一定程度上提高了负荷预测的精度和速度。

关键词: 短期电力负荷预测; 粗糙集; 属性约简

Short-term load forecasting model based on extended rough set

LIU Xue-qin¹, WU Yao-hua², CUI Bao-hua¹

(1. Department of Electrical Engineering, North China Baoding Electric Power VOC.&TECH. College, Baoding 071051, China;
2. Department of Electrical Engineering, Shaanxi University of Technology, Hanzhong 723003, China)

Abstract: There are many factors that influence short-term load forecasting(STLF), how to justify and select the correlative factors is the key to improve the performance of load forecasting. A reduction algorithm based on rough set theory is proposed to mine more correlative attributes in the pending forecasting components. A reduction algorithm through classification reliability algorithm which with certain noise and having very good cover ability and generalizable ability through set classification reliability- β is introduced to overcome the large computational complexity of conventional reduction algorithm.

Key words: STLF; rough set; attribute reduction

中图分类号: TM732

文献标识码: A

文章编号: 1674-3415(2010)05-0025-04

0 引言

预测模型输入变量的选择不论是对于预测精度还是对于模型的收敛性能、计算速度的影响都很大, 在现有的文献中, 为了得到精确的预测结果往往将气象因素及各种相关信息都作为神经网络的输入参数, 这在一定程度上提高了预测精度, 但输入参数过多造成了网络结构复杂、训练费时, 同时容易陷入局部最小。如何选择输入变量是神经网络建模的关键之一, 目前文献多是根据经验来选择, 无法体现出各输入变量对于输出值影响的重要程度, 不能恰当地反映输入输出之间的实际关系。

粗糙集理论是由上近似集和下近似集组成的一种处理方法, 从一个新的角度将知识定义为对论域的划分能力, 并将其引入数学中的等价关系来进行讨论, 从而为数据分析, 特别是不精确、不完整数据分析提供了一套新的数学方法。研究表明: 将

粗糙集理论应用于电力负荷预测, 为预测模型输入变量的选取提供依据, 这样可以有效地考虑影响负荷预测的因素, 压缩输入变量。许多研究将粗糙集理论和预测模型相结合, 并取得了一定的成果^[1-2]。

本文提出了一种结合扩展粗糙集模型和神经网络模型的新预测方法, 首先采用基于扩展粗糙集原理的属性约简算法来寻找与负荷最直接相关的影响因素, 然后用得到的结果作为 BP 网络的输入进行训练预测。这样, 既全面考虑了影响负荷预测的各种因素, 又避免了由于输入变量过多而导致的神经网络拓扑结构复杂、训练时间过长的不足。通过粗糙集理论的属性约简算法, 既全面考虑了影响负荷预测的各种因素, 又避免了由于输入变量过多而导致的神经网络拓扑结构复杂、训练时间过长的不足。而且由于在属性约简过程中设置了分类可信度 β , 因而对数据具有了一定的容错性和泛化能力, 增强了抗噪声能力^[3-4]。

1 粗糙集理论基础

粗糙集理论作为一种处理不精确与不完全数据的新的数学理论,最初由波兰数学家 Pawlak 于 1982 年提出,到 90 年代初已引起各国学者的注意。粗糙集理论建立在分类机制的基础上,将知识理解为对数据的划分,是在特定空间上由等价关系构成的划分。粗糙集理论与其它处理不确定和不精确问题理论最显著的区别是:粗糙集理论无需提供问题所需处理的数据集合之外的任何先验信息,对问题的不确定性的描述或处理比较客观,又由于这个理论未包含处理不精确或不确定原始数据的机制,所以该理论与概率论、模糊数学、证据理论等其它处理不精确或不确定问题的理论有很强的互补性。

2 基于扩展粗糙集理论的负荷预测模型

一个准确的负荷预测模型应该能够准确地描述与负荷最直接相关的各种因素。由粗糙集理论的基本概念可以看出,粗糙集可以通过约简剔出不必要的信息,有利于信息的归类和简化。为了找到与负荷值直接相关的条件,以便改善预测精度和计算速度,本章利用粗糙集理论对影响负荷的各因素进行属性约简,寻找到与电力负荷直接相关的必要条件,将它们作为神经网络的输入矢量^[3-6]。

本文提出的基于粗糙集理论建立神经网络负荷预测模型的步骤如下:

- 1) 由负荷数据及相关信息历史数据建立初始信息表;
- 2) 对原始数据离散化处理,建立实数值型决策表;
- 3) 对建立的决策表进行属性约简,得到与预测负荷相关的最佳条件属性集;
- 4) 将此最佳条件属性集作为神经网络的输入对网络进行训练;
- 5) 若神经网络拟合误差满足要求,则结束。

2.1 初始信息表的建立

以日负荷数据、各气象因子等可能与待预测变量具有相关性的各种变量作为条件属性,得到条件属性集 $C = \{C_1, C_2, \dots, C_m\}$, 以待预测负荷作为决策属性集 $D = \{y\}$, 建立用于电力系统短期负荷预测的信息系统, 则任一时期负荷影响因素和负荷的历史数据构成一个对象 $x_i = \{C_{i,1}, C_{i,2}, \dots, C_{i,m}; y_i\}$, 论域为 $U = \{x_1, x_2, \dots, x_n\}$, 这时建立了一个决策数据模型 $S = \{U, C \cup D\}$ 。

值得注意的是, 由于气象影响具有累积效应,

连续多日高温和某日单独高温对该日的负荷影响程度大不相同。因此, 在选择负荷预测的条件属性集时不应只关注当日的气象因子, 而是要综合预测日前几日内的气象因素。本文选择预测日的日平均温度、日最高温度、日最低温度和日平均湿度以及预测日之前 13 天内每日的日最大负荷、日平均温度、日最高温度、日最低温度和日平均湿度共 69 个属性作为条件属性, 决策属性变量为预测日最大负荷。

2.2 属性值离散化

由于粗糙集能处理的是离散数据, 在应用约简算法前需要将连续值转化为离散值, 离散化方法很多, 如等距离或等频率划分法、最小信息熵离散化方法、基于布尔逻辑和基于属性重要性的离散化方法。本文采用建立模糊隶属函数的方法来对数据进行离散化, 其依据是各种数据所表现出来的特性。温度较低情况下的隶属度函数采用偏小型梯形分布为:

$$\mu_{low} = \begin{cases} 0 & t > 10 \\ (10-t)/(10-0) & 0 \leq t \leq 10 \\ 1 & t < 0 \end{cases}$$

温度适中情况下的隶属函数采用中间型梯形分布为

$$\mu_{avg} = \begin{cases} 0 & t > 25 \text{ 或 } t < 5 \\ (t-5)/(10-5) & 5 \leq t \leq 25 \\ (25-t)/(25-15) & 15 < t < 25 \end{cases}$$

温度较高情况下的隶属度函数采用偏大型梯形分布为

$$\mu_{high} = \begin{cases} 0 & t < 20 \\ (t-20)/(40-20) & 20 \leq t \leq 40 \\ 1 & t > 40 \end{cases}$$

温度的模糊隶属函数如图 1 所示。用类似的方法得到平均湿度的模糊隶属函数如图 2 所示。由于最大负荷的模糊隶属函数无明显的领域知识帮助决策, 所以本文中采用等距离划分, 其模糊隶属函数如图 3 所示。

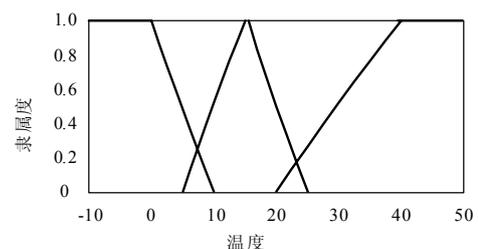


图 1 温度模糊隶属函数

Fig.1 Fuzzy membership function of temperature

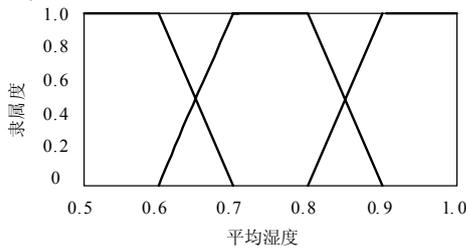


图 2 平均湿度模糊隶属函数

Fig.2 Fuzzy membership function of average humidity

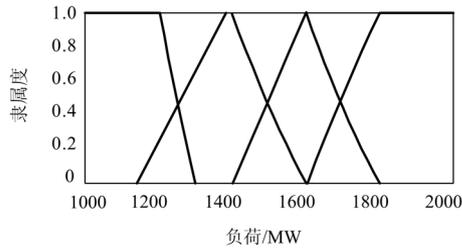


图 3 负荷模糊隶属函数

Fig.3 Fuzzy membership function of load

2.3 影响因素约简

分析影响负荷预测的主要因素,是在保持影响因子与负荷值之间的依赖关系不发生变化的前提下,通过对决策表中所有可能影响因素的约简,消除决策表中冗余或影响较小的因素,寻求负荷变化的主要因素,根据本文提出的基于扩展分类的粗糙集模型对影响负荷预测的各影响因子进行属性约简,其具体操作步骤如下:

1) 给定分类可信度 $0.5 \leq \beta \leq 1$ (本文取 0.85),

对于条件属性集 C , 计算 $\gamma_C^\beta(D)$;

2) 计算可分辨矩阵并计算出每个属性出现的次数 t_{a_i} ;

3) 利用公式 $f(a_i) = \sum_m^n t_{a_m} / m$ 计算每个属性频率函数的重要性 $f(a_i)$;

4) 令 $Red = \emptyset$;

5) 选择属性重要性 $f(a_i)$ 最大的属性记做 ρ ,

$$Red = Red \cup \{\rho\};$$

6) 如果 Red 满足 $\gamma_{Red}^\beta(D) = \gamma_C^\beta(D)$ 则终止, 否则转 5);

7) 最后得到的 Red 就是 C 相对于 D 的一个约简。

3 实例分析

结合陕西某市的电力负荷历史数据及天气状况,利用本文提出的方法对该市 2001 年 8 月份的负荷建立模型进行预测。

首先,将该市 2001 年 5 月到 7 月的数据(包括日最大负荷、日平均温度、日最高温度、日最低温度和日平均湿度)进行归一化处理作为基本数据建立属性值决策表。建立的初始决策表格式如下所述:数据 1~5 为预测日前 13 天(d-13 日)的日最大负荷及气象因子(气象因子包括日平均温度、日最高温度、日最低温度和日平均湿度);数据 6~65 为预测日前 12 天(d-12 日)到前一天(d-1 日)的日最大负荷及气象因子;数据 66-69 为预测日的气象因子。上述 69 个数据组成了初始决策表的条件属性,而决策属性变量取为预测日的日最大负荷。最终形成的决策表为一个 97 行(样本数)、70 列(条件属性个数及决策属性个数之和)的表格。

表 1 约简后得到的条件属性集

Tab.1 Conditional attributes after reduction

输入变量	属性重要性
预测日的最高温度	0.038 3
预测日的平均温度	0.150 6
预测日前 1 天的最高温度	0.178 6
预测日前 1 天的日最大负荷	0.205 1
预测日前 2 天的平均湿度	0.026 8
预测日前 5 天的最高温度	0.056 2
预测日前 7 天的平均湿度	0.132 6
预测日前 7 天的日最大负荷	0.196 0

其次,利用前文所述的方法对属性表进行离散化并约简,约简结果见表 1。由表 1 可以看出:虽然最初的条件属性多达 69 个,经过本文的算法约简后,属性个数大大减少,仅为 8 个。

最后,利用前述的三层 BP 神经网络模型对 2001 年 8 月份的负荷进行预测。具体为:对 8 月份的每一日单独建立一个 BP 网络,共建立 31 个。每个独立的 BP 网络输入层神经元个数取为 8,隐层神经元个数依据 Kolmogrov 定理,取为 17 个,输出层取为 1。训练时,将连接权及阈值进行初始化,随机设置使其范围处于 $[-0.5, +0.5]$ 之间。动量系数取为 0.95,学习速率依据试验确定取为 $0.428^{[3]}$,限制误差为 0.000 1,最大训练次数 5 000。

表 2 预测结果与精度对照
Tab.2 Comparison of the results and precision

预测日	A1	A2	实际值	E1	E2
01	1 754.773	1 771.778	1 797.5	2.377	1.431
02	1 781.919	1 804.398	1 807.0	1.388	0.144
03	1 744.975	1 735.126	1 761.8	0.955	1.514
04	1 739.091	1 759.491	1 792.6	2.985	1.847
05	1 908.234	1 944.828	1 999.7	4.574	2.744
06	1 578.981	1 602.402	1 619.7	2.514	1.068
07	1 449.246	1 473.471	1 475.3	1.766	0.124
08	1 146.070	1 148.683	1 150.8	0.411	0.184
09	1 179.799	1 192.638	1 198.8	1.585	0.514
10	1 172.684	1 176.723	1 198.5	2.154	1.817
11	1 174.125	1 193.811	1 212.2	3.141	1.517
12	1 221.710	1 214.180	1 244.7	1.847	2.452
13	1 285.558	1 285.132	1 292.2	0.514	0.547
14	1 318.115	1 339.280	1 356.7	2.844	1.284
15	1 247.227	1 263.221	1 266.4	1.514	0.251
16	1 136.984	1 163.818	1 166.7	2.547	0.247
17	1 097.591	1 124.279	1 137.6	3.517	1.171
18	1 172.641	1 165.174	1 187.1	1.218	1.847
19	1 224.128	1 228.273	1 260.0	2.847	2.518
20	1 160.443	1 172.940	1 179.0	1.574	0.514
21	1 255.872	1 262.990	1 266.6	0.847	0.285
22	1 352.365	1 355.665	1 369.4	1.244	1.003
23	1 393.074	1 410.217	1 423.8	2.158	0.954
24	1 432.073	1 432.218	1 450.7	1.284	1.274
25	1 237.585	1 249.556	1 269.5	2.514	1.571
26	1 385.097	1 405.676	1 434.1	3.417	1.982
27	1 445.623	1 461.554	1 477.9	2.184	1.106
28	1 374.388	1 392.294	1 395.7	1.527	0.244
29	1 421.566	1 417.732	1 425.1	0.248	0.517
30	1 392.233	1 402.629	1 418.1	1.817	1.091
31	1 510.495	1 524.891	1 544.6	2.208	1.276

网络预测的结果与精度对照表如表 2。其中，A1 和 E1 分别表示未采用属性约简算法，仅按经验选取当日最高温度、当日最低温度、当日平均温度以及当日平均湿度为基本气象信息进行负荷预测时的日最大负荷(MW)及相对误差(%), A2 和 E2 分别表示采用由本文属性约简算法得到的最小属性集进行预测时的日最大负荷(MW)及相对误差(%)。

由表 2 可以看出，采用本文所提出的预测模型对负荷预测精度的提高有了很大的改善，大部分日的预测误差维持在 2%以下，只有 5 日、12 日、19

日的预测误差保持在 2%到 4%以内，究其原因是一些点大多属于周末负荷，因为影响周末负荷的因素不只是文中所列的这些因素，更多还有许多没有被发掘的人为因素等影响。

4 结论

引入粗糙集理论中的属性约简算法挖掘与待预测测量相关性大的属性作为模型输入量，充分保证了预测模型输入参数的合理性，可以解决神经网络模型输入参数的确定问题。

由于传统粗糙集属性约简算法计算量大、泛化能力差、无法抵御噪声对属性约简过程的干扰，本文采用了基于分类可信度的启发式属性约简算法来挖掘与预测测量相关性的属性。该算法在属性约简过程中设置了分类可信度 β ，因而对数据具有了一定的容错性和泛化能力，增强了抗噪声能力。

经过对实际数据的计算分析，证实了本文提出的方法在一定程度上提高了负荷预测的精度和速度。

参考文献

[1] 刘同明, 等. 数据挖掘技术及其应用[M]. 北京: 国防工业出版社, 2001.

[2] Pawlak Z. Rough Sets: Theoretical Aspects of Reasoning about Data[M]. Dordrecht Kluwer Academic Publishers, 1991.

[3] 钟波, 周家启, 肖智. 基于粗糙集与神经网络的电力负荷新型预测模型 [J]. 系统工程理论与实践, 2004 (6): 113-119.
ZHONG Bo, ZHOU Jia-qi, XIAO Zhi. A New Power Load Forecasting Model Based on Rough Set and Artificial Neural Network[J]. Systems Engineering Theory & Practice, 2004(6): 113-119.

[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 Set Theory to Establish Artificial Neural Networks for Short Term Load Forecasting[J]. Proceedings of the CSEE, 2003, 23(11): 1-4.

[5] 李元诚, 方廷健. 一种基于粗糙集理论的 SVM 短期负荷预测方法 [J]. 系统工程与电子技术, 2004, 26 (2): 187-190.
LI Yuan-cheng, FANG Ting-jian. Approach to Forecast Short Term Load of SVM Based on Rough Sets[J]. Systems Engineering and Electronics, 2004, 26 (2): 187-190.

(下转第 38 页 continued on page 38)

生影响, 将地理因子直接引入目标函数, 并不另外给经济适应度和地理适应度设定权重比例, 使之更为合理。

需要说明的是本文考虑地理因子的影响更多的是基于城市配网的角度来分析, 所取地理因子比较接近。对于可能具有复杂地形的农村配网, 不同地理因子的选取对总的费用的影响差别将很大。且在将地理因子各因素由定性描述转化为定量计算的过程中具有较大的可塑空间, 如何提高计算精确度有待进一步研究。

参考文献

[1] 姚建刚, 曹一家, 江全元. 电力系统工程学[M]. 北京: 清华大学出版社, 2007.
YAO Jian-gang, CAO Yi-jia, JIANG Quan-yuan. Power System and Its Engineering[M]. Beijing: Tsinghua University Press, 2007.

[2] 董永峰, 杨彦卿, 宋洁, 等. 基于改进粒子群算法的变电站选址规划[J]. 继电器, 2008, 36(5): 32-35.
DONG Yong-feng, YANG Yan-qing, SONG Jie, et al. Optimal Planning of Substation Locating Based on Improved PSO Algorithm[J]. Relay, 2008, 36(5): 32-35.

[3] 刘自发, 张建华. 基于改进多组织粒子群体优化算法的配电网变电站选址定容[J]. 中国电机工程学报, 2007, 27(1): 105-111.
LIU Zi-fa, ZHANG Jian-hua. Optimal Planning of Substation Locating and Sizing Based on Refined Multi-team PSO Algorithm[J]. Proceedings of the CSEE, 2007, 27(1): 105-111.

[4] Miguez E, Diaz-dorado E, Cidras J. An Application of An Evolution Strategy in Power Distribution System Planning[A]. in: Evolutionary Computation Proceedings of IEEE World Congress on Computational Intelligence[C]. Anchorage(AK,USA): 1998. 241-246.

[5] 王成山, 魏海洋. 变电站选址定容两阶段优化规划方法[J]. 电力系统自动化, 2005, 29(4): 62-66.
WANG Cheng-shan, WEI Hai-yang. Two-phase Optimization Planning Approach to Substation Locating and Sizing[J]. Automation of Electric Power Systems, 2005, 29(4): 62-66.

[6] 陈颖, 徐晓辉, 李志全. 基于免疫克隆原理的改进粒子群优化算法的研究[J]. 系统仿真报, 2008, 20(6): 1471-1474.
CHEN Ying, XU Xiao-hui, LI Zhi-quan. Study of Modified Particle Swarm Optimization Algorithm Based on Immune Clone Principle[J]. Journal of System Simulation, 2008, 20(6): 1471-1474.

[7] 胡旺, 李志蜀. 一种更简化而高效的粒子群优化算法[J]. 软件学报, 2007, 18(4): 861-868.
HU Wang, LI Zhi-shu. A Simpler and More Effective Particle Swarm Optimization Algorithm[J]. Journal of Software, 2007, 18(4): 861-868.

[8] 闫丽梅, 许爱华, 任爽, 等. 一种变电站选址的新方法[J]. 高电压技术, 2007, 33(9): 75-79.
YAN Li-mei, XU Ai-hua, REN Shuang, et al. New Model of Selecting Location of Transformer Substation[J]. High Voltage Engineering, 2007, 33(9): 75-79.

收稿日期: 2009-03-26; 修回日期: 2009-05-04

作者简介:

严司玮(1985-), 男, 硕士研究生, 研究方向为电力市场相关软件开发、电网规划及电力系统稳定性; E-mail: enosfrank@yahoo.com.cn

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、高压外绝缘等方面的研究与教学工作;

李丰涛(1983-), 男, 硕士研究生, 研究方向为电网规划与电力市场。

(上接第 28 页 continued from page 28)

[6] 康重庆, 夏清, 张伯明. 电力系统负荷预测研究综述与发展方向的探讨[J]. 电力系统自动化, 2004, 28(17): 1-11.
KANG Chong-qing, XIA Qing, ZHANG Bo-ming. The Study of Power Load Forecasting Overview and Development Direction[J]. Automation of Electric Power

Systems, 2004, 28(17): 1-11.

收稿日期: 2009-03-10; 修回日期: 2009-04-09

作者简介:

刘学琴(1973-), 女, 硕士, 讲师, 主要从事电力系统自动化教学与研究。E-mail: liu2291292@126.com