

基于负荷特性聚类的样本自适应神经网络台区短期负荷预测

方芳¹, 卜凡鹏², 田世明², 齐林海³, 李夏威⁴

1. 国网北京市电力公司昌平供电公司, 北京 102200
2. 中国电力科学研究院, 北京 100192
3. 华北电力大学控制与计算机工程学院, 北京 102206
4. 华北电力大学电气与电子工程学院, 北京 102206

摘要 介绍了批量处理时间序列数据情况下,基于台区负荷特性聚类的样本自适应反向传播神经(BP)神经网络预测短期电力负荷的方法,通过对历史数据的预处理、初始聚类中心的设置以及最优聚类数目的确定,建立典型日负荷曲线的聚类预测模型。基于历史数据的聚类结果及待预测日的温度、湿度、气压、风速、星期等相关参数,使用BP神经网络算法得出待预测日负荷曲线预测结果。通过实例验证,基于台区负荷特性聚类的样本自适应神经网络短期负荷预测能够得到较为准确的预测结果。

关键词 聚类分析;人工神经网络;电力负荷预测;数据挖掘

负荷预测是电力部门进行电力运行研究和配电网规划的重要组成部分,是确保电力系统安全有效和经济运行的根本,是电力规划建设的先决条件^[1-4]。低压台区负荷预测是供电部门的新兴工作之一,通过准确的负荷预测,可以经济合理地调整运行方式,减少上级电站备用容量、合理安排检修计划、降低运营成本、提高经济效益^[5-8]。

电力负荷传统预测方法包含时间序列法、回归分析法等,在现代负荷预测中,用于负荷预测的方法有人工神经网络,如反向传播神经(BP)网络、径向基函数(RBF)网络等^[9-13]。台区负荷具有波动性、随机性较大等特点,为台区的负荷预测带来难度。本文提出一种考虑多台区负荷特性聚类的样本自适应神经网络短期负荷预测模型,同时考虑待预测日的温度、湿度、气压、风速、星期、节假日等相关参数,得出典型日负荷曲线预测结果。通过大数据分析平台实例验证,该方法能够得到较为准确的预测结果。

1 负荷预测的理论计算

随着智能电网、通信网络技术和传感器技术的发展,电力公司建设了营销业务、用电信息采集、能量管理等信息系

统对电网的运行进行监控和管理,这些系统的建设积累了大量电网运行数据。数据呈现数据量大、数据种类多、数据所蕴藏的价值大等特点^[14-17]。

提高预测精度的途径有多种:其一可以通过扩大用于负荷预测的数据数量与数据密度,如增大用电信息采集系统的采样频率,将更多历史数据及更多方面的气象、经济数据考虑进预测系统中;其二可以引入更多先进可用的预测算法,并结合电力负荷自身特性,更新优化传统的负荷预测模型;其三可以通过聚类、分类等手段,将地区内负荷按照各自各时期的特性分为不同类别,再依据不同类的负荷分别加以建模预测,再综合汇总得到一个较为精确的预测结果。

本文通过聚类分析的方式对多台区历史数据样本进行分类处理,建立典型日负荷曲线的聚类BP神经网络算法预测模型,实现台区较准确的负荷预测。

1.1 聚类计算

K-means 算法接受参数 k ;然后将事先输入的 n 个数据对象划分为 k 个聚类以便使得所获得的聚类满足:同一聚类中的对象相似度较高;而不同聚类中的对象相似度较小。聚类相似度是利用各聚类中对象的均值所获得一个“中心对象”

收稿日期:2017-11-02;修回日期:2017-11-24

基金项目:国家电网公司科技项目(52094017002U)

作者简介:方芳,工程师,研究方向为智能用电,电子信箱:411912023@qq.com;卜凡鹏(通信作者),工程师,研究方向为智能用电,电子信箱:bupanpeng@epri.sgcc.com.cn

引用格式:方芳,卜凡鹏,田世明,等.基于负荷特性聚类的样本自适应神经网络台区短期负荷预测[J].科技导报,2017,35(24):66-70;doi:10.3981/j.issn.1000-7857.2017.24.008

(引力中心)进行计算的。

K-means 需要优化的目标函数:设共有 N 个数据点需要分为 K 个 cluster, K-means 要做的就是最小化

$$J = \sum_{n=1}^N \sum_{k=1}^K \tau_{nk} \|x_n - \mu_k\| \quad (1)$$

式中, τ_{nk} 在数据点 n 被归类到 cluster k 时为 1, 否则为 0。

通常直接找 τ_{nk} 和 μ_k 来最小化 J 并不容易, 本文采取迭代的办法: 先固定 μ_k , 选择最优的 τ_{nk} , 很容易得出, 只要将数据点归类到离其最近的中心便能保证 J 最小。下一步则固定 τ_{nk} , 再求最优的 μ_k 。将 J 对 μ_k 求导并令导数等于 0, 很容易得到 J 最小时, μ_k 应该满足

$$\mu_k = \frac{\sum_n \tau_{nk} x_n}{\sum_n \tau_{nk}} \quad (2)$$

亦即 μ_k 的值应当是所有 cluster k 中的数据点的平均值。由于每一次迭代都是取到 J 的最小值, 因此 J 只会不断地减小 (或者不变), 而不会增加, 这保证了 K-means 最终会到达一个极小值。

K-means 算法的步骤如下:

- 1) 选定 K 个中心 μ_k 的初值。
- 2) 将每个数据点归类到离它最近的中心点所代表的 cluster 中。
- 3) 采用式(3)计算每个 cluster 的新的中心点

$$\mu_k = \frac{\sum_n \tau_{nk} x_n}{\sum_n \tau_{nk}} \quad (3)$$

- 4) 重复第 2 步, 一直到迭代到最大的步数或者前后的 J 的值相差小于给定阈值为止。

1.2 BP 神经网络

BP 神经网络的过程主要分为 2 个阶段, 第 1 阶段是信号的前向传播, 从输入层经过隐含层, 最后到达输出层; 第 2 阶段是误差的反向传播, 从输出层到隐含层, 最后到输入层, 依次调节隐含层到输出层的权重和偏置, 输入层到隐含层的权重和偏置。

需要依据信号的前向传播和误差的反向传播来构建整个网络。

- 1) 网络的初始化。

假设输入层的节点个数为 n , 隐含层的节点个数为 l , 输出层的节点个数为 m 。输入层到隐含层的权重 w_{ij} , 隐含层到输出层的权重为 w_{jk} , 输入层到隐含层的偏置为 a_j , 隐含层到输出层的偏置为 b_k 。学习速率为 η , 激励函数为 $g(x)$ 。其中激励函数为取 Sigmoid 函数。形式为

$$g(x) = \frac{1}{1 + e^{-x}} \quad (4)$$

- 2) 隐含层的输出。

如图 1 所示的 3 层 BP 网络所示, 隐含层输出为

$$H_j = g\left(\sum_{i=1}^n w_{ij} x_i + a_j\right) \quad (5)$$

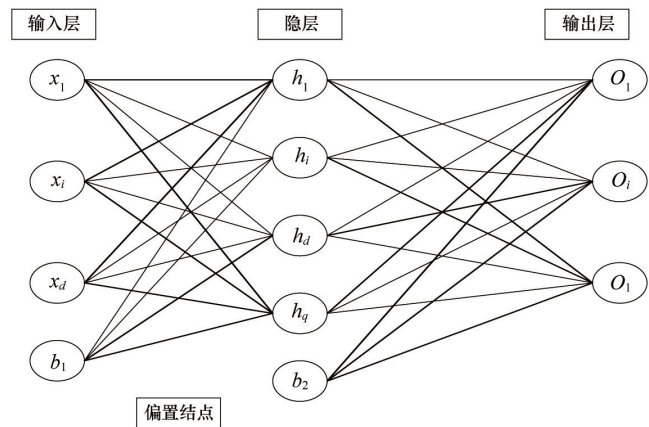


图 1 BP 神经网络

Fig. 1 BP neural network diagram

- 3) 输出层的输出。

$$O_k = \sum_{j=1}^l H_j w_{jk} + b_k \quad (6)$$

- 4) 误差的计算。

取误差为

$$E = \frac{1}{2} \sum_{k=1}^m (Y_k - O_k)^2 \quad (7)$$

式中, Y_k 为期望输出。记 $Y_k - O_k = e_k$, 则可以表示为

$$E = \frac{1}{2} \sum_{k=1}^m e_k^2 \quad (8)$$

式中, $i=1, 2, \dots, n; j=1, 2, \dots, l; k=1, 2, \dots, m$ 。

- 5) 权值的更新。

权值的更新公式为

$$\begin{cases} w_{ij} = w_{ij} + \eta H_j (1 - H_j) x_i \sum_{k=1}^m w_{jk} e_k \\ w_{jk} = w_{jk} + \eta H_j e_k \end{cases} \quad (9)$$

2 实验过程

2.1 聚类分析

- 1) 从输入的数据点集合中随机选择一个点作为第一个聚类中心。
- 2) 对于数据集中的每一个点 x , 计算它与最近聚类中心的距离 D 。
- 3) 选择一个新的数据点作为新的聚类中心, 选择原则为: $D(x)$ 较大的点, 被选取作为聚类中心的概率较大。
- 4) 重复步骤 2 和 3, 直到 k 个聚类中心被选出来。
- 5) 利用这 k 个初始的聚类中心运行标准的 K-means 算法。

从上面的算法描述上可以看到, 算法的关键是第 3 步, 如何将 $D(x)$ 反映到点被选择的概率上, 方法如下:

- 1) 先从数据库随机挑个随机点当“种子点”。
- 2) 对于每个点, 都计算距其最近的一个“种子点”的距离 $D(x)$ 并保存在一个数组里, 把这些距离相加得到 $\text{Sum}(D(x))$ 。

3) 再取一个随机值,用权重的方式取计算下一个“种子点”。这个算法的实现方法是,先取一个能落在 $\text{Sum}(D(x))$ 中的随机值 Random , 然后用 $\text{Random}=D(x)$, 直到 $\text{Random} \leq 0$, 此时的点就是下一个“种子点”。

4) 重复步骤 2 和 3, 直到 k 个聚类中心被选出来。

2.2 建模

1) 虚拟预测日相似日选择样本。

根据聚类结果和影响因素的选择结果,确定了预测日的特征值为温度、湿度、气压、风速,将其确定为研究对象,并计算预测日与历史日气象的因素相似度。总体可由 2 部分构成,即基础负荷部分和气象敏感性负荷。气象敏感性负荷主要受气象因素的影响,同时还会受到假期、星期、月份等时间因素的影响。采用皮尔逊相关性系数计算负荷与各影响因素的相关性系数:

$$\gamma_{xy} = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2}} \quad (10)$$

(1) 相关系数的值介于 -1 与 +1 之间,即 $-1 \leq r \leq +1$ 。

(2) 当 $r > 0$ 时,表示两变量正相关,当 $r < 0$ 时,表示两变量为负相关。当 $|r|=1$ 时,表示两变量为完全线性相关即函数关系。当 $r=1$ 时,表示两变量间完全正相关,而当 $r=-1$ 时,表示两变量间完全负相关。当 $r=0$ 时,表示两变量间无线性相关关系。

根据相关性系数,最终挑选出与虚拟预测日相似的样本数据集。

2) 样本归一化。

对挑选的样本数据集按列进行 L2 范数归一化。

向量 $\mathbf{x}=(x_1, x_2, \dots, x_n)$ 的 L2 范数定义为: $\text{norm}(\mathbf{x}) = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$, 要使得 x 归一化到单位 L2 范数,即建立一个从 x 到 x' 的映射,使得 x' 的范数为 1, 则

$$\begin{aligned} 1 &= \text{norm}(\mathbf{x}') = \frac{\sqrt{x_1'^2 + x_2'^2 + \dots + x_n'^2}}{\text{norm}(\mathbf{x})} \\ &= \frac{\sqrt{x_1^2 + x_2^2 + \dots + x_n^2}}{\text{norm}(\mathbf{x})^2} \\ &= \sqrt{x_1'^2 + x_2'^2 + \dots + x_n'^2} \end{aligned} \quad (11)$$

即

$$\mathbf{x}'_i = \frac{x_i}{\text{norm}(\mathbf{x})}$$

3) BP 神经网络模型。

(1) 初始化 BP 网络,使用 BP 神经网络对样本数据进行批处理学习。

(2) 设置迭代次数及迭代步长,进行 BP 网络学习。

(3) 设置预测精度,计算预测结果,进行参数修正。

(4) 符合要求,输出结果。

流程如图 2 所示。

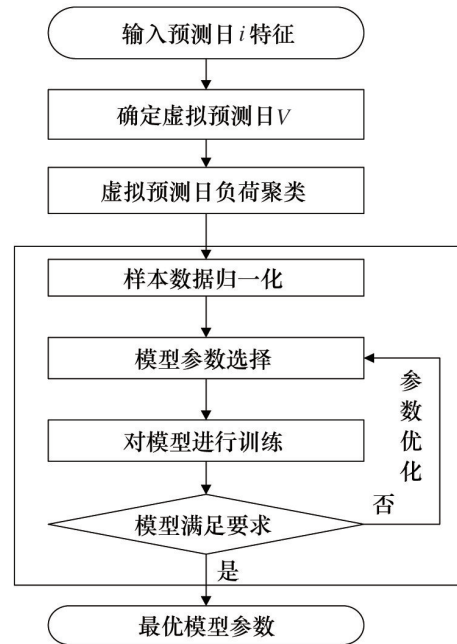


图 2 模型流程

Fig. 2 sample adaptive model

3 结果与分析

3.1 聚类结果分析

图 3 为某地 12400 台区负荷数据的聚类结果,通过类中心选取函数多次聚类最终把预测日当天的台区负荷聚为 5 类。第 1 类台区呈现白天高峰、晚上高峰的双峰特点;第 2 类台区用电相对稳定;第 3 类台区呈现晚上用电高峰的特点;第 4 类台区呈现晚上用电,白天不用电的特点;第 5 类台区呈现白天用电高峰,晚上用电低峰的特点。

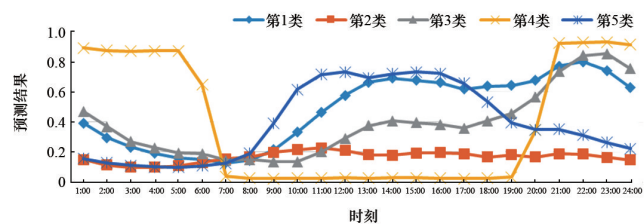


图 3 台区负荷聚类结果

Fig. 3 Clustering results of load data

3.2 预测结果分析

基于某地区 2014—2017 年 12400 个台区抽取 80% 的历史负荷数据作为训练样本,建立自适应样本的 BP 神经网络预测模型及多元线性回归模型。同时以聚类结果中其中一类负荷剩余 20% 历史负荷数据进行测试,预测结果如图 4 所

示。其中基于自适应样本的BP神经网络预测的准确率为93.5%，而基于多元线性回归模型预测的准确率为87.5%，结

果显示本文所建立的自适应样本BP神经网络模型科学有效，且较传统方法预测有较高的准确率。

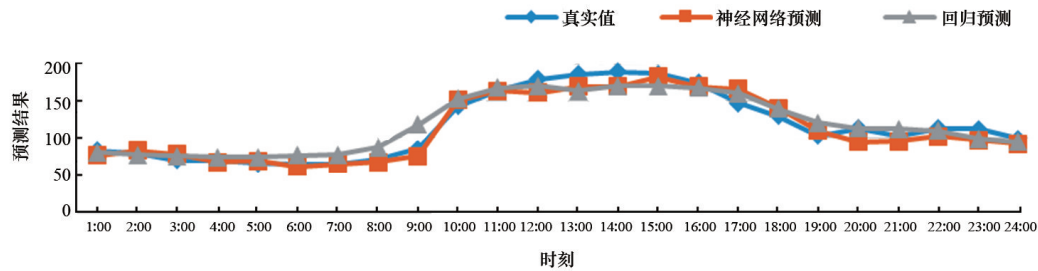


图4 预测结果对比

Fig. 4 Comparison of prediction results

4 结论

本文介绍了基于台区负荷特性聚类的样本自适应BP神经网络预测短期负荷的方法，通过对历史负荷数据的处理、初始聚类中心的设定和最优聚类数目的确定，建立了典型日负荷曲线的自适应样本的BP神经网络预测模型。通过实例对比验证，该模型在实际负荷预测中科学有效，且较多元线性回归模型预测具有较高的准确率。

参考文献 (References)

- [1] 陈宏义, 李存斌, 施立刚. 基于聚类分析的短期负荷智能预测方法研究[J]. 湖南大学学报(自然科学版), 2014, 41(5): 94-98.
Chen Hongyi, Li Cunbin, Shi Ligang. A new forecasting approach for short-term load intelligence base on cluster method[J]. Journal of Hunan University(Natural Sciences), 2014, 41(5): 94-98.
- [2] 张素香, 刘建明, 赵丙镇, 等. 基于云计算的居民用电行为分析模型研究[J]. 电网技术, 2013, 37(6): 1542-1546.
Zhang Suxiang, Liu Jianming, Zhao Bingzhen, et al. Cloud computing-based analysis on residential electricity consumption behavior[J]. Power System Technology, 2013, 37(6): 1542-1546.
- [3] 张璨, 王蓓蓓, 李扬. 典型行业用户需求响应行为研究[J]. 华东电力, 2012, 40(10): 1701-1705.
Zhang Can, Wang Beibei, Li Yang. Demand response behavior of typical industrial users[J]. East China Electric Power 2012, 40(10): 1701-1705.
- [4] 吴潇雨, 和敬涵, 张沛, 等. 基于灰色投影改进随机森林算法的电力系统短期负荷预测[J]. 电力系统自动化, 2015, 39(12): 50-55.
Wu Xiaoyu, He Jinghan, Zhang Pei, et al. Power system short-term load forecasting based on improved random forest with grey relation projection[J]. Automation of Electric Power Systems, 2015, 39(12): 50-55.
- [5] 朱健峥. 短期台区电力负荷预测方法探讨[J]. 机电信息, 2010(18): 120-121.
Zhu Jianzheng. Discussion on short term power load forecasting method in taiqu[J]. Electromechanical Information, 2010(18): 120-121.
- [6] 张晓虎, 罗隆福, 刘洁. 基于负荷实测的配电网无功优化及其降损节能效益分析[J]. 电网技术, 2012, 36(5): 226-231.
Zhang Xiaohu, Luo Longfu, Liu Jie. Load measurement based reactive power optimization for distribution network and analysis on its effects in loss reduction and energy conservation[J]. Power System Technology, 2012, 36(5): 226-231.
- [7] 王新, 孟玲玲. 基于EEMD-LSSVM的超短期负荷预测[J]. 电力系统保护与控制, 2015, 43(1): 61-66.
Wang Xin, Meng Lingling. Ultra-short-term load forecasting based on EEMD-LSSVM[J]. Power System Protection and Control, 2015, 43(1): 61-66.
- [8] Tucci M, Crisostomi E, Giunta G, et al. A multi-objective method for short-term load forecasting in European countries[J]. IEEE Transactions on Power Systems, 2016, 31(5): 3537-3547.
- [9] 张刚, 刘福潮, 王维洲, 等. 电网短期负荷预测的BP-ANN方法及应用[J]. 电力建设, 2014, 35(3): 54-58.
Zhang Gang, Liu Fuchao, Wang Weizhou, et al. BP-ANN Method for power grid short-term load forecasting and its application[J]. Power Construction, 2014, 35(3): 54-58.
- [10] 刘念, 张清鑫, 刘海涛. 基于核函数极限学习机的微电网短期负荷预测方法[J]. 电工技术学报, 2015, 30(8): 218-224.
Liu Nian, Zhang Qingxin, Liu Haitao. Online short-term load forecasting based on ELM with kernel algorithm in micro-grid environment [J]. Transactions of China Electrotechnical Society, 2015, 30(8): 218-224.
- [11] Ul Islam B, Baharudin Z, Nallagownden P. Short term electric load forecasting with back propagation neural network and simulated annealing genetic algorithm[J]. Applied Mechanics & Materials, 2015, 785: 14-18.
- [12] 金义雄, 段建民, 徐进, 等. 考虑气象因素的相似聚类短期负荷组合预测方法[J]. 电网技术, 2007, 31(19): 60-64.
Jin Yixiong, Duan Jianmin, Xu Jin, et al. A combinational short-term load forecasting method by use of similarity clustering and considering weather factors[J]. Power System Technology, 2007, 31(19): 60-64.
- [13] Duan Q H, Liu J R, Zhao D F. Short term electric load forecasting using an automated system of model choice[J]. International Journal of Electrical Power & Energy Systems, 2017, 91: 92-100.
- [14] Baliyan A, Gaurav K, Mishra S K. A review of short term load forecasting using artificial neural network models[J]. Procedia Computer Sci-

- ence, 2015, 48: 121-125.
- [15] 康重庆, 周安石, 王鹏, 等. 短期负荷预测中实时气象因素的影响分析及处理策略[J]. 电网技术, 2006(7): 5-10.
Kang Chongqing, Zhou Anshi, Wang Peng, et al. Impact analysis of hourly weather factors in short-term load forecasting and its processing strategy[J]. Power System Technology, 2006(7): 5-10.
- [16] 方芳, 田世明, 卜凡鹏, 等. 一种自适应选择样本的用电负荷预测方法[J]. 电子技术应用, 2017, 43(11): 18-21, 26.
Fang Fang, Tian Shiming, Bu Fanpeng, et al. An electric load forecasting method based on adaptive selection of samples[J]. Application of Electronic Technique, 2017, 43(11): 18-21, 26.
- [17] 王德文, 孙志伟. 电力用户侧大数据分析与并行负荷预测[J]. 中国电机工程学报, 2015, 35(3): 527-536.
Wang Dewen, Sun Zhiwei. Big data analysis and parallel load forecasting of electric power user side[J]. Proceedings of the CSEE, 2015, 35(3): 527-536.

A neural network for short term load forecasting based on Sample self adapted of load characteristics clustering

FANG Fang¹, BU Fanpeng², TIAN Shiming², QI Linhai³, LI Xiawei⁴

1. State Grid Beijing Changping Electric Power Supply Company, Beijing 102200, China
2. China Electric Power Research Institute, Beijing 100192, China
3. School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China
4. School of Electrical and Electronic Engineering, North China Electric Power University, Beijing 102206, China

Abstract This paper introduces the methods and the steps of predicting the power load by the BP neural network with cluster optimization in batch processing time series. Through the preconditioning of historical data, the setting of the initial clustering center and the determination of the optimal number of clusters, a clustering prediction model of the load curve is established based on the clustering results of the historical data and the relevant parameters such as the temperature, the humidity, the air pressure, the wind speed and the time (the current week). The results show that with the clustering algorithm, the related factors and the BP network adaptive rate can be comprehensively considered, while the training speed is improved, to obtain more accurate prediction results.

Keywords cluster analysis; neural networks; power load forecasting; data mining

(责任编辑 刘志远)