

基于 EUNITE 竞赛数据的中期电力负荷预测

李 炎¹, 翟永杰², 周 倩², 韩 璞²

(1. 华北电力大学 后勤与资产管理处, 河北 保定 071003;
2. 华北电力大学 控制科学与工程学院, 河北 保定 071003)

摘要: 应用自组织模糊神经网络 (SOFNN) 算法, 基于欧洲智能技术网络 (EUNITE) 竞赛数据进行了中期电力负荷预测的应用研究。算法能够自动决定神经模型的结构并得出模型的参数, 具有很好的实用价值。研究了训练数据选取和输入特征向量编码等实际应用问题, 结果表明负荷预测精度高, 优于竞赛的优胜者, 之后提出了结合周平均负荷预测修正日负荷预测的方法, 精度得到进一步地提高。

关键词: 模糊神经网络; 自组织; 负荷预测

中图分类号: TM715 **文献标识码:** A **文章编号:** 1007-2691(2007)04-0022-05

Mid-term load forecasting based on EUNITE competition data

LI Yan¹, ZHAI Yong-jie², ZHOU Qian², HAN Pu²

(1. Department of Logistics and Assets Administration, North China Electric Power University, Baoding 071003, China;
2. School of Control Science and Engineering, North China Electric Power University, Baoding 071003, China)

Abstract: The self-organizing fuzzy neural network (SOFNN) algorithm was applied for load forecasting with the data of EUNITE competition. The algorithm can automatically determine the model structure and identify the model parameters. The training data selection and input variables encoding problems were also discussed in details. The results show that SOFNN can outperform winners' models by providing very promising prediction accuracy. Further more, the average load value in each week was forecasted to revise the daily load. Through this procedure, better forecasting accuracy can be achieved.

Key words: fuzzy neural networks; self-organizing; load forecasting

0 引 言

电力负荷预测在电力工业中是一个非常重要的研究课题。根据不同的预测时间长度, 可以分为短期、中期和长期, 短期预测一般指由日到周的预测, 中期为由月到年的预测, 长期则是几年甚至更长时间的预测。

由于较长时间的预测 (中期或长期) 可能会因为预测误差的累积而变得不可靠, 因而更多的研究总是集中在短期负荷预测, 尤其是提前一天的负荷预测。然而, 中期负荷预测和长期负荷预测也具有实际意义, 仍然是值得深入研究的课

题^[1]。

2001 年, 欧洲智能技术网络 (European Network on Intelligent Technologies, EUNITE) 组织了一次中期电力负荷预测竞赛。竞赛目的是预测一个月中每一天的最大负荷。竞赛提供了一些信息, 包括历史负荷数据、节假日数据和历史气象数据。竞赛中, Chen, Chang^[1] 使用 Support Vector Machines (SVM) 取得了优胜。连同其他方法的结果, 如神经网络、模糊模型、ARIMA 模型等都提供在竞赛报告中^[2]。

在这些方法中, 神经网络具有非常好的非线性逼近能力, 能够逼近任何复杂的非线性系统, 具有良好的自学习能力。因此很适合于电力负荷预测问题, 是在国际上得到认可的实用预测方法之一^[3]。

然而, 由于神经网络预测模型的模型结构 (网络的层数和神经元的个数) 的选取大多凭借

收稿日期: 2007-03-30.

基金项目: 华北电力大学博士学位教师科研基金资助项目
(200512014).

经验, 需要使用者具有深入的神经网络方面的理论基础和应用经验, 因此限制了神经网络的应用。总体来看, 目前的研究达不到实用的程度^[3], 解决其结构选择和参数取值问题可以有效提高其实用性。

近年来大量的自组织神经网络算法被提出, 本文使用的在线自组织模糊神经网络 (SOFNN) 算法便是其一。算法由 Gang Leng^[4] 于 2004 年提出, 其主要优点在于: 能够自动地决定网络结构并且给出模型参数, 具有很好的预测精度。

本文将 SOFNN 算法应用于 EUNITE 竞赛数据预测。尽管竞赛已经结束了, 但仍然有很多基于这些数据的研究工作在开展, 并和优胜者的结果进行对比^[5~7]。同时, 以往的研究者已经取得了很多的经验和有价值的结果, 因此, 可以在采纳一些经验和结论的前提下, 开展进一步的研究, 将研究重点集中于算法应用方面。

1 竞赛任务描述

EUNITE 电力负荷预测竞赛的组织者为竞赛参与者提供了以下的数据^[2]:

- (1) 从 1997 年到 1998 年的每 30 min 一次的电力负荷数据记录。
- (2) 从 1995 年至 1998 年的每日平均气温。
- (3) 从 1997 年到 1999 年的节假日。

竞赛的任务是预测 1999 年 1 月的每日最大负荷, 评判标准主要包括: 平均绝对百分误差 (MAPE) 和最大误差值 (ME)。

平均绝对百分误差 (MAPE) 定义如式 (1):

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{L_i - \hat{L}_i}{L_i} \right| \times 100\%$$

$i = 1, 2, \dots, 31, n = 31$ (1)

最大误差 (ME) 定义如式 (2):

$$ME = \max |L_i - \hat{L}_i| \quad i = 1, 2, \dots, 31$$
(2)

式中: L_i 和 \hat{L}_i 分别为 1999 年 1 月第 i 天最大负荷的真实值和预测值, n 为 1999 年 1 月的天数, $n = 31$ 。

2 算法及应用基础

2.1 SOFNN 算法

SOFNN 的结构为 5 层模糊神经网络, 其学习过程包括结构学习和参数学习两部分。结构学习通过自组织方法得到经济型的网络规模; 参数

学习通过在线递推最小二乘算法使得网络快速收敛于一点。SOFNN 算法的最大优点在于可以自动识别所需要的神经元个数, 能够自动地决定神经模型的结构并得出模型的参数; 无需对神经网络和模糊系统有深入的理论知识, 无用像现有的许多神经网络和模糊学习方法那样需要进行反复试验, 因此更具有实用性和应用价值。算法的详细内容请参考文献 [4]。

2.2 一些重要的结论

本文的研究采纳了文献 [1] 中的一些结论。

(1) 日期属性是非常重要的特征。电力负荷按每周呈现周期性特征, 而且, 在周末时负荷要低于工作日。因此, 对星期日和节假日属性编码作为输入特征向量是非常重要的。

(2) 气温属性可以不用考虑。在气温和负荷之间有着一定的直接联系, 因此在大部分短期负荷预测 (STLF) 研究中, 使用了气温属性。然而, 在竞赛中, 1999 年 1 月份的实际气温并没有提供。也就是说, 如果要使用气温属性, 就要首先使用气温的历史数据来预测 1999 年 1 月份的气温值。实际上, 气温预测是比负荷预测更为复杂的问题, 而且依据如此有限的历史气温数据来预测气温, 比预测负荷本身更加困难, 所以, 在模型中不使用气温信息。

(3) 时间序列方案更好一些。过去的负荷数据会影响着未来的负荷值, 这是时间序列的概念, 应用这些历史负荷数据会对负荷预测有益。

预测使用滚动预测方式, 例如, 使用 1998 年 12 月 25 至 31 日的数据得到 1999 年 1 月 1 日的数据, 然后, 使用 12 月 26 至 31 日的数据和预测得到的 1999 年 1 月 1 日的数据, 预测得到 1999 年 1 月 2 日的数据, 连续进行这种滚动预测, 直到得到 1 月 31 日的值。

(4) 无监督数据分段技术非常有用。无监督数据分段技术对于时间序列预测非常有用, 通过该技术, 在不提供任何季节信息的情况下, 能将冬天的负荷数据和夏天的负荷数据很自然地分开, 这说明冬天和夏天的负荷数据具有不同的特征。依据分段的结果, 选择不同的训练样本, 与需要预测的时间段相对应, 预测的精度将会更好。

3 应用及结果

3.1 日最大负荷预测

(1) 训练数据集选择。根据无监督数据分段

的结果^[1]，选择 1997 年及 1998 年 1 月和 2 月的数据作为预测 1999 年 1 月负荷的训练集，构成新的时间序列，如图 1。

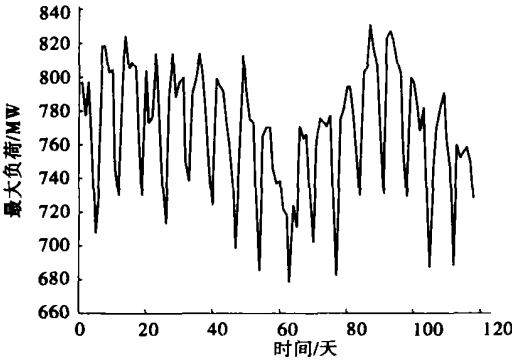


图 1 日负荷预测时训练数据

Fig. 1 Training data for day load forecasting

(2) 时间序列模型阶数的选择。电力负荷属于随机过程，在随机过程研究中，自相关系数能够显示随机过程是否平稳以及选择合适的模型阶数^[8]。对训练数据集进行样本自相关系数计算，如图 2。从样本峰值和峰谷数值得到阶数为 7，也就是选取前 7 天的历史数据作为特征向量。

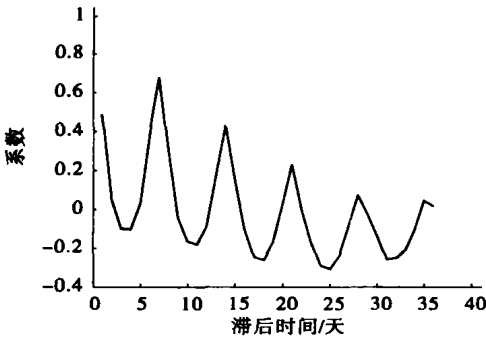


图 2 训练数据的自相关系数

Fig. 2 Autocorrelation coefficients of training data

(3) 特征编码。在选择有用的信息和合适的分段数据之后，对输入特征向量进行编码如下：

[星期属性 历史负荷]

星期属性使用 7 位 0/1 编码来实现，代表预测日为星期几，如 [1 0 0 0 0 0] 代表星期一，[0 1 0 0 0 0] 代表星期二，依次类推。历史负荷数据使用归一化数值表示过去 7 天的负荷数据。训练输出数据为预测日的最大负荷。

(4) 训练及预测结果。训练结果显示于表 1 及图 3~6 中，训练数据为归一化值。图 3 显示训练过程中误差的变化，其中虚线表示误差限制值，图 4 表示结构自组织过程中神经元个数的变化。

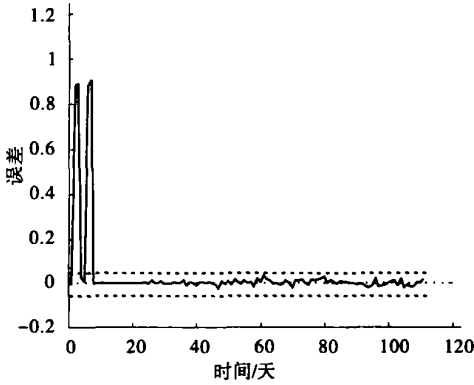


图 3 训练过程中的误差变化

Fig. 3 Errors during training

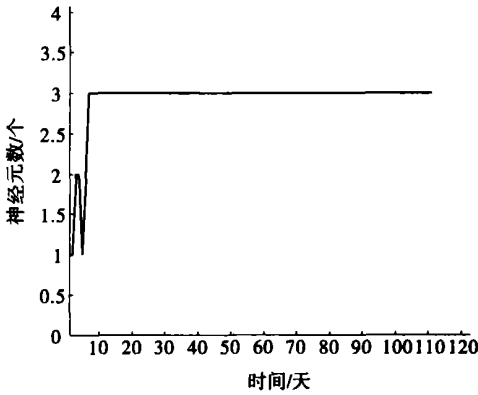


图 4 训练过程中神经元个数变化

Fig. 4 Growth of neurons during training

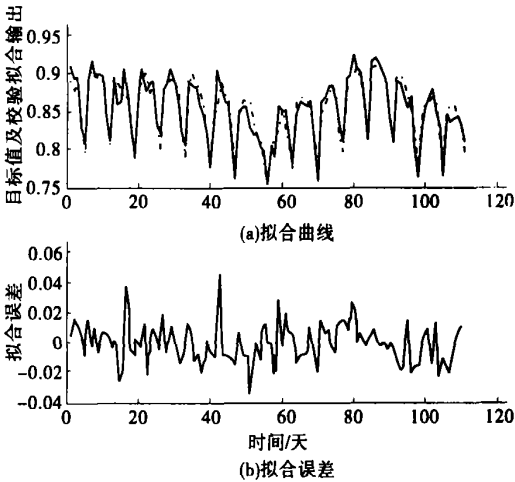


图 5 训练结果及误差

Fig. 5 Result and error of training

训练完成后，建立了一个 3—EBF—神经网络，使用归一化后的训练数据基于该模型进行数据拟合校验，图 5 为校验曲线及误差，其中实线为实际值，虚线为预测值。之后使用该模型进行日最大负荷预测，得到预测结果如图 6，图中

实线为实际值, 虚线为预测值。

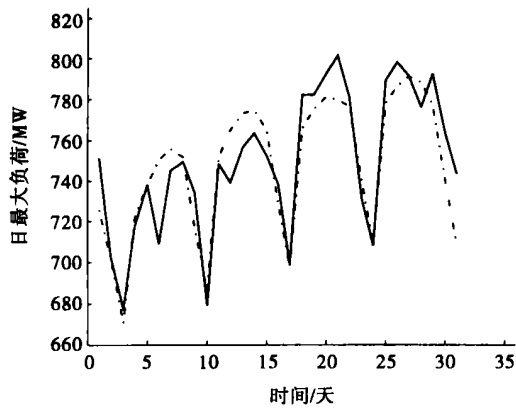


图 6 日最大负荷预测结果

Fig. 6 Result of forecasting

预测的性能指标如表 1。

表 1 SOFNN 训练和预测精度

Tab. 1 Accuracy of training and forecasting

Method	数据集	MAPE/ (%)	ME
SOFNN	训练数据集	1.33	44.13
	预测数据集	1.78	50.04
EUNITE 优胜者算法 winner	预测数据集	1.95	50—60*

*报告中 ME 值没有准确提供, 但可以从报告中估计得出约 50~60。完整的 EUNITE 网络竞赛原始数据可从 EUNITE 网站获得 (<http://neuron.tuke.sk/competition/index.php>)。

EUNITE 竞赛中最好的结果为: MAPE 值 1.95%^[2]。因此, 从表 1 结果上看出, 本文的方法能够得到更好的预测结果。

3.2 周平均最大负荷预测

在应用 SOFNN 进行日最大负荷预测并取得较好结果后, 又进行了进一步的修正。基本思路是: 由于中期负荷是基于预测值滚动预测的, 因此预测误差会不断累积, 另一方面, 日负荷预测时历史数据采用前 7 天数据, 预测结果按每周特征循环, 不包含长期趋势, 因此提出使用周平均最大负荷数据预测值进行修正的方法。周负荷预测的滚动预测次数少, 因而累积误差小, 同时能够体现长期特征, 从而弥补日负荷预测值的不足。

(1) 训练数据选择。依据分段结果, 选择冬天数据作为训练数据, 包括每年 1 月到 3 月以及 10 月到 12 月的样本, 一方面冬天数据与需要预测的 1 月份数据有相同的特征模型^[1], 另一方

面, 完整的冬天数据, 能够提供更多的训练样本和长期特征。提取这些样本后, 对每一周最大负荷值进行平均值计算, 得到周平均最大负荷时间序列。

(2) 模型阶数选取。得到周平均最大负荷时间序列之后, 依据自相关系数值, 得到阶数为 26, 即使用过去 26 周的历史数据预测当前周的值。

(3) 选择完数据之后, 进行编码如下:

[周数, 历史周平均负荷值]

周数使用 5 位 0/1 编码来表示, 代表该周在一年中的第几周, 如第 1 周为 [0 0 0 0 1], 第 10 周为 [0 1 0 1 0]。历史周平均负荷由过去 26 周负荷值表示。

建立模型后进行预测, 得到预测值为 [716.2, 739.7, 757.7, 781.3], 图 7 中虚线所示, 实际的周平均值 [720.1, 738.2, 763.7, 767.7], 图 7 中实线所示。

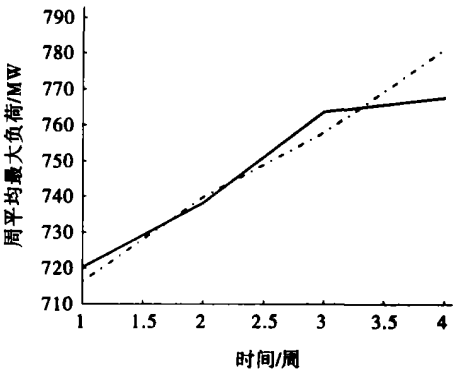


图 7 周平均最大负荷预测结果

Fig. 7 Forecasting result of week average max-load

3.3 使用周平均最大负荷修正日负荷

(1) 首先计算基于 SOFNN 进行预测得到的日负荷预测值的每周平均值, 如第 j 周为: WL_j 。

(2) 使用基于 SOFNN 进行周负荷预测得到的每周平均值 WL_j' 减去 WL_j , 得到每周的修正值 $\text{delta } WL_j$, 即 $\text{delta } WL_j = WL_j' - WL_j$ 。

(3) 在第 j 周每日最大负荷预测值 Y_{ij} 上都增加 $k \times \text{delta } LW_j$, $Y_{ij}' = Y_{ij} + k \times \text{delta } LW_j$, 得到最终修正值 Y_{ij}' 。其中 k 为修正系数, 取 $k = 0.5$, 相当于认为周负荷预测和日负荷预测的可信度一致。

最后, 得到修正预测结果如图 8, 图中实线为实际负荷值, 虚线为原算法预测值, 星号为修

正后的预测值。

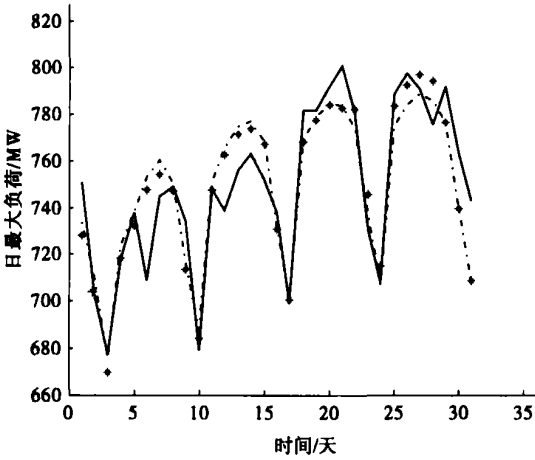


图 8 修正后的日最大负荷预测结果
Fig. 8 Forecasting results after revised

表 2 修正后 SOFNN 预测精度
Tab. 2 Accuracy of forecasting

Method	MAPE/ (%)	ME
原 SOFNN 算法	1.78	50.04
周负荷修正算法	1.59	41.95
EUNITE 优胜者算法	1.95	50—60

从图 8 及表 2 中可以看出，周负荷具有更低的累积误差，同时弥补日负荷预测长期特征不足的缺点，修正之后，MAPE 和 ME 指标更好，因此，修正是有效的。

4 结 论

在进行中期负荷预测的应用研究中，使用不同的训练集、模型阶数和特征编码对 1999 年 1 月进行日最大负荷预测和周平均最大负荷预测，结果显示基于 SOFNN 算法能够取得比竞赛最佳结果更好的预测值。之后用周平均最大负荷预测值修正日负荷预测值，以改善预测性能，结果表明经过修正后，预测精度得到了进一步的提高。

参考文献:

[1] Chen B J, Chang M W, Lin C J. Load forecasting using support vector machines: a study on EUNITE competition 2001 [J]. IEEE transactions on power systems, 2004. 19 (4): 1821—1830.

[2] Company behind East—Slovakia Power Distribution Company. World wide competition within the EUNITE network, EUNITE competition report [R].

[3] 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.

[4] Leng G, Prasad G, McGinnity T M. An on-line algorithm for creating self-organizing fuzzy neural networks. Neural Networks, 2004, (17): 1477—1493.

[5] Ortiz-Arroyo D, Skov M K, Huynh Q. Accurate Electricity Load Forecasting with Artificial Neural Networks [C]. Proceedings of the 2005 International Conference on Computational Intelligence for Modelling, Control and Automation, and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC' 05), 2005.

[6] Tao X. Input dimension reduction for load forecasting based on support vector machines [C]. Hong Kong: 2004 IEEE International Conference of Electric Utility deregulation, restructuring and power technologies, 2004.

[7] Hsu C C. Dynamically Optimizing Parameters in Support Vector Regression: An Application of Electricity Load Forecasting [C]. Hawaii: Proceedings of the 39th Hawaii International Conference on System Sciences, 2006.

[8] Pankratz A. Forecasting with Univariate Box-Jenkins Models [M]. John Wiley & Sons, 1983.

作者简介: 李炎 (1972—), 女, 华北电力大学工程师, 主要研究方向为模式识别与智能系统。