经验模态分析综合法在负荷预测中的应用

余 林，舒 勤，马 哲
(四川大学电气信息学院，四川 成都 610065)

摘 要: 针对电力负荷的周期相似性，提出一种基于经验模态分析法的综合负荷预测方法。先对原始数据进行统计，使用 EMD 将统计的时间序列分解为有限个固有模态分量，对固有模态分量使用模糊 C 均值聚类，再采用 ARMA 将聚类后的固有模态分量预测，最后把每个分量预测值求和得到负荷预测值。实例仿真计算表明，该算法比直接使用 ARMA 模型进行预测具有更高的预测精度，是一种有效短时预测方法。

关键词: 负荷预测; 经验模态分解; 自回归滑动平均; 聚类

Abstract: According to the similarity of power load，an integrated load forecasting method based on empirical mode decomposition (EMD) is proposed. Firstly，an artificial statistical is done for the raw data，and the statistical time series is decomposed into different intrinsic modes by EMD，then the intrinsic mode components are clustered by fuzzy clustering. Then，these different clustered components are predicted by aoturegressive moving average (ARMA) model. Finally，the forecasted load is obtained by adding together the predicted values of each component. The experiment simulations show that the proposed algorithm has a higher forecasting accuracy than the direct use of ARMA model，which is an effective short – term forecasting method.

Key words: load forecasting; empirical mode decomposition; aoturegressive moving average; cluster
中图分类号: TM714 文献标志码: A 文章编号: 1003 – 6954(2015) 02 – 0040 – 05
DOI:10.16527/j.cnki.cn51-1315/tm.2015.02.010

0 引 言

电力系统短期负荷预测对电力系统控制、电网安全、经济运行、系统规划、优化调度及电能质量等方面起着十分重要的作用，是能量管理系统的重要组成部分[1]。随着电力生产和消费日益市场化，准确的进行电力系统短期负荷预测可合理地规划地市电网用电的供需平衡，有效地增强电网运行的安全性及可靠性，提高电力企业的经济效益和社会效益。因此，短期负荷预测已成为电力系统中一个重要研究领域。

短期负荷预测的关键问题是如何提高预测日期的精度。国内外许多专家和学者在此方面做了大量的研究工作，提出了很多预测方法[2]，可分为传统预测方法和人工智能方法两大方面。传统预测方法包括回归分析法、时间序列法等，回归分析法是根据历史数据变化规律及影响负荷变化因素建立回归方程，确定参数来预测负荷值，该方法对历史数据要求高，结构过于简单，精度不高; 时间序列法是按一定时间间隔对电力负荷进行采样得到的时间序列，建立模型，虽然该算法较为成熟，但只适用于负荷变化较为均匀的短期负荷，且没有考虑天气、节假日等因素。人工智能方法包括专家系统[5]、模糊逻辑方法[6]和人工神经网络方法[7]等。专家系统是基于知识的程序设计方法建立计算机系统，根据专家知识和经验丰富系统并预测，其只适用于单一系统，且容易出现人为错误; 模糊逻辑法是建立在模糊数学理论上的新技术，但在实际应用中较困难; 人工神经网络具有学习及自适应能力，很适合电力短期负荷预测，但其存在收敛速度慢、泛化能力差等不足。近年，支持向量机(SVM)[8]是建立在统计学习理论上的一种新方法，其提高了泛化能力，具有预测能力强、收敛速度快等特点，但依然存在核函数和参数的最优选择等问题。

这里首先对一定时间间隔采集到的时间序列电力负荷数据对其进行曲线拟合，之后利用采用经验模态分解(empirical mode decomposition，EMD) 对负荷数据进行分解。对提取的分量聚类处理，再对聚类后的分量分别进行自回归滑动平均(aoturegressive moving average, ARMA) 预测，最终得到预测日一天的负荷。通过试验验证了该方法的有效性。

1 经验模态分解(EMD)

由于电力负荷模型是一个受到多种不确定因素影响的非线性大系统，其系统的高度复杂性和不确定性，导致数据成分复杂且不满足广义平稳信号的条件：①其均值为常数，与时间变量无关；②自相关函数仅为时间差的函数；因此是一种典型的非平稳信号。其信号成分的不同代表不同的特征，确定的成分代表负荷的主要趋势，非确定的成分代表着负荷在主要趋势附近的波动。EMD 方法对非平稳性信号有较大的优越性与适应性，已经成为处理非线性、非平稳信号的一种有效工具[9-12]。经验模态分解可以将原始信号分解为一系列固有模态分量(intrinsic mode function, IMF)。EMD 分解能够起到放大关键部位数据的效果，经过分解信号得到的各个固有模态分量放大了数据的局部特征，对它们进行分析，可以更准确有效地把握原数据的特征信息。因此对电力负荷的统计数据做 EMD 分解可以有效、准确得到信号的瞬时特征，通过对 IMF 分量的分析能够更清楚地发现负荷的本质和变化趋势。EMD 分解后的每个 IMF 需要满足两个条件：①在整个数据长度上，必须有相同数目的极值点和过零点，或者二者之差不能多于一个；②在任意时刻，上包络线(由局部极大值点形成)与下包络线(由局部极小值点形成)的均值为零；经过 EMD 分解得到的各个 IMF 分量从高频到低频依次分布，每个 IMF 分量的频率成分与信号自身变化、采样频率等因素有关，因此该方法具有自适应性，适合应用在复杂信号的研究中。EMD 的本质就是一个筛分过程，设定一个信号由多个 IMF 组成，步骤如下[13]。

(1) 确定信号 $x(t)$ 的所有局部极值点，并求出上下包络线。所有原信号数据都应该在这两条包络线内；

(2) 将这两条包络线的均值记为 $m(t)$，并求出 $y_1(t) = x(t) - m(t)$；

(3) 判断 $y_1(t)$ 是否为 IMF，若 $y_1(t)$ 不满足 IMF 条件，则将 $y_1(t)$ 视为新的 $x(t)$，重复以上步骤，直到 $y_1(t)$ 满足 IMF 条件，此时，记 $y_1(t) = c_1(t)$，则 $c_1(t)$ 为信号 $x(t)$ 的第一个 IMF 分量，它代表信号 $x(t)$ 中最高频率的分量；(4) 将 $c_1(t)$ 从 $x(t)$ 中分离出来，即得到一个去掉高频分量的差值信号 $r_1(t)$，即有 $r_1(t) = x(t) - c_1(t)$；

(5) 重复以上步骤，得到 $c_2(t) \cdots c_n(t)$，直到 $c_n(t)$ 或者 $r_n(t)$ 满足给定的终止条件(通常使 $r_n(t)$ 成为一个单调函数)时终止筛选，由此可得 $x(t)$ 的分解式为

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t)$$

式中，$r_n(t)$ 为残余函数，代表信号的平均。而各 IMF 分量 $c_1(t)$，$c_2(t)$，…，$c_n(t)$ 代表了信号的高频分量到低频分量的变化，各分量频率由高频到低频依次分布。

2 模糊 C 均值聚类(FCM)

模糊 C 均值聚类算法(fuzzy c-means algorithm, FCM)[14]的基本思想为：将数据集 $X = \{x_1, x_2, \cdots, x_n\} \in R$ 划分为 c 类，$u_{ik} \in R$ 为数据集 X 中任意数据对 i 类的隶属度，模糊隶属度矩 $U = \{u_{ik}\} \in R$ 来表示分类结果，隶属度矩阵满足以下条件。

$$\begin{gathered} u_{ik} \in [0,1], \forall i,k \\ 0 < \sum_k u_{ik} < n, \forall i \\ \sum_k u_{ik} = 1, \forall K \end{gathered} \tag{1}$$

模糊 C 均值聚类是通过迭代的方法不断更新隶属度矩阵 U 和聚类中心 V 以达到目标函数 $J_m(U,V)$ 收敛来实现的，目标函数如式(2)。

$$J_m(U,V) = \sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ik})^m d_{ik}^2(x^k, v^i) \tag{2}$$

其中，$U = \{u_{ik}\}$ 为满足条件式(2)的隶属度矩阵；$V = \{v_1, v_2, \cdots, v_c\} \in R$ 为聚类中心点集；m 为加权指数，其范围为 $[1, \infty)$，一般情况下 $m = 2$ 是较理想的取值。第 k 个样本到第 i 类中心点的距离定义为式(3)。

$$d_{ik}^2(x_k, v_j) = \| x_k - v_i \|^2_A = (x_k - v_i)^T A (x_k - v_i) \tag{3}$$

A 为 $p \times p$ 的正定矩阵，当 $A = 1$ 时，d_{ik}^2 为欧氏距离。模糊 C 均值聚类就是通过反复迭代更新隶属度和中心来对目标函数式(2)进行优化，具体步骤如下。

(1) 初始化聚类中心 $V = \{v_1, v_2, \cdots, v_c\}$；

(2) 计算隶属度矩阵为

$$u_{ik} = \left[\sum_{j=1}^{c}\left[\frac{d_{ik}(x_k, v_i)}{d_{jk}(x_k, v_j)}\right]^{2/(m-1)}\right]^{-1} \tag{4}$$

(3)更新聚类中心为

$$v_i=\frac{\sum_{k=1}^{n}(u_{ik})^m x_k}{\sum_{k=1}^{n}(u_{ik})^m}\quad i=1,2,\cdots,c \tag{5}$$

(4)重复步骤(2)、(3)直到式(2)收敛。

3 时间序列方法模型

电力负荷的历史数据是按一定时间间隔进行采样记录下来的有序集合,因此它是一个时间序列,电力负荷时间序列预测方法就是根据负荷的历史资料设法建立一个时间序列的数学模型,用这个模型一方面来描述电力负荷时间序列变化的规律,另一方面在该模型的基础上建立负荷预测的数学表达式,对未来的负荷进行预测。传统的时间序列模型[15,16]有以下几种。

3.1 自回归模型(AR模型)

在p阶的AR模型中,时间序列的当前值表示为序列中过去的p个值与随机噪声的线性组合。

$$y_1=\varphi_1 y_{t-1}+\cdots\cdots+\varphi_p y_{t-p}+a_t \tag{6}$$

式中,y_t为第t时刻的最大或最小负荷;$y_{t-i}(i=1,2,\cdots,p)$为$t-i$时刻的最大或最小负荷;a_t为白噪声。

设B为时间后移算子

$$Bz_t=z_{t-1},B^m z_t=z_{t-m}$$

$$\varphi(B)=1-\varphi_1 B-\varphi_2 B^2-\cdots\cdots\varphi_p B^p \tag{7}$$

则式(6)可以改写为

$$\varphi(B)y_t=a_t \tag{8}$$

自回归过程可能是平稳的或是非平稳的。平稳的条件是:$\varphi(B)=1-\varphi_1 B-\varphi_2 B^2-\cdots\cdots\varphi_p B^p$看作为$B$的$p$次多项式,$\varphi(B)$所有根的绝对值都必须大于1,即所有根都在单位圆以外。

3.2 滑动平均模型(MA模型)

在q阶MA模型中,时间序列的当前值表示为过去的q个随机噪声的线性组合为

$$y_t=a_t-\theta_1 a_{t-1}\cdots\cdots\theta_q a_{t-q} \tag{9}$$

且乘在a上的权数$1,\theta_1,\theta_2,\ldots\ldots,\theta_q$不必总和为1,也不必是正数。令

$$\theta(B)=1-\theta_1 B-\theta_2 B^2-\cdots\cdots\theta_q B^q \tag{10}$$

则式(9)可写为

$$y_t=\theta(B)a_t \tag{11}$$

由于$\theta(B)=1-\theta_1 B-\theta_2 B^2-\cdots\cdots\theta_q B^q$为有限$q$项,所以$\theta(B)$无条件收敛,即该过程是无条件平稳。

3.3 自回归滑动平均模型(ARMA模型)

在实际时间序列的拟合中,将自回归和滑动平均两个模型组合起来一同纳入一新模型,可得到更一般的(p,q)阶自回归滑动平均模型为

$$y_t=\varphi_1 y_{t-1}+\cdots\cdots+\varphi_p y_{t-p}+a_t-\theta_1 a_{t-1}\cdots\cdots\theta_q a_{t-q}$$

即 $\varphi(B)y_t=\theta(B)a_1$ (12)

ARMA模型的阶数主要由准则函数进行判定,选取合适的阶数使得模型对原始数据拟合程度最高,即在此时准则函数达到最小值。实际应用中ARMA模型的阶数通常都比较低,因此可以从低到高逐渐选取p值和q值,并判断每种取值情况下的准则函数值是否达到极小值。常用的准则函数有AIC准则、BIC准则、FPE准则等。ARMA模型具有随机差分方程的形式,利用它不仅能揭示负荷时序数据本身的结构与规律,也可以定量地了解观测数据之间的线性相关特性,预测其未来值的大小。

4 EMD与ARMA负荷预测模型算法

对电力负荷信号进行EMD分解,形成若干个固有模态分量,去除其中干扰分量,然后对分量进行模糊C聚类,最后对分类后的分量建立ARMA预测模型来进行负荷预测。具体算法如下。

(1)应用第1节中的EMD分解步骤进行负荷数据分解;

(2)分析分解后的分量,去除其中的随机干扰分量;

(3)对IMF分量进行模糊C均值聚类,设定聚类个数为n;

(4)对聚类后的分量无量纲化处理;

(5)对分量建立ARMA模型进行预测;

(6)将预测结果进行反无量纲化处理;

(7)将全部的预测数据结果累加得到负荷的预测结果。

5 实例仿真以及结果分析

采用某地市地区的负荷数据,每15 min采集一次数据,从零点15分开始到第2天零点,一天共96个负荷值。选择历史负荷的天数为21天,来预测第22天的负荷,即根据连续21天的共2 016个负荷值

数据来预测下一天的 96 个负荷数据。21 天的负荷数据如图 1 所示。

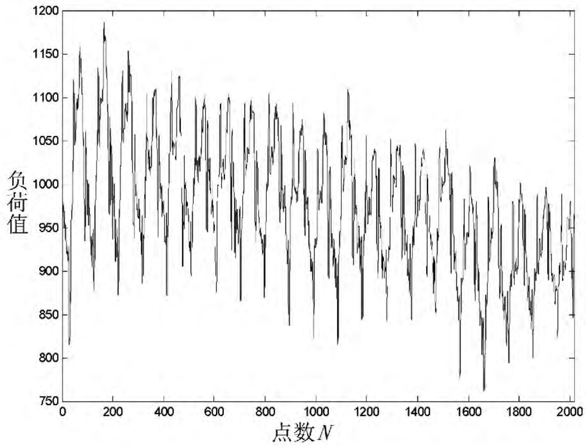

图 1　21 天电力负荷数据

5.1　EMD 分解

利用所提方法，对要预测日前面 21 天的数据进行 EMD 分解，分解的结果如图 2 所示，共有 12 个 IMF，从上至下分别为 IMF1 到 IMF12，第一个模态分量主要为随机干扰，在预测中去除，最后一个为残余量，表示平均趋势。由于电力负荷数据存在周期性的相似性质成分，故将分解后的 IMF 采用模糊 C 均值聚类，将聚类数设定为 6 类，结果如图 3 所示。

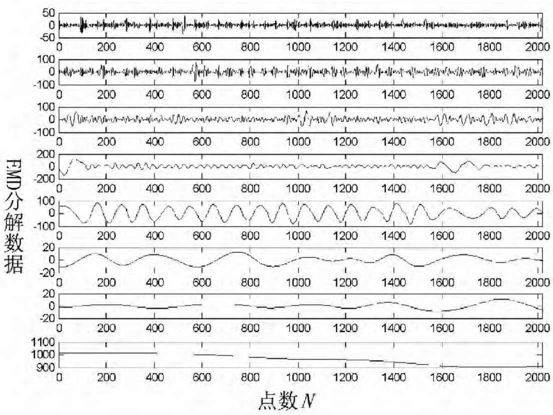

图 2　电力负荷 EMD 分解

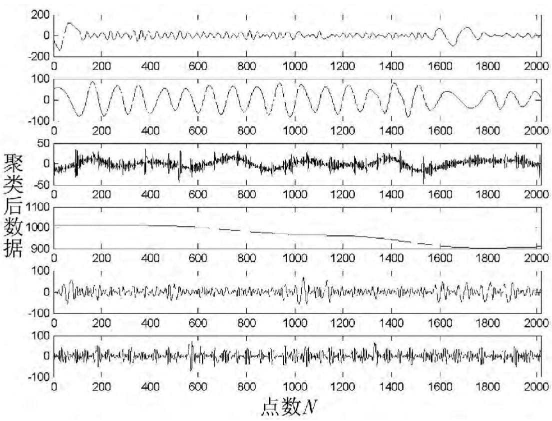

图 3　经过模糊聚类后的分量

5.2　IMF 的预测

利用 ARMA 模型对聚类后的分量分别进行预测，将预测结果累加得到负荷值。前面采用一步预测，对每一个分量采用 AIC 准则，使得信息量达到最小来确定 ARMA 模型的阶数，具体的分类如表 1 所示。取 2013 年 1 月 3 日至 1 月 24 日的电力负荷数据作为训练样本数据，以 2013 年 1 月 25 日的负荷数据作为预测样本数据，对 2013 年 1 月 25 日的负荷数据进行预测。利用 Matlab 实现前面所提方法预测，结果如图 4 所示。

表 1　聚类后每个 ARMA 模型阶数(p,q)

类别数	n_1	n_2	n_3	n_4	n_5	n_6
P 值	5	4	2	4	2	1
Q 值	6	4	4	4	3	5

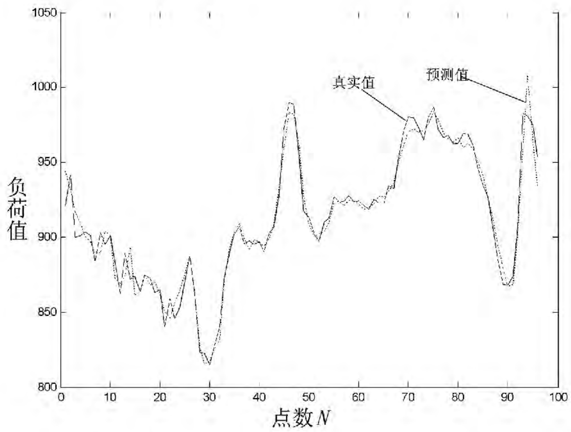

图 4　EMD－ARMA 预测与实际值的比较

为了比较所提方法的优越性，将其用直接的 ARMA 进行预测的方法进行比较。直接使用 ARMA 方法对样本进行训练后，预测的负荷值与实际值比较结果如图 5 所示。

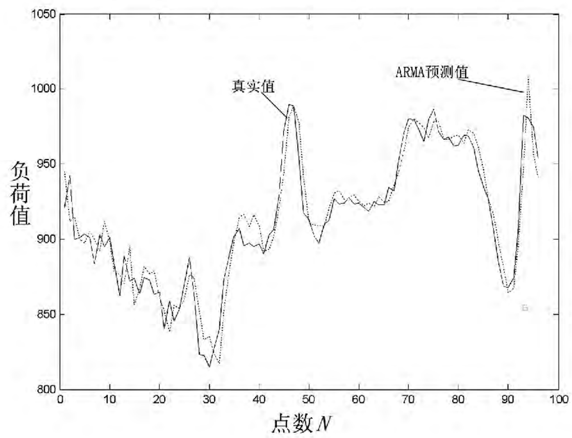

图 5　ARMA 预测与实际值的比较

为更加可靠、方便对比，引入如下为一些误差的计算公式。

1) 平均绝对误差

$$MAE = \frac{1}{N}\sum_{i=1}^{N}(|Y_1 - \hat{Y}_i|) \tag{13}$$

2) 均方误差

$$MSE = \frac{1}{N}\sqrt{\sum_{i=1}^{N}|Y_i - \hat{Y}|^2} \tag{14}$$

3) 平均绝对百分比误差

$$MAPE = \frac{1}{N}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| \tag{15}$$

其中，Y_i 表示的是第 i 个负荷实际值；$\hat{Y}$ 表示的是第 i 个负荷预测值；N 代表的是需要预测的负荷个数，这里是预测一天96个点的负荷量，所以 N 为96。

表2中给出了两种方法的平均绝对误差，均方误差和平均绝对百分比误差

表2 两种方法的平均绝对误差、均方误差和平均绝对百分比误差的对比

	平均绝对误差	均方误差	平均绝对百分比误差
时间序列法	11.623 1	1.581 3	0.012 8
所提方法	6.223 3	0.842 8	0.011 2

从表中可以验证所提方法的有效性，误差小于直接用ARMA方法，具有更高的预测精度。

6 结　语

提出了一种基于短时负荷的预测方法，采用EMD分解和时间序列模型进行负荷预测。经验模态分解方法将负荷信号序列分解成多个本征模态分量，把分解后的每个分量进行模糊聚类，再就聚类后的分量构造时间序列模型进行预测，然后将各个子模型预测值相加获得电力负荷预测值。从具体的地市地区的负荷数据进行验证，预测下一天的负荷量。仿真结果和预测误差表明，该方法性能较好，优于一般的时间序列预测方法，可以有效地对负荷数据进行短时预测。

参考文献

[1] 康重庆，夏清，刘梅. 电力系统负荷预测[M]. 北京：中国电力出版社，2007.

[2] 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.

[3] 雷绍兰，孙才新，周湶，等. 电力短期负荷的多变量时间序列线性回归预测方法研究[J]. 中国电机工程学报，2006，26(2)：25－29.

[4] 万昆，柳瑞禹. 区间时间序列向量自回归模型在短期电力负荷预测中的应用[M]. 电网技术，2012，36(11)：77－81.

[5] Ho Ku－Long，Hsu Yuan－Yih，Lee C E，et al. Short－term Load Forecasting of Taiwan Power System Using a Knowledge－based Expert System[J]. IEEE PWRS，1990，5(4)：46－57.

[6] Chow Mo－yuen，Zhu Jin－xiang，Hahn Tram. Applications of Fuzzy Multi－objective Decision Making in Spatial Load Forecasting[J]. IEEE Trans. on Power Systems，1998，13(3)：1185－1190.

[7] Park D C，EI－SharKawi M. A，Marks J，et al. Electric Load Forecasting Using a Neural Network[J]. IEEE Trans on Power System，1991，6(2)：442－449.

[8] 赵登福，王蒙，张讲社，等. 基于支撑向量机的短期负荷预测[J]. 中国电机工程学报，2002，22(4)：26－30.

[9] 杨宇，于德介，程军圣. 基于EMD与神经网络的滚动轴故障诊断方法[J]. 振动与冲击，2005，24(1)：85－86.

[10] 陈小强，王琦，汪同庆，等. 基于EMD的轻轨锚固螺杆故障诊断方法研究[J]. 机械强度，2009，31(4)：548－552.

[11] Huang N E，Shen Z，Long S R，et al. The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non－stationary Time Series Analysis[J]. Proc R Soc Lond A，1998，454：903－905.

[12] 刘霖雯，刘超，江成顺. EMD新算法及其应用[J]. 系统仿真学报，2007，19(2)：446－447.

[13] 罗向龙，牛国宏，潘若禹. 交通流量经验模态分解与神经网络短时预测方法[J]. 计算机工程与应用，2010，46(26)：212－214.

[14] 杜海顺，汪凤泉. 一种快速的模糊C均值聚类彩色图像分割方法[J]. 计算机工程与应用，2009，45(33)：138－140.

[15] 张冉，赵成龙. ARIMA模型在网络流量预测中的应用研究[J]. 计算机仿真，2011，28(2)：171－174.

[16] 叶瑰昀，罗耀华，刘勇. 基于ARMA的电力系统负荷预测方法研究[J]. 信息技术，2002(6)：74－76.

作者简介：

余　林(1988)，硕士，研究方向为电力系统信号检测与处理、现代信号处理；

舒　勤(1958)，教授，博士，研究方向为现代信号处理、智能电网；

马　哲(1989)，硕士，研究方向为电力系统信号检测与处理、现代信号处理。

(收稿日期：2014－10－25)