

# 基于模糊聚类与函数小波核回归的短期负荷预测方法

祖向荣,田 敏,白 焰

(华北电力大学 控制与计算机工程学院,北京 102206)

**摘要:** 短期负荷预测日益成为智能电网的重要课题。将历史日负荷序列表示成等负荷段组成的子序列集合,基于模式相似性方法,采用模糊聚类与函数型小波核非参数回归(FWKNR)相结合的组合预测算法。FWKNR将预测日负荷表示成历史日相应负荷段的加权平均值,将与参考段更相似的段赋予更高权重,并基于离散小波变换的形状相似性度量,采用N-WE计算权重;由预测日各分段预测快速完成日预测。模糊聚类针对单一用户历史负荷进行典型负荷模式的分类预处理,并识别与预测日有更相似行为模式约减的有效训练样本集合参与模型预测。基于某地区实际负荷数据,实验比较分析验证了组合算法的优越性。

**关键词:** 短期负荷预测; 模式相似性方法; 模糊聚类; 非参数小波核回归; 非线性时间序列; 智能电网

**中图分类号:** TM 715

**文献标识码:** A

**DOI:** 10.16081/j.issn.1006-6047.2016.10.020

## 0 引言

短期负荷预测 STLF(Short-Term Load Forecasting)主要用于预测未来几小时、一天甚至几天的负荷,STLF的精度与速度直接影响电力系统运行的安全性、经济性与供电质量。智能电网新能源高度渗入与电动汽车新消费模式等技术创新,赋予了 STLF 崭新的研究内容与挑战<sup>[1-3]</sup>,对负荷预测的准确性、实时性、可靠性与智能性提出更高的要求。在用户侧智能电表将逐步按细粒度间隔(如 15 min)采集海量用户负荷数据,为智能用电细粒度数据分析与电力运行精细化管理提供基础。如何利用海量智能电表数据建立可靠的预测模型、提出高效的预测方法是 STLF 的重要课题<sup>[4-6]</sup>。

电力负荷具有非线性、时变性和不确定性的特点,具有复杂的非线性特征,现代负荷预测的一个基本问题是对非线性动态复杂系统的探索。STLF 的复杂性体现在:负荷时间序列是一个非平稳的随机过程,呈现多种趋势分量、多重周期变化(日,周,年)与随机噪声,同时,负荷受多种外部变量,诸如气象、温度、时间、人口、经济、电力价格、地理环境、消费类型与习惯等的影响,其中气象与温度是重要外部变量影响因素,使负荷序列呈现典型的非线性时间序列特征。STLF 作用的大小主要取决于预测精度,如何提高预测精度是目前研究 STLF 理论与方法的重点与难点。

电力负荷预测是智能电网的重要课题,作为一个重要分支,以神经网络为代表的人工智能与机器学习方法得到了不断的发展与应用。文献[7]综述了当前负荷预测的主要方法与发展方向。本文研究方向则为非线性时间序列分析方法,故重点沿着其 STLF 应用进

行了简要综述,篇幅所限,本文重点研究基于核方法的非参数模型。近年来非线性时间序列 STLF 研究获得广泛重视,并呈现与多种高级人工智能算法相融合的趋势。文献[8]研究了多变量时间序列 STLF 方法,根据单变量时间序列的延时重构对由历史负荷序列及其相关因素序列所构成的多变量时间序列进行相空间重构与计算;文献[9]考虑气温等外界因素对负荷的非线性影响,组合模糊逻辑与时间序列算法,构造具有非线性特性的传递函数模型,跟踪负荷突变趋势进行预测;文献[10-13]针对目前负荷特性的复杂化,使负荷序列呈现更为显著的非线性波动特征,认为波动不对称性各种特征的刻画是认识与把握负荷时间序列的重要环节,文献对重要的波动性模型之一,即 GAR-CH(Generalized Auto Regressive Conditional Heteroskedasticity model)族模型进行了深入研究,提出相关 STLF 算法,并给予详细实验验证;文献[14-15]研究了基于混沌时间序列的预测方法与应用技术。

本文采用了模糊聚类与基于核方法的非参数模型最新成果,以模式相似性方法 PSM(Pattern Similarity-based Method)为理论基础,PSM 以最小距离方法为通用原则,形成了机器学习与模式识别若干方法。文献[16]分析了各种基于 PSM 的 STLF 预测模型 PSBFMs(Pattern Similarity-Based Forecasting Models),包括基于核估计的模型 KEM(Kernel Estimation-based Model)、最近邻估计的模型(nearest neighbor estimation-based model),以及基于模式聚类的模型(pattern clustering-based models)。其中有传统聚类分析方法与新的人工免疫系统(artificial immune systems),其公共特征是直接从数据中学习,并利用负荷序列周期变化中呈现的模式相似性构造模型。PSBFMs 主要优点是其简单性:具有清晰的结构和易于理解的操作原则、低数量的参数,以及快速的优化与学习过程。文献基于各种 PSBFMs 方法的实验比较验证,结论是

收稿日期:2016-01-18;修回日期:2016-08-15

基金项目:国家自然科学基金资助项目(61571189)

Project supported by the National Natural Science Foundation of China(61571189)

基于核方法 KM(Kernel Method)的 N-WE(Nadaraya-Watson Estimator)权估计方法与模糊邻域模型 FNM(Fuzzy Neighborhood Models)预测性能最佳,精度很高,且参数少,使得运行时复杂度最低。这些新的方法基于模式相似性与局部参数回归,简化了预测问题,能够开发高效的预测模型,是富有前景的应对多周期变化时间序列预测问题的方法。本文研究的基础模型即为基于核方法(采用 N-WE 权估计)的函数型非参数核回归方法 FNKRM(Functional Nonparametric Kernel Regression Method)。

非参数核回归方法是近几年兴起的一种适合不确定性的、非线性的动态系统的建模方法,在电力预测中有诸多应用。文献[17]较早通过结合温度的周负荷预测实验,论证了 FNKRM 与人工神经网络 ANN(Artificial Neural Network)方法有非常相近的性能,且 FNKRM 模型简单可靠,可直接由历史数据驱动计算,模型参数少,且由交叉验证容易获得。文献[18]分析了多种定性与定量外部环境变量对 STLF 精度的影响,针对重要外部变量温度如何在 FNKRM 进行表达的问题,通过由历史负荷与温度数据构造最接近预测日负荷行为特征的虚拟基准参考段(virtual reference segment)于模型的相似性度量,从而在模型中合理考虑了温度的影响,提高预测精度,与多种算法对比实验,证明其具有更优越的性能,也证明了 FNKRM 的易于扩展性。文献[19-20]在中长期日负荷预测、短期风电功率预测方面应用非参数回归取得了良好的预测效果。

聚类分析是自动发现具有相似行为序列的重要方法,其既可作为发现相似用电模式、特殊用电模式、异常数据识别与修正等相关分析工具,也为进一步分析提供预处理过程,作为重要的可视化分类方法,在基本保留重要信息的条件下降低数据处理量,从中提取有效样本,并进行模式识别(pattern identification),亦是电力海量数据处理的重要工具之一。文献[4-6, 21-23]将聚类分析分别与不同的负荷预测模型相结合,显著提高了预测效率与精度。模糊聚类分析是根据客观事物间的不同特征、亲疏程度与相似性等关系,通过建立模糊相似关系对客观事物进行分类的数学方法,由于能很好地符合电力负荷数据的非线性特征,外部环境变量对负荷曲线影响的定性与定量表达的不确定性、随机性特征,证明是一种很好的聚类方法,模糊 C 聚类 FCM(Fuzzy C-Means)是其中一种健壮的聚类技术,它采用 ISODATA(Iterative Self-Organizing Data Analysis Techniques Algorithm)技术实现优化,对聚类中心的优化计算尤其有效,由于考虑了电力负荷序列的模糊特性,FCM 被证明在 STLF 中具有良好的分类效果,是本文样本预

处理的工具。

本文也应用了离散小波变换 DWT(Discrete Wavelet Transform)的相似性度量方法,DWT 用于电力负荷分段数据的降噪预处理<sup>[17]</sup>,提高相似性度量效果与计算效率。

综上所述,本文研究基于模式相似性与局部参数回归方法,将日负荷曲线视为函数型数据,从历史负荷数据本身挖掘负荷变动的内在随机分布规律,建立基于数据驱动的非参数模型。本文根据智能电表等间隔采集原理,将历史日负荷样本表示成等采样负荷段组成的子序列集合,基于 DWT 的形状相似性分段度量,应用 N-WE 权估计方法计算各负荷段在未来日估计中的权重,形成函数型小波核非参数回归 FWKNR(Functional Wavelet-Kernel Nonparametric Regression)方法,但非参数模型预测精度易受样本数据平稳性影响,本文应用 FCM 分析对单一用户历史分段样本进行模式识别与分类预处理,提取与未来预测日高度相关的有效样本,降低计算量,同时使训练样本具有更好的平稳性,从而提高了预测精度,基于聚类分析的 FWKNR 记为 C-FWKNR 方法。

## 1 FWKNR 预测方法

### 1.1 短期负荷分段预测问题描述

设对连续随机过程变量  $X(t)$  进行观察测量,按时序在时刻  $t_1, t_2, \dots, t_p$  采样得到的离散序列集合  $X(t_1), X(t_2), \dots, X(t_p)$  称为  $X(t)$  的时间序列。本文定义某地区用电负荷为连续的随机过程量  $X=(X(t); t \in \mathbf{R})$ , 预测其未来的发展趋势,预测问题描述为在时间间隔  $[0, T]$  内观察测量  $X$ , 预测在时间间隔  $[T, T+\delta]$  ( $\delta$  为时间间隔,  $\delta > 0$ )  $X$  的值。本文依据智能电表负荷采样过程,将反映负荷历史的时间间隔  $[0, T]$  分割成各子时间段  $[l\delta, (l+1)\delta]$  ( $l=0, 1, \dots, k-1; k=T/\delta$ ), 由此将离散时间过程  $S=(S_n; n \in \mathbf{N})$  ( $\mathbf{N}=\{1, 2, \dots\}$ ) 定义为:

$$S_n(t)=X(t+(n-1)\delta) \quad n \in \mathbf{N}, \forall t \in [0, \delta) \quad (1)$$

令  $S_n(t_i)$  为日负荷  $S_n$  在观察点  $t_i$  ( $i=1, 2, \dots, p$ ) 的值,则  $S_n=[S_n(t_1), S_n(t_2), \dots, S_n(t_p)]$  ( $n \in \mathbf{N}$ ) 表示一个由各观测负荷段集合组成的完整的日负荷曲线  $S_n$ 。本文根据智能电表的日负荷采样频率确定采样间隔  $\delta$ 。实际上,用电负荷序列被记录在一个数量有限的等距离时间点  $t_1, t_2, \dots, t_p$  上,若  $\delta$  为 0.5 h 则  $p$  取 48,若  $\delta$  为 1 h 则  $p$  取 24。由此定义短期日负荷分段预测问题为:根据第 1 天至第  $L$  天的所有历史负荷段样本  $S_1, S_2, \dots, S_L$  集合,由所提分段预测模型计算未来第  $L+1$  天所有负荷段集合  $S_{L+1}$ ,即分段预测确定  $S_{L+1}=[S_{L+1}(t_1), S_{L+1}(t_2), \dots, S_{L+1}(t_p)]$  日负荷曲线。



## 1.2 FWKNR 预测方法

本文采用基于核方法的非参数回归预测模型,利用基于最近观测负荷段  $S_L$  的有条件核平滑(kernel smoothing)方法预测  $S_{L+1}$ ,即对历史负荷段的带权加权均值(weighted average)进行计算,其权估计基于历史段与当前段  $S_L$  的相似性度量来确定权重,该模型具有很好的估计性能<sup>[6-7,11-12]</sup>。因此负荷曲线段相似性度量方法在预测  $S_{L+1}$  中起着重要作用。

(1)基于离散小波变换的形状相似性度量方法。

DWT 很好地揭示了测量信号受多因素事件影响呈现的内在多尺度(multi-scale)特点,成为时序数据预处理分析与建模的重要方法,本文在负荷曲线形状相似性度量计算中,首先将历史日负荷序列按 DWT 分解得到离散小波系数,并去除与高频振荡分量有关的高尺度(scale)部分;由于离散小波系数以各尺度表示了用电行为特征因素,提供了直观的数据约简与降噪预处理,分尺度计算曲线形状相似性改进了相似性度量效果,并降低了计算量。特别地,本文为了计算  $S_n$  和  $S_m(n \neq m)$  的相似性,分别记  $\theta_{j,k}^{(n)}$  和  $\theta_{j,k}^{(m)}$  是  $S_n$  和  $S_m$  在尺度  $j(j=j_0, j_0+1, \dots, J-1)$  和位移  $k(k=0, 1, \dots, 2^j-1)$  上的离散小波系数,其中  $j_0 \geq 0$  是给定的初始尺度,  $J$  是最大分解尺度,两者均由应用确定<sup>[24]</sup>。按照如下步骤计算  $S_n$  和  $S_m$  的相似性。

a. 在每个尺度  $j$  上,通过离散小波系数计算欧几里得距离:

$$d_j(\theta^{(n)}, \theta^{(m)}) = \left[ \sum_{k=0}^{2^j-1} (\theta_{j,k}^{(n)} - \theta_{j,k}^{(m)})^2 \right]^{1/2}$$

b. 量化  $S_n$  和  $S_m$  之间的相似性,按照如下距离公式计算在所有尺度上的距离:

$$D(S_n, S_m) = \sum_{j=j_0}^{J-1} 2^{-j/2} d_j(\theta^{(n)}, \theta^{(m)})$$

该相似性计算方法具有良好的度量效果<sup>[24]</sup>,能很好地表达日负荷曲线的形状的相似性,进而反映各日负荷曲线用电行为特征相似程度。

(2)函数小波核非参数回归预测方法。

本文采用基于核方法的非参数回归预测模型<sup>[6,16,18]</sup>;由于在历史日负荷  $S_1, S_2, \dots, S_L$  中,对于预测  $S_{L+1}$  包含最多有效信息的是  $S_L$ ,因此最近一天的日负荷  $S_L$  (称为基准参考段)对预测  $S_{L+1}$  起着很大的决定作用,  $S_{L+1}$  预测的结果被表示成历史负荷段的加权平均数,与  $S_L$  更相似的曲线被设置有更高的权重。预测负荷  $\hat{S}_{L+1}$  计算公式定义如下<sup>[18]</sup>:

$$\hat{S}_{L+1}(t_i) = \sum_{l=1}^{L-1} \omega_{L,l} S_{L+1}(t_i) \quad i=1, 2, \dots, p \quad (2)$$

此处权重  $\omega_{L,l} = \omega(S_L, S_l)$ , 称为 N-W 权(Nadaraya-Watson weights),定义如下:

$$\omega_{L,l} = \frac{K_h(D(S_L, S_l))}{\sum_{i=1}^{L-1} K_h(D(S_L, S_i))} \quad l=1, 2, \dots, L-1 \quad (3)$$

权重  $\omega_{L,l}$  应满足:对于  $l=1, 2, \dots, L-1, \omega_{L,l} \geq 0$ ,

$\sum_{l=1}^{L-1} \omega_{L,l} = 1$ 。且对称函数  $K$  要满足  $K_h(\cdot) = h^{-1} K(\cdot/h)$ , 在

映射  $\mathbf{R} \rightarrow \mathbf{R}$  中  $K(x) \geq 0$ ,  $\int K(x) dx = 1$  且  $\int x^2 K(x) dx < \infty$ 。模型更多细节参见文献[25]。

综上所述,本文将基于上述两步形成的预测模型称为 FWKNR。

(3)模型中外部环境变量的扩展。

电力系统中的负荷受多种外部环境变量的影响,如温度因素、日期因素、突发事件等。其中温度因素与日期因素对负荷的影响最为明显,本模型对外部变量的考虑,可以通过计算虚拟基准参考段(virtual reference segment)<sup>[18]</sup>,代替本文的最近日参考段  $S_L$ ,使虚拟基准参考段更确切地表达预测日负荷行为特征<sup>[18]</sup>;采用模糊关联聚类分析更准确地提取与预测日负荷相似的有效训练样本<sup>[22]</sup>,预计扩展模型会有更高的预测精度;由于目前环境变量较难获取,故作为本文的下一步工作目标。

## 2 FCM 改进的预测算法流程

### 2.1 FCM

FCM 由于考虑了事物之间的模糊特性,分类的效果很好,在电力系统有着很好的应用背景。所以本文采用 FCM 改进 FWKNR。本文采用如下方法实现 FCM。

定义  $c \times n$  阶的样本曲线隶属度矩阵  $U = [u_{ij}]$ ,  $c$  为聚类数量,  $n$  为样本数量,  $u_{ij}$  表示样本中第  $j$  个样本隶属于第  $i$  个类的隶属程度,且  $u_{ij}$  满足:  $\sum_{i=1}^c u_{ij} = 1$  ( $\forall j=1, 2, \dots, n$ )。  $u_{ij}$  的求解方法为:

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left( \frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)}} \quad (4)$$

其中,  $i=1, 2, \dots, c$ ;  $j=1, 2, \dots, n$ ;  $m \in [1, \infty)$  为模糊指数;  $d_{ij}$  为  $x_j$  到聚类中心  $V_i$  的距离。本文中的距离计算方法为欧几里得距离。  $V_i$  的计算方法是:

$$V_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m} \quad i=1, 2, \dots, c \quad (5)$$

FCM 步骤如下。

(1)确定聚类数  $c$ 、模糊指数  $m$  及循环停止条件  $\varepsilon$  的值,随机选择一组初始的聚类中心  $V_i(i=1, 2, \dots, c)$ ,初始化隶属度矩阵  $U^{(0)}(U=[u_{ij}])$ 。

(2)第  $k$  轮迭代:根据式(5)计算中心向量  $V^{(k)}$ 。

(3)根据式(4)更新隶属度矩阵  $U^{(k)}$  及  $U^{(k+1)}$ 。

(4)如果  $\|U^{(k+1)} - U^{(k)}\| < \varepsilon$  则算法停止,输出聚类中心  $V$  和隶属度矩阵  $U$ ;否则返回步骤(2)继续进

行迭代计算。

## 2.2 FCM 改进的预测算法流程

本文提出基于 FCM 的 FWKNR 组合模型,即 C-FWKNR,其预测方法如下:(1)使用 FCM 方法,将某用户的  $L$  个历史样本聚类到  $M$  个类中,每个类的日负荷曲线具有形状相似性关系,代表了一组具有某种相似消费行为的公共日模式,即某种类型的日负荷曲线样本集合;(2)识别参考段曲线  $S_L$  所在的类,此类中的历史样本与  $S_L$  具有曲线相似性关系;(3)应用步骤(2)中获得的  $S_L$  所在的类的样本,采用 FWKNR 模型预测  $S_{L+1}$ 。

下面详细描述 C-FWKNR 模型对某一用户的预测算法。

(1)采用 FCM 方法对用户曲线分类:假定有  $L$  条历史曲线  $S_1, S_2, \dots, S_L$ ,用 FCM 方法自动地将  $L$  条曲线分配到  $M$  个类中,记作  $G_1, G_2, \dots, G_m, \dots, G_M$ 。

(2)曲线识别:根据步骤(1),确定  $S_1, S_2, \dots, S_L$  所在的类,参考段  $S_L$  所在的类即为所识别的类,记作  $G_m = \{S_1^{(m)}, S_2^{(m)}, \dots, S_{K(m)}^{(m)}\}$ ,其中  $S_d^{(m)} (\forall d = 1, 2, \dots, K(m))$  是类  $G_m$  的样本曲线,  $K(m)$  是类  $G_m$  中的曲线数量。

(3)预测:用步骤(2)中包含的结果建立样本曲线  $\{(S_d^{(m)}, S_{d+1}^{(m)})\}_{d=1,2,\dots,K(m)}$ ,  $S_d^{(m)}$  (相对于第  $d$  天的负荷曲线)属于类  $G_m$ ,  $S_{d+1}^{(m)}$  是第  $d+1$  天的观测曲线,  $S_{d+1}^{(m)}$  不一定属于类  $G_m$ 。参考式(2),用式(6)来预测  $S_{L+1}$  的所有负荷段:

$$\hat{S}_{L+1}(t_i) = \sum_{d=1}^{K(m)} \omega_{L,d}^{(m)} S_{d+1}^{(m)}(t_i) \quad i=1,2,\dots,p \quad (6)$$

对于所有的  $d=1,2,\dots,K(m)$ ,  $\omega_{L,d}^{(m)} = \omega(S_L, S_d^{(m)})$ ,  $\omega(\cdot, \cdot)$  由式(3)定义。

## 3 算例结果与分析

### 3.1 数据样本的来源及特征

实验数据取自文献[26],数据为 2004 年至 2008 年 20 个地区的真实用电负荷数据,其中包含 33001 条负荷记录,采样间隔  $\delta$  为 1 h ( $p=24$ )。为了检测模型的预测能力,从 20 个地区中随机选择某地区的日负荷数据为实验数据,此处选择地区 1。实验时选择该地区 2007 年 1 月 1 日至 2008 年 5 月 31 日每隔 1 h 的日负荷曲线,将原始负荷曲线按照  $p=24$  进行划分以满足 C-FWKNR 模型的预测要求,令  $S_1, S_2, \dots, S_{517}$  为 517 d 的日负荷。

### 3.2 预测精确度的比较

为进行相关实验验证,将 517 个日负荷分为两部分:  $A = \{S_1, S_2, \dots, S_{181}\}$  为预测训练样本,即 2007 年 1 月 1 日至 2007 年 6 月 30 日的日负荷曲线集合;  $B = \{S_{182}, S_{183}, \dots, S_{517}\}$  为测试样本,用于检测 C-FWKNR 模型

的预测精确度。预测  $S_{L+1}$  的负荷值,则需要将  $S_1 \sim S_{L-1}$  作为历史日负荷,  $S_L$  作为参考段。在预测集  $B$  重复这个过程,直到集合  $B$  中所有的预测段都按照上述方式完成计算。

下面用 C-FWKNR 模型进行预测。

(1)聚类个数:采用无监督 FCM 方法,对  $S_1 \sim S_{181}$  进行聚类。模糊指数  $m=2$ 、停止条件  $\varepsilon=0.01$ 、聚类数  $c=4$  时的训练样本聚类见图 1,可以观察到每个类中在 07:30 左右都出现了用电峰值,而类 1 与类 2 在 20:00 出现的峰值低于 07:30 时刻的峰值,类 3 与类 4 在 19:30 的峰值高于 07:30 时刻的峰值。

(2)确定参考段  $S_L$  所属的类:将参考段  $S_L$  分配到一个合适的类,该实验中,  $S_L$  所属类为聚类 4,聚类 4 中  $S_{181}$  (本例的预测天为  $S_{182}$ ) 如图 1(d) 中所示。  $S_{181}$  在 05:00 用电量最低,在 19:00 达到用电最高峰,与聚类 4 反映出的用电习惯刚好吻合。

(3)预测与验证标准:为了预测  $S_{182}$  每小时的日负荷曲线,选择的样本数据为  $\{(S_d^{(m)}, S_{d+1}^{(m)})\}_{d=1,2,\dots,40}$ , 40 为聚类 4 中曲线数。按照式(4)来预测  $S_{182}$ 。本文核函数  $K$  选择高斯核函数。为更好地说明本文的预测方法,需要根据以上步骤将预测集  $B$  中的 336 d 全部进行预测,如图 2 所示,给出了  $S_{182} \sim S_{188}$  一周的预测值与真实值的比较结果。

C-FWKNR 模型的预测精确度定义与计算如下。

设  $S_d(t_i)$  和  $\hat{S}_d(t_i)$  分别表示  $t_i$  时刻的实际负荷值和预测负荷值,则定义每小时的绝对误差 OHAE (One-Hour Absolute Error) 为:

$$\text{OHAE}_d(t_i) = \left| \hat{S}_d(t_i) - S_d(t_i) \right|$$

其中,  $i=1,2,\dots,24$ ;  $d=1,2,\dots,336$ 。

定义每小时的相对误差 OHRE (One-Hour Relative Error) 为:

$$\text{OHRE}_d(t_i) = \left| \frac{\hat{S}_d(t_i) - S_d(t_i)}{S_d(t_i)} \right| \times 100\%$$

由此定义每天的平均相对误差 DMRE (Daily Mean Relative Error) 为:

$$\text{DMRE}_d = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{S}_d(t_i) - S_d(t_i)}{S_d(t_i)} \right| \times 100\%$$

如上  $n=24$ ,一天按小时分为 24 段计算误差。

图 3 为地区 1 在 2007 年 6 月 30 日至 2008 年 5 月 31 日中随机选择某一周(2007 年 7 月 30 日至 2007 年 8 月 5 日)的日负荷的 OHRE。由图 3 可看出,周一到周日每个小时的负荷预测精度都比较高,周内的 OHRE 平均值达到 1.01%,且 OHRE 在 2% 以下的达到了 83%。

图 4 为地区 1 在 2007 年 6 月 30 日至 2008 年 5 月 31 日年内日负荷预测的 DMRE,可以看出

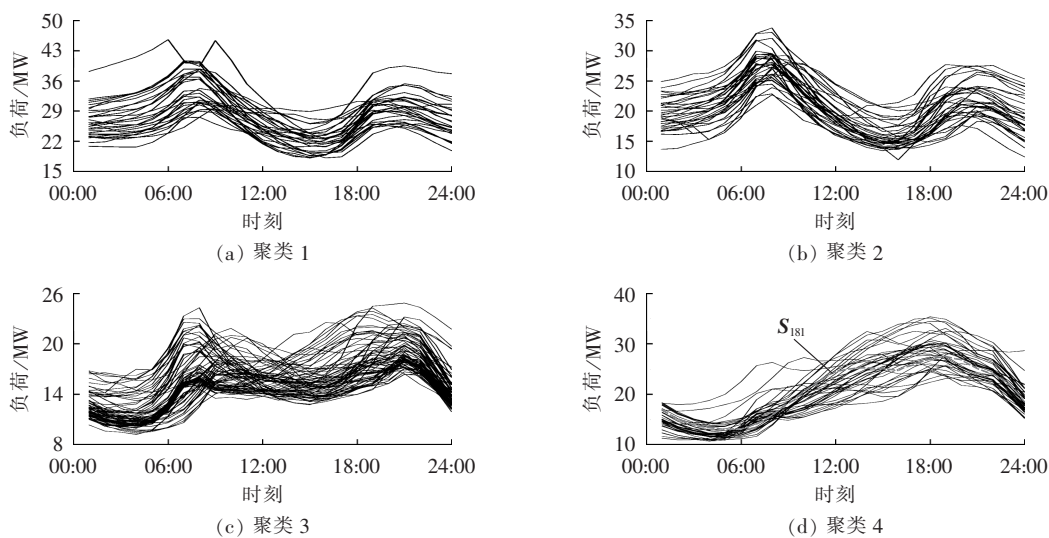


图 1 FCM 聚类算法将原日负荷曲线样本聚成典型 4 类

Fig.1 Original daily load curves are classified to four typical patterns by FCM clustering algorithm

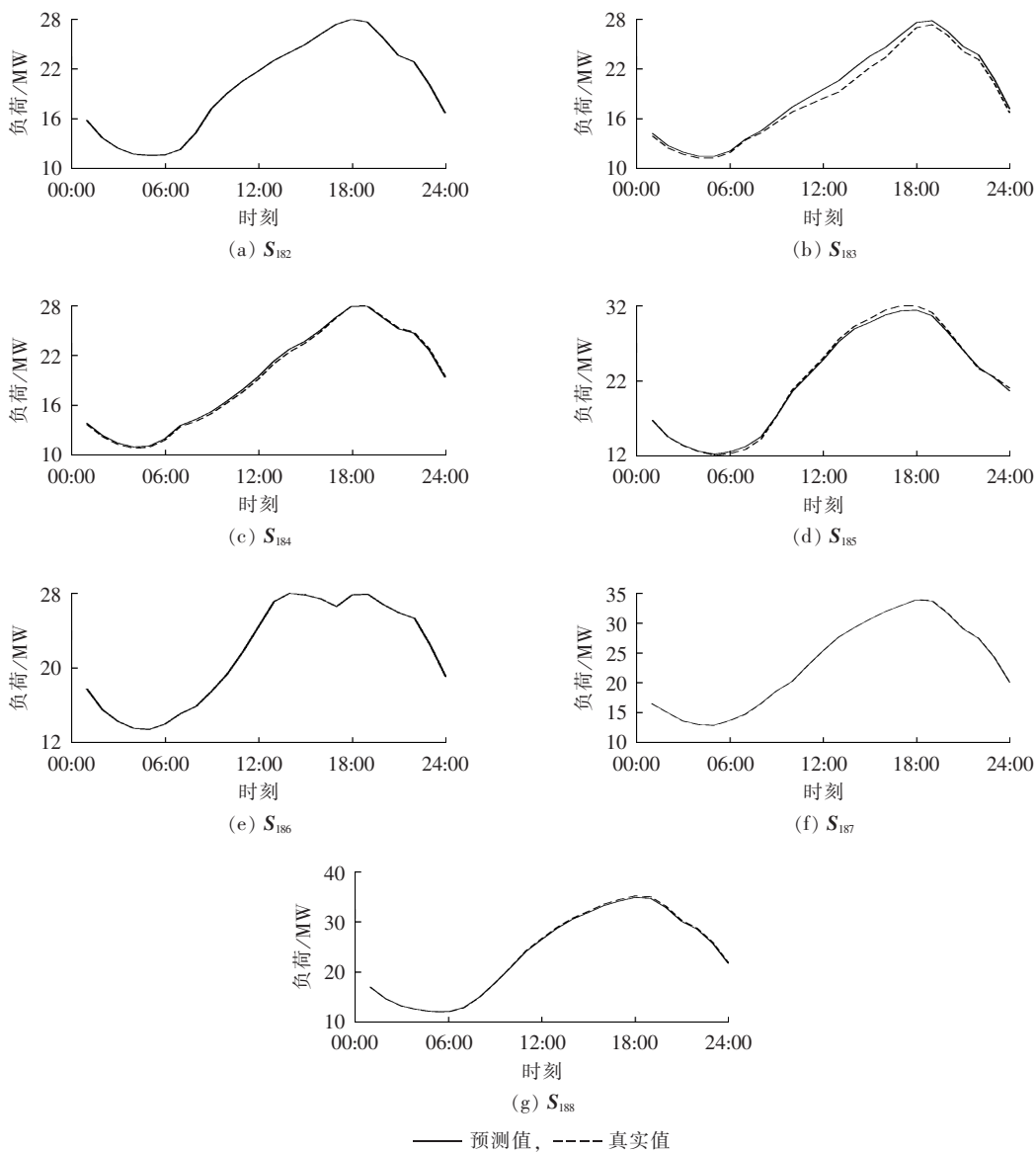


图 2  $S_{182}$ — $S_{188}$  一周日负荷预测值与真实值比较图

Fig.2 Comparison between predicted and actual daily load curve for a week from  $S_{182}$  to  $S_{188}$

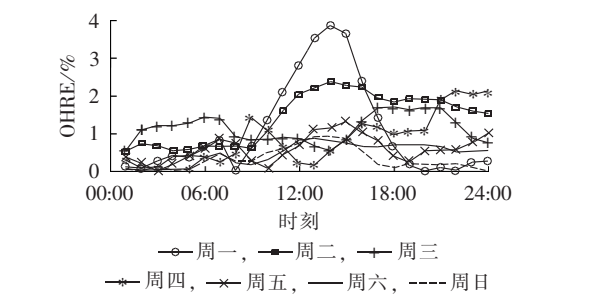


图 3 一周内每天的负荷预测精度  
Fig.3 OHRE for a week

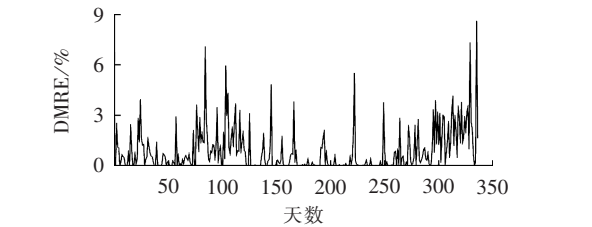


图 4 年内日负荷预测精度  
Fig.4 DMRE for a year

C-FWKNR 的预测精度总体较高,其 DMRE 平均值为 0.87%。表 1 为地区 1 在 2007 年 6 月 30 日至 2008 年 5 月 31 日内日负荷预测的 DMRE 在每个预测精度区间内的天数与比例,通过表 1 可以清楚看出大部分误差在 1% 以下,由此可说明 C-FWKNR 预测模型的预测精度较高。

表 1 年内各预测精度区间所占天数及比例  
Table 1 Days and proportion of DMRE intervals for a year

| DMRE/% | 天数  | 比例/% | DMRE/% | 天数 | 比例/% |
|--------|-----|------|--------|----|------|
| [0,1]  | 239 | 71.1 | (3,4]  | 18 | 5.4  |
| (1,2]  | 46  | 13.7 | (4,9]  | 8  | 2.4  |
| (2,3]  | 25  | 7.4  |        |    |      |

如图 5 所示,实验给出 FWKNR 与 C-FWKNR 这 2 个模型的月累计平均相对误差比较图,由于 2 个模型的误差都很小,采用 336 d 的误差不能明显表示比较结果,故选择按月对误差求平均值。表 2 比较了 FWKNR 与 C-FWKNR 在 1 月到 11 月(12 月缺少原始数据)的月相对误差平均值。显然,C-FWKNR 算法精度明显优于 FWKNR。

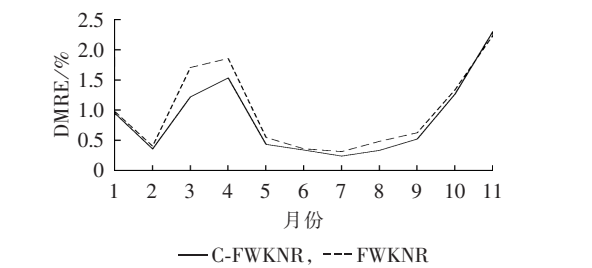


图 5 FWKNR 与 C-FWKNR 的误差比较图  
Fig.5 Comparison of DMRE between FWKNR and C-FWKNR

表 2 FWKNR 与 C-FWKNR 在 1 月到 11 月的平均误差比较表

Table 2 Comparison of monthly DMRE between FWKNR and C-FWKNR from January to November

| 月份 | DMRE/%  |       | 月份 | DMRE/%  |       |
|----|---------|-------|----|---------|-------|
|    | C-FWKNR | FWKNR |    | C-FWKNR | FWKNR |
| 1  | 0.96    | 0.98  | 7  | 0.25    | 0.32  |
| 2  | 0.36    | 0.41  | 8  | 0.34    | 0.49  |
| 3  | 1.22    | 1.71  | 9  | 0.53    | 0.63  |
| 4  | 1.53    | 1.86  | 10 | 1.27    | 1.34  |
| 5  | 0.44    | 0.55  | 11 | 2.30    | 2.30  |
| 6  | 0.35    | 0.37  |    |         |       |

4 结论

STLF 日益成为智能电网的重要问题,智能电表按固定间隔采集负荷数据,形成海量非线性时间序列大数据,为智能用电细粒度数据分析与运行精细化管理提供基础。由于受各种外部变量,诸如天气、温度、假期、周内/周末、经济等因素影响,通常认为 STLF 是一个困难问题。数据驱动的基于核方法的非参数模型能较好地表示历史样本数据所呈现的非线性特征,模型简单可靠,自适应性强,在 STLF 中获得了较好的应用效果,本文基于负荷序列分段下的模式相似性度量原理,采用 FCM 方法与 FWKNR 组合预测模型,即 C-FWKNR,提取与未来预测日高度相关的有效样本,降低计算量,同时使训练样本具有更好的平稳性,从而提高了预测精度;本文针对某地区实际负荷数据进行了较为详细的实验验证与误差分析,对比实验证明了 C-FWKNR 算法的优越性。该模型提供了一种简洁的应用智能电表海量数据分段预测的新颖方法,并且模型易扩展,考虑外部环境变量的影响因素,进一步提高预测精度;另外,模型具有普适性,可推广应用到诸如短期风电功率预测等智能电网时序大数据短期预测问题。

参考文献:

[1] 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):527-537.  
WANG Dewen,SUN Zhiwei. Big data analysis and parallel load forecasting of electric power user side[J]. Proceedings of the CSEE,2015,35(3):527-537.

[2] 钟清,孙闻,余南华,等. 主动配电网规划中的负荷预测与发电预测[J]. 中国电机工程学报,2014,34(19):3050-3056.  
ZHONG Qing,SUN Wen,YU Nanhua,et al. Load and power forecasting in active distribution network planning[J]. Proceedings of the CSEE,2014,34(19):3050-3056.

[3] DA SILVA P G,ILIC D,KARNOUSKOS S. The impact of smart grid prosumer grouping on forecasting accuracy and its benefits for local electricity market trading[J]. IEEE Transactions on Smart Grid,2014,5(1):402-410.

[4] QUILUMBA F L,LEE W J,HUANG H,et al. Using smart meter data to improve the accuracy of intraday load forecasting



- considering customer behavior similarities[J]. IEEE Transactions on Smart Grid, 2015, 6(2): 911-918.
- [5] KWAC J, FLORA J, RAJAGOPAL R. Household energy consumption segmentation using hourly data[J]. IEEE Transactions on Smart Grid, 2014, 5(5): 420-430.
- [6] CHAOUCH M. Clustering-based improvement of nonparametric functional time series forecasting: application to intra-day household-level load curves[J]. IEEE Transactions on Smart Grid, 2014, 5(1): 411-419.
- [7] FEINBERG E A, GENETHLIOU D. Applied mathematics for restructured electric power systems: optimization, control, and computational intelligence: load forecasting[M]. New York, USA: Springer, 2005: 269-285.
- [8] 雷绍兰, 孙才新, 周淦, 等. 一种多变量时间序列的短期负荷预测方法研究[J]. 电工技术学报, 2005, 20(4): 62-67.  
LEI Shaolan, SUN Caixin, ZHOU Quan, et al. Method of multivariate time series of short-term load forecasting[J]. Transactions of China Electrotechnical Society, 2005, 20(4): 62-67.
- [9] 王晔, 张少华. 一种应用时间序列技术的短期电力负荷预测模型[J]. 上海大学学报(自然科学版), 2002, 8(2): 133-136.  
WANG Xian, ZHANG Shaohua. An improved method for short-term electric load forecasting using time series techniques[J]. Journal of Shanghai University(Natural Science), 2002, 8(2): 133-136.
- [10] 陈昊. 基于广义自回归条件异方差模型的负荷预测新方法[J]. 电力系统自动化, 2007, 31(15): 51-54.  
CHEN Hao. A new load forecasting method based on autoregressive conditional heteroscedasticity model[J]. Automation of Electric Power Systems, 2007, 31(15): 51-54.
- [11] 陈昊. 基于不对称自回归条件异方差模型的短期负荷预测[J]. 电网技术, 2008, 32(15): 84-89.  
CHEN Hao. Short-term load forecasting based on asymmetric autoregressive conditional heteroscedasticity models[J]. Power System Technology, 2008, 32(15): 84-89.
- [12] 陈昊, 王玉荣. 基于随机波动模型的短期负荷预测[J]. 电力自动化设备, 2010, 30(11): 86-89.  
CHEN Hao, WANG Yurong. Short-term load forecasting based on SV model[J]. Electric Power Automation Equipment, 2010, 30(11): 86-89.
- [13] 陈昊, 万秋兰, 王玉荣. 基于自回归条件密度模型的短期负荷预测方法[J]. 东南大学学报(自然科学版), 2014, 44(3): 561-566.  
CHEN Hao, WAN Qiulan, WANG Yurong. Short term load forecasting method based on autoregressive conditional density model[J]. Journal of Southeast University(Natural Science), 2014, 44(3): 561-566.
- [14] 温权, 张勇传, 程时杰. 负荷预报的混沌时间序列分析方法[J]. 电网技术, 2001, 25(10): 13-16.  
WEN Quan, ZHANG Yongchuan, CHENG Shijie. Chaotic time series analysis to load prediction[J]. Power System Technology, 2001, 25(10): 13-16.
- [15] 张步涵, 刘小华, 万建平, 等. 基于混沌时间序列的负荷预测及其关键问题分析[J]. 电网技术, 2004, 28(13): 32-35.  
ZHANG Buhan, LIU Xiaohua, WAN Jianping, et al. Load forecasting based on chaotic time series and analysis of its key factors[J]. Power System Technology, 2004, 28(13): 32-35.
- [16] DUDEK G. Pattern similarity-based methods for short-term load forecasting-part 2: models[J]. Applied Soft Computing, 2015, 37(C): 277-287.
- [17] CHARYTONIUK W, CHEN M S, VAN OLINDA P. Nonparametric regression based short-term load forecasting[J]. IEEE Transactions on Power Systems, 1998, 13(3): 725-730.
- [18] PAPARODITIS E, SAPATINAS T. Short-term load forecasting: the similar shape functional time series predictor[J]. IEEE Transactions on Power Systems, 2012, 28(4): 3818-3825.
- [19] 许梁, 孙涛, 徐箭, 等. 基于函数型非参数回归模型的中长期日负荷曲线预测[J]. 电力自动化设备, 2015, 35(7): 89-94.  
XU Liang, SUN Tao, XU Jian, et al. Mid and long term daily load curve forecasting based on functional nonparametric regression model[J]. Electric Power Automation Equipment, 2015, 35(7): 89-94.
- [20] 王彩霞, 鲁宗相, 乔颖, 等. 基于非参数回归模型的短期风电功率预测[J]. 电力系统自动化, 2010, 34(16): 78-82.  
WANG Caixia, LU Zongxiang, QIAO Ying, et al. Short-term wind power forecast based on non-parametric regression model[J]. Automation of Electric Power Systems, 2010, 34(16): 78-82.
- [21] 张平, 潘学萍, 薛文超. 基于小波分解模糊灰色聚类与 BP 神经网络的短期负荷预测[J]. 电力自动化设备, 2012, 32(11): 121-125.  
ZHANG Ping, PAN Xueping, XUE Wenchao. Short-term load forecasting based on wavelet decomposition, fuzzy gray correlation clustering and BP neural network[J]. Electric Power Automation Equipment, 2012, 32(11): 121-125.
- [22] 姚李孝, 宋玲芳, 李庆宇, 等. 基于模糊聚类分析与 BP 网络的电力系统短期负荷预测[J]. 电网技术, 2005, 29(1): 20-23.  
YAO Lixiao, SONG Lingfang, LI Qingyu, et al. Power system short-term load forecasting based on fuzzy clustering analysis and BP neural network[J]. Power System Technology, 2005, 29(1): 20-23.
- [23] ZHANG T, ZHANG G, LU J, et al. A new index and classification approach for load pattern analysis of large electricity customers[J]. IEEE Transactions on Power Systems, 2012, 27(27): 153-160.
- [24] ANTONIADIS A, PAPARODITIS E, SAPATINAS T. A functional wavelet-kernel approach for continuous-time prediction[J]. Journal of the Royal Statistical Society, 2005, 68(5): 837-857.
- [25] FERRATY F, VIEU P. Companion manual of the book: non-parametric functional data analysis: theory and practice[M]. New York, USA: Springer-Verlag, 2006.
- [26] DEEPAK S D, SINGH S N. Electrical load profile analysis and peak load assessment using clustering technique[C]//PES General Meeting(Conference & Exposition. IEEE, 2014. [S.l.]: IEEE, 2014: 1-5.

#### 作者简介:



祖向荣

祖向荣(1963—),女,河北沧州人,副教授,博士研究生,研究方向为智能电网信息与通信技术、数据挖掘(**E-mail**:zuxrmail@163.com);

田敏(1989—),女,河北廊坊人,硕士研究生,主要研究方向为智能电网信息与通信技术、人工智能、数据挖掘(**E-mail**:0621tianmin@sina.com);

白焰(1954—),男,辽宁沈阳人,教授,博士研究生导师,博士,研究方向为控制理论与控制工程、智能电网、模式识别与智能系统(**E-mail**:by@ncepu.edu.cn)。

(下转第165页 continued on page 165)

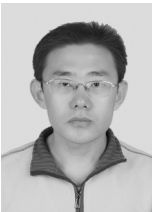
[22] 樊昌信,曹丽娜. 通信原理[M]. 北京:国防工业出版社,2007.

[23] VEDDER J D,何克明. 误差函数及其反函数的简单近似[J]. 物理实验,1989,9(3):114-115.

[24] HOOIJEN O G. A channel model for the residential power circuit used as a digital communications medium [J]. IEEE Transactions on Electromagnetic Compatibility,1998,40 (4):331-336.

[25] HOOIJEN O G,VINCK A J H. On the channel capacity of a European-style residential power circuit[C]//IEEE International Symposium on Power Line Communications and Its Applications. St Louis, Missouri, USA:IEEE, 1998:31-44.

作者简介:



胡正伟

胡正伟(1978—),男,山东昌邑人,讲师,博士,主要研究方向为智能电网通信及电力线通信关键技术 (E-mail:hzwizard@hotmail.com);

谢荣圆(1988—),女,河北清河人,硕士研究生,主要研究方向为电力线通信网络组网关键技术;

谢志远(1964—),男,河北辛集人,教授,博士研究生导师,博士,主要研究方向为配网自动化与电力线通信技术。

Mapping of PLC channel status to QoS parameter

HU Zhengwei,XIE Rongyuan,XIE Zhiyuan

(Department of Electronic and Communication,North China Electric Power University,Baoding 071003,China)

**Abstract:** A method of mapping the PLC(Power Line Communication) channel status to the QoS(Quality of Service) parameters is proposed,which applies the mathematical derivation to make the PLC channel status equivalent to QoS parameters,such as time delay,channel capacity,BER(Bit Error Rate) and delay jitter. The obtained QoS parameters are then used to determine whether the PLC channel satisfies the requirements of service transmission. A simulation system of P2P PLC in specific conditions is established based on the proposed method and the simulative results show its feasibility and effectiveness.

**Key words:** power line communication; channel status; QoS; service transmission; communication; mapping

.....

(上接第 140 页 continued from page 140)

Short-term load forecasting based on fuzzy clustering and functional wavelet-kernel regression

ZU Xiangrong,TIAN Min,BAI Yan

(Department of Control & Computer Engineering,North China Electric Power University,Beijing 102206,China)

**Abstract:** STLF(Short-Term Load Forecasting) is an important issue of smart grid. The historical daily load curve is expressed as a set of hourly load segments and a hybrid forecasting algorithm based on the pattern similarity method is applied,which combines the fuzzy clustering with the FWKNR(Functional Wavelet-Kernel Nonparametric Regression). FNWKR is applied to express the load curve of the predicted day as a set of weighted average of the corresponding hourly load segments of historical days,which assigns higher weight to the segment with higher similarity and uses N-WE(Nadaraya-Watson Estimator) to calculate the weight based on the shape-similarity measurement of discrete wavelet transform. The daily load is predicted by the quick forecasting of load segments. Fuzzy clustering is used to pre-classify the historical loads to typical load patterns for a particular customer and recognize the effective reduced training sample set with more similar behaviour pattern to the predicted day for the model forecasting. Based on the practical load data of a region,the experimental analysis verifies the superiority of the proposed algorithm.

**Key words:** short-term load forecasting; pattern similarity method; fuzzy clustering; wavelet-kernel nonparametric regression; nonlinear time series; smart grid