

阶梯电价机制下居民用电中期负载预测研究

实行阶梯电价已成为缓解我国能源供应紧缺、资源交叉补贴不合理等矛盾的重要举措。准确地刻画阶梯电价下的动态用户需求，有利于提高负载预测准确度。拟从现行递增阶梯电价动态需求的特征出发，提出一个新的分类负载预测模型。首先通过模糊聚类分析对某地区居民用电行为进行聚类；然后针对不同用户类别进行数据分类，再分别建立基于自组织模糊神经网络的负载预测模型；最后对各个模型进行加总。通过某地区实际居民数据检验，验证了分类负载预测模型的有效性。相比于传统的单一模型，新分类模型的预测准确度得到了提高，进一步保证了电网公司为居民提供高质量稳定的供电服务，同时帮助降低了电网公司的运营成本，提高了经济效益。

蔡秀雯 王 铮/国网泉州供电公司 傅 馨/厦门大学管理学院
曾晓军 冷 钢/School of Computer Science, University of Manchester, UK

外的29个省份也于2012年7月起全面试行居民用电阶梯电价。

居民阶梯电价自实施以来已取得了初步的成效，并在一定程度上增强了居民的节能意识，改变了一些不良的用电习惯。在传统的单一低价的电力价格制度下，不同特征用户群之间的用电行为（而非用电量）差异并不明显。然而，在新的阶梯电价制度下，不同特征（如收入、家庭结构和生活习惯等）的用户群对阶梯电价将产生不同的响应，由此所引起的用电行为差异将逐步凸显出来，这也使阶梯电价下的负载预测变得更加复杂。准确的负载预测不仅能保证电网系统安全稳定的运行，同时还能降低电力运营成本，提高经济效益和社会效益。负载预测按照时间范畴可分为短期、中期与长期。短期一般指未来几小时、一天到几天的预测；中期是指未来数周、数月的预测；而长期预测则是对未来数年、甚至更长时间的预测。相较于短期预测而言，中、长期负载时间跨度较长，需要的基础数据量较大，并且容易受多种因素的干扰，使预测误差累积而变得不可靠。然而，准确的中、长期负载能为配电网规划提供重要的决策支持，是实现智能电网的一个重要模块。

阶梯电价的科学理论基础通过市场细分的方法区分出不同特征用户群，针对不同用户群采用不同电价机制，以提高资源配置效率。我国尚处于全面实行阶梯电价的初步阶段，对我国阶梯电价下用电消费行为分析方面的研究还比较稀少，在理论和实践上均存在较

0 引言

近年来，随着我国社会经济的发展，人民生活水平的不断提高，对电力的需求也越来越大，随之而来的电力供应和减排压力也越来越大。长期以来，我国对居民电价实行单一形式的低电价政策，通常国家电网通过工业和商业用电的收益来对居民用电成本进行交叉补贴。这种现状造成了用电量越多的高收入用户，实际上享受了越多的电力补贴。此补贴机制不但不利于体现社会公平性，而且也不利于节能减排和资源的合理利用。为了缓解当前能源压力和改善我国电价结构，自2004年起，浙江、福建和四川等地区陆续推行了居民生活用电的阶梯电价，全国除新疆和西藏以

国家自然科学基金(71301133)，教育部人文社会科学基金项目(13YJC630033)。



蔡秀雯/工程师

关键词/Keywords

阶梯电价机制 ·
分类负载预测模型 ·
模糊C均值聚类 ·
自组织模糊神经网络 ·

大的研究空白。现有研究中用户细分方法一般都是基于家庭收入、用电量等的单一变量，将用户群简单分为高、中和低收入的用户群，同时需要事先确定用户群的数目。近期不少研究已经指出影响电力消费的因素纷繁复杂，更为丰富的影响变量需要在用户细分中考虑。另一方面，国内外关于电力负载预测的研究已经非常多了，以往研究总是将某个区域内用户的用电总量作为目标进行预测。传统的总量预测建立的是用户的平均行为模型，这类模型无法揭示各个顾客群体的不同行为，同时忽略了不同类型用户用电行为的差异性。因此，高效准确的用户聚类并且捕捉各类用户用电行为规律是在阶梯电价机制下智能电网规划急需改进的两个重要方面。

传统的负载预测是根据总量的历史数据来进行预测的，近年来，随着智能电能表的普及，现行的居民实时用采系统可以方便快捷地捕捉阶梯电价下更丰富和精细的居民用电数据，这为识别不同类型用户的用电特征提供了强有力的数据支持。本研究拟提出一个以聚类分析与负载预测算法为核心的分类预测模型。首先，提取出阶梯电价机制下居民用电行为的属性特征，通过聚类分析识别出阶梯电价机制下居民不同的用电行为特征，将具有相似用电行为特征的用户聚成一类，再针对每个用户类别分别建立起对应的中期负载预测模型。本文所提出的模型，结合了模糊C均值聚类（FCM）与自组织模糊神经网络（SOFNN）等方法，不仅能捕捉阶梯电价机制下用户用电行为的不同特征性，而且对于总体的中期负载预测准确度有所提高。准确的中期负载预测，有助于进一步为科学合理地开展智能配电网规划建设提供一系列决策支持。

相较于以往的研究，本文创新之处在于：
①提出了一个新的分类负载预测方法，从模型机理上与传统的总量负载预测模型是不同的；
②创新性地从阶梯电价相关的指标引入聚类模型中；
③利用智能电能表提供更准确、更全面的数据，更好地进行中后期负载预测。同时，相比传统的每15 min数据采集，在保证准确度要求的前提下，本

文利用较少的数据采样点（每日数据）进行中后期预测。

1 分类负载预测模型

1.1 模型框架

自20世纪80年代中期以来，国内外学者开展了大量的以各种电力负载预测模型和方法为主体的电力负载预测研究。但绝大多数为单一负载预测模型，主要包括回归分析、时间序列、神经网络与支持向量机等模型。单一负载预测模型通过分析影响区域用电量的关键因素，捕获变量间的相关关系，并以此构建出模型进行预测。然而，几乎所有的单一负载预测模型只是将用户的用电量进行简单加总，以区域用电总量为目标建立该区域的平均行为模型，而忽略了不同类型用户用电行为的特征性。

分类负载预测模型通过对单一负载预测模型所含有的信息进行分类，不同的行为采用不同的预测模型，这有助于提供更准确的预测结果。如有两个不同的用户群，分别为价格敏感度高和价格敏感度低的用户群，分类预测可以知道每个用户群的准确行为，而传统总量的平均模型则会导致对价格敏感度高用户群预测过高，而对价格敏感度低用户群预测过低。

目前，常见的组合负载预测模式是对同一输入输出数据集建立多个不同的预测模型，然后将预测结果组合起来，或按适当的权重进行加权平均，采用更复杂的非线性组合模型，最后选择拟合度最佳或标准离差最小的组合模型。本文所提出的分类负载预测模型则采用不同的思路，其出发点是先将数据按不同的行为特征分成 N 类，其结果是一组 N 个具有不同输入输出的数据集，然后对每个数据集建立起相应的模型，最后对预测结果进行汇总。具体来说，本文对所收集的数据进行预处理后，利用聚类算法捕捉和识别不同用户在分析期内用电行为的特征性，将用电行为类似的用户聚成一类。然后对数据进行分类，即每个数据类的输入和输出都不同。然后对不同类型的

用户（即不同数据集）分别建立最适合的负载预测模型。进行预测时，首先要识别预测用户属于哪个聚类，然后选择对应的预测模型进行负载预测，最后汇总单个预测结果得到最终总体预测结果。具体的框架结构如图1所示。

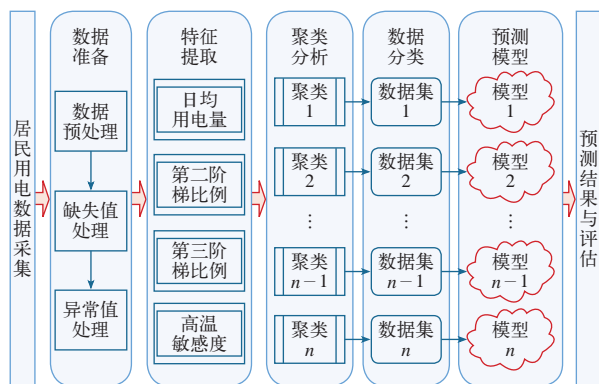


图1 分类负载预测模型框架

本文使用模糊C均值聚类（FCM）与自组织模糊神经网络（SOFNN）模型结合的方法进行分类负载预测。相较于传统的K-means聚类算法，FCM加入了模糊的概念，使每一个输入向量（用户特征）不再仅隶属于某一个特定的聚类，而是以其隶属度来表示其属于不同聚类的程度。另外，SOFNN的优点在于：①SOFNN能自动确定模型的结构，识别模型的参数；②预测准确度较高。

同时，本文所提出的分类负载预测模型还具有通用性与兼容性好的特点。其他常用的聚类方法（如K均值、自组织特征映射神经网络）与负载预测模型（如回归、时间序列和支持向量机）也可方便地在本框架下自由组合。此外，不同的用户聚类可以采用完全不同的预测模型来进行负载预测，这有助于对每个用户聚类采用最适合的预测模型，极大地提高了分类负载预测模型的灵活性和预测准确度。本文所提出的分类预测模型和传统的组合预测不同之处在于：①分类模型中，每一个子模型都是预测一类负载，但所预测的输入输出都是不同的；②对各个子模型的预测结果只是作了汇总，并没有组合预测模型中线性加权或者非线性组合的建模过程；③分类模型可以方

便地获得两类输出，除了传统的总量预测结果外，还可以获得各个用电类别的预测结果，但组合预测模型只能获得总量预测结果。

1.2 特征提取

特征提取主要包括聚类属性选择与预测输入变量提取两个部分。经过数据分析和预实验，本文提取了各用户日均用电量、第二阶梯比例、第三阶梯比例与高温敏感性四组聚类属性，来反映居民一定时间段内的负载变化规律。日均用电量主要取决于家庭所拥有的各类电器设备的数目，借此可推测居民的收入水平，而收入水平接近的居民用电行为往往比较相似。第二阶梯比例、第三阶梯比例则可以反映用户在过去几个月内用电的波动性，捕获阶梯电价机制下各用户长期的用电规律。此外，研究表明气象因素尤其是温度，往往对居民用电变化产生影响。尤其是夏季持续高温，空调机等家用电器使用频率较高，日均用电量往往大幅提高。高温敏感性指标反映用户在高温天气用电量的波动情况。四组指标的计算方法为

$$\text{日均用电量} = \text{总用电量} / \text{采样总天数} \quad (1)$$

$$\text{第二阶梯比例} = \text{到达第阶梯月份数} / \text{总月份数} \quad (2)$$

$$\text{第三阶梯比例} = \text{到达第阶梯月份数} / \text{总月份数} \quad (3)$$

$$\text{高温敏感性} = \text{高温日平均用电量} / \text{日均用电量} \quad (4)$$

其中，本文所定义的高温是指平均温度大于等于25℃。

同时，所有预测模型的输入变量主要提取出过去7天的用电量与当天的温度共计8个输入属性。其中，应用历史负荷数据对预测有益，因为可以使用滚动预测方式进行。而如果温度未知且需要预测，可以使用天气预报数据或该地区过去几年同一日的平均温度进行估计。

1.3 聚类分析

FCM是由Dunn于1973年提出的，其基本思想是通过不断优化目标函数来获得各样本点对于所有聚类中心的隶属度，进而确定样本点的类属，最终达到自动对样本数据聚类的目的。

假设样本集合为 $Z = \{z_1, z_2, \dots, z_N\}$ ， N 为样本总

数。将其分为 C 个模糊聚类组，并且求出聚类中心集为 $V = \{v_1, v_2, \dots, v_c\}$ ，依据最小二乘法原理，采用以下优化目标函数来划分数据

$$J(U, V) = \sum_{c=1}^C \sum_{n=1}^N (u_{cn})^m \|z_n - v_c\|^2 \quad n=1, 2, \dots, N \quad c=1, 2, \dots, C \quad (5)$$

式中， m 为模糊调节参数； u_{cn} 是第 n 个样本第 c 类的隶属度，且 $\sum_{c=1}^C u_{cn} = 1, 0 \leq u_{cn} \leq 1, U = [u_{cn}]$ 是 $C \times N$ 维的矩阵。

此外，在模糊聚类之前，需将提取出的特征属性进行归一化，即将这些的属性值映射到 $[0, 1]$ 之间，以去除不同量级对用户用电量特征的影响。通常采用极大极小值法对数据集进行归一化处理，处理方法如下

$$z'_n = \frac{z_n - z_{n_{\min}}}{z_{n_{\max}} - z_{n_{\min}}} \quad (6)$$

式中， z'_n 为采用极大极小值法归一化后的第 n 样本数据； $z_{n_{\max}}$ 和 $z_{n_{\min}}$ 分别为数据序列的最大值和最小值。

1.4 自组织模糊神经网络

Leng于2004年提出SOFNN模型如图2所示，该模型是由输入层、椭球基函数(EBF)层、标准化层、加权平均层和输出层构成。其中：

1) 输入层中的神经元 $i = 1, 2, \dots, r$ 表示输入变量 x_i 。

2) EBF层中的神经元 $j = 1, 2, \dots, u$ 表示一个模糊规则的前提，各神经元将其中所有隶属函数的值相乘作为输出结果 Φ_j ，具体算法如下

$$\Phi_j = \exp \left[- \sum_{i=1}^r \frac{(x_i - c_{ij})^2}{2\delta_{ij}^2} \right] \quad (7)$$

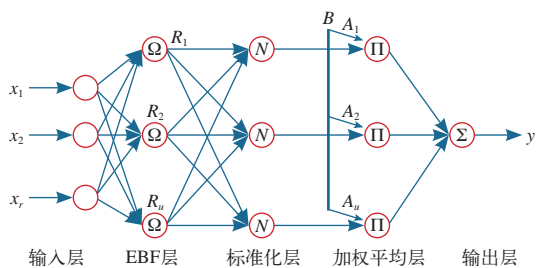


图2 自组织模糊神经网络基本框架

式中， c_{ij} 为隶属函数的中心； δ_{ij} 为隶属函数的宽度。

3) 标准化层中的神经元个数一般与EBF层相同，其对应的输出结果 Ψ_j 为

$$\Psi_j = \frac{\Phi_j}{\sum_{k=1}^u \Phi_k} = \frac{\exp \left[- \sum_{i=1}^r \frac{(x_i - c_{ij})^2}{2\delta_{ij}^2} \right]}{\sum_{k=1}^u \exp \left[- \sum_{i=1}^r \frac{(x_i - c_{ij})^2}{2\delta_{ij}^2} \right]} \quad (8)$$

4) 加权平均层中各个神经元的输出是将标准化层输出结果乘以加权误差 w_2 ，该层神经元输出结果为 $f_j = w_2 \Psi_j$ 。

5) 输出层中的每个神经元表示通过加总加权平均层中的输出结果得到的变量。因此，该层的输出结果 y 为

$$y = \sum_{j=1}^u f_j = \sum_{j=1}^u w_2 \Psi_j = \sum_{j=1}^u w_2 \frac{\exp \left[- \sum_{i=1}^r \frac{(x_i - c_{ij})^2}{2\delta_{ij}^2} \right]}{\sum_{k=1}^u \exp \left[- \sum_{i=1}^r \frac{(x_i - c_{ij})^2}{2\delta_{ij}^2} \right]} \quad (9)$$

SOFNN学习过程主要包括参数学习与结构学习。参数学习通过在线递推最小二乘算法使网络快速收敛；结构学习则是通过自动地增加、修改或删减EBF层中神经元的自组织方式，找到最为合适的神经网络规模。结构学习主要包括三个关键步骤：①增加神经元；②修剪神经元；③合并隶属函数与规则。而通过SOFNN中的参数学习与结构学习算法，可以为每个聚类寻找最优的网络结构并进行预测。

2 算例实验

2.1 数据和数据准备

本文以某地区533户居民为对象进行实例分析，由于该地区抄表日为11日，以2014年4月11日~2015年1月10日的用电数据作为训练集，2015年1月11日~2月10日的用电数据作为测试集来进行实验。

数据准备阶段主要任务包括：①数据预处理，由于所收集的智能电能表记录的数据均为累计值，因此计算各用户当日用电量需要将当日的

用电累计值减去前一日的用电累计值；②缺失值处理，预处理后需检测数据中是否存在缺失的现象，通过计算用电缺失时间段前后一天的累计电量之差，并按照缺失的天数进行平均，将缺失的数据补齐；③异常值处理，对超出设定指标阈值范围的数据样本进行过滤，如路灯、工业用户与长期无主的住户等。

2.2 FCM聚类分析

传统的FCM算法需要用户事先确定聚类个数。在聚类分析中，聚类个数的确定对聚类结果将产生很大影响。为了客观地确定合适的聚类个数，本文主要通过计算每一个试探类别数的误差二次方和（SSE）和MIA值，通过对比分析找出最优的聚类个数参数。SSE和MIA计算公式如下

$$SSE = \sum_{c=1}^C \sum_{k=1}^{n_c} \|z_{ck} - v_c\|^2 \quad (10)$$

$$MIA = \sqrt{\frac{1}{C} \sum_{c=1}^C \left[\frac{1}{n_c} \sum_{k=1}^{n_c} (z_{ck} - v_c)^2 \right]} \quad (11)$$

式中， n_c 为第 c 类中的样本数据的个数； z_{ck} 为第 c 类中的第 k 个样本。

根据指标变化走势选取最优聚类数，所得到的聚类结果如图3所示。当类数超过6时，随着类数的增加，曲线越来越平坦，SSE和MIA指标值的减小趋势明显减弱。

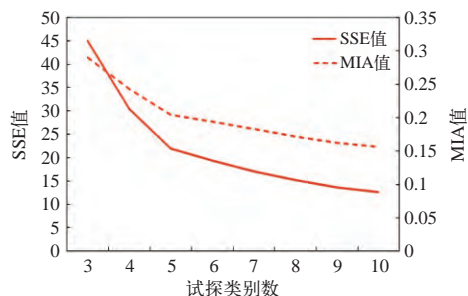


图3 不同类别的SSE和MIA指标测试结果

为了保证每一个聚类中心均有一定数量的样本，因此将聚类个数设定为6，聚类结果如表1所示。观察K-means聚类中心的数值，结合每一类的样本特征，可以总结如下：

1) 第一类用户平均用电量最少，每一类用电

量依次增加，第六类用户平均用电量最多。

2) 第一类用户用电量基本停留在第一阶梯电价规定的用电量之内，到达第二、第三阶梯的用电量很少；第二类用户用电量平均分布在第一、第二阶梯电价规定的用电量之内；第三类用户用电量基本停留在第二阶梯电价规定的用电量之内，第一、第三阶梯的用电量很少；第四、五类的用户则主要停留第二、三阶梯中。其中，第四类用户的用电量更经常落在第二阶梯，而第五类用户更经常落在第三阶梯；第六类用户用电量基本停留在第三阶梯电价规定的用电量之内，第一、第二阶梯的用电量极少。

3) 可以看出，对于第一类、第三类和第六类用户，阶梯电价并不能很理想地减少其用电量，原因在于这类人每个月的用电量较为稳定，几乎处于某一阶梯用电量之内不变，阶梯电价的微调并不能引起他们对生活用电量的显著改变。对于第二类、第四类、第五类用户，阶梯电价对其影响可能比较明显，原因在于这三类用户的用电量分布在不同的阶梯电价内，一般会出现阶梯跨度。当由某一阶梯用电量跳入另一阶梯用电量时，根据理性经济学人假设，价格敏感度高的用户都会相应地减少其用电量，避免在更高的电价下面使用更多的用电量。不过这些结论只是基于聚类数据初步总结出来的，如果想要得到更确切的结论，还需要进一步的研究。

表1 6个聚类中心结果

聚类	数量	日均用电量	第二阶梯比例	第三阶梯比例	高温敏感性
1	75	0.057 2	0.027 7	0.004 6	0.353 5
2	113	0.161 6	0.379 4	0.093 1	0.614 8
3	81	0.196 5	0.825 3	0.071 2	0.421 2
4	118	0.294 4	0.545 7	0.416 6	0.594 6
5	89	0.338 4	0.397 6	0.548 5	0.598 8
6	57	0.469 5	0.064 5	0.927 2	0.522 4

同时，根据原始数据集分别描绘出各聚类中居民用电量曲线簇如图4所示。从中可以发现类别1中的用户每天的用电量都很少，并且非常平均，基本集中在10 kW·h以下，对高温天气不敏感；类别3中的用户同样对高温天气不敏感，而其余类别

均在夏季用电量有明显幅度的增加，且对应的高温敏感性指标较高，这些类别中的用户对于高温天气比较敏感。

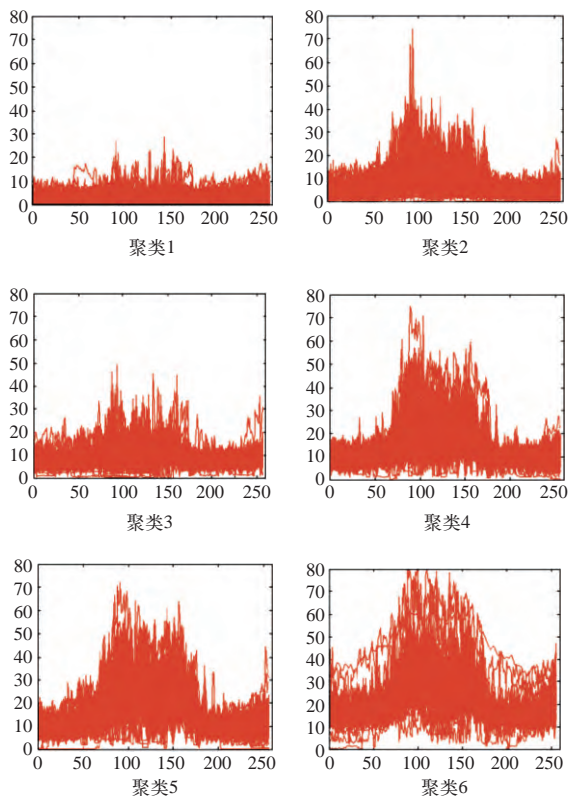


图4 聚类后的居民用电量曲线簇

2.3 SOFNN负载预测

将居民用电数据按照聚类结果进行加总，再分别搭建起相应的SOFNN模型。通过多次模拟实验与跟踪误差矫正，估计较优的SOFNN预实验参数为 $\delta = 0.01$ ， $\sigma_0 = 0.1$ ， $k_{rmse} = 0.01$ 和 $k_d(i) = 0.01$ ($i = 1, 2, \dots, 8$)。

预测的准确性由常用的平均绝对百分比误差 (MAPE) 与最大绝对百分比误差 (ME) 进行衡量，计算公式如下

$$MAPE = \frac{\sum_{n=1}^{\bar{N}} \frac{|y_n - \hat{y}_n|}{y_n}}{\bar{N}} \times 100\% \quad (12)$$

$$ME = \max \left| \frac{y_n - \hat{y}_n}{y_n} \right| \times 100\% \quad (13)$$

式中， y_n 和 \hat{y}_n 为该区域用电总量的实际值与预测

值； \bar{N} 为预测的天数。

各个聚类所对应模型的滚动误差如表2所示。整体来看，除了模型1外，各模型的滚动预测误差均在可接受的范围内。原因在于模型1所对应的聚类1中存在着一些用户，其用电行为随机性特别强，用电规律难以甄别。在分类预测模型中，这类用户的用电行为和其他有规律的用户用电行为差别较大，往往会将这类用户归为一类。但由于随机性强等特点，这类用户的用电行为往往无法准确预测。不过，模型1所预测的用电量相对于总用量的比例较小，对分类预测模型的预测准确度影响并不大。

表2 聚类所对应模型的滚动预测误差

误差指标	模型1	模型2	模型3	模型4	模型5	模型6
MAPE(%)	22.87	7.27	5.14	3.86	2.79	3.29
ME(%)	75.34	17.43	21.34	9.80	14.25	8.90

将各模型预测结果进行汇总，得到最终分类负载预测模型结果如图5所示。单一负载预测模型与分类负载预测模型的MAPE值分别为3.34%与2.78%，均在4%以内，准确度结果令人满意。相比于单一负载预测模型，分类负载预测模型整体预测准确度提高了0.56%。此外，在第14天时（即2015年1月25日）两个模型预测准确度同时下降至90%以下，查阅了相关历史数据发现，其原因是由于区域性停电导致实际用电量突然大幅降低。对于这种突发事件，预测模型难以及时对此进行响应和修正，这在一定程度上影响了总体的预测准确度。

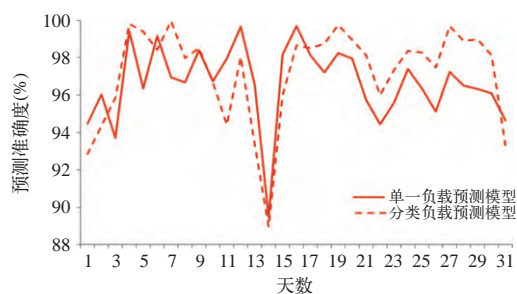


图5 测试集预测结果比较

3 结束语

本文通过关注阶梯电价机制对居民用电行为

的影响,提出了分类负载预测模型并进行区域中期负载预测。该模型组合了模糊C均值(FCM)与自组织模糊神经网络(SOFNN)等方法,并对某地区533户居民的2014年4月~2015年2月的用电数据进行建模与预测分析。结果表明,该模型不仅能够高效、准确地进行用户聚类,捕捉阶梯电价机制下各类用户用电行为的特征性,还能显著提高该区域总体负载预测的准确度。但是,目前本文的数据样本大多来自同一个或者邻近小区,用户特征差异不是特别大,未来可以多采集地域特征和用户特征差异更为显著的用户数据进行分析。另外,此模型可以用来建立某地区更准确的需求预测模型,为今后的阶梯电价参数优化模型提供坚实基础。

参考文献

[1] 黄海涛,张粒子,乔慧婷.基于变密度聚类的居民阶梯分段电量制定方法[J].电网技术,2010,34(11):111-116.

[2] 陈耀武,汪乐宇,龙洪玉.基于组合式神经网络的短期电力负荷预测模型[J].中国电机工程学报,2001,21(4):79-82.

[3] Papalexopoulos A D, Hesterberg T C. A regression-based approach to short-term system load forecasting [J]. IEEE Transactions on Power Systems, 1990, 5(4): 414-423.

[4] 李玲玲,朱博.基于混沌时间序列的短期电力负荷预测[J].信息技术,2009(3):44-46.

[5] 翟永杰,王静娴,周黎辉.基于模糊支持向量机的电力系统中期负荷预测[J].华北电力大学学

报:自然科学版,2008,35(2):70-73.

[6] 程旭,康重庆,夏清.短期负荷预测的综合模型[J].电力系统自动化,2000,24(9):42-44.

[7] 赵宇红,唐耀庚,张韵辉.基于神经网络和模糊理论的短期负荷预测[J].高压电技术,2006,32(5):107-110.

[8] Dunn J C. Well-separated clusters and optimal fuzzy partitions [J]. Journal of Cybernetics, 1974, 4(1): 95-104.

[9] Mao H, Zeng X J, Leng G, et al. Short-term and midterm load forecasting using a bilevel optimization model [J]. IEEE Transactions on Power Systems, 2007, 24(2): 1080-1090.

[10] 李炎,翟永杰,周倩.基于EUNITE竞赛数据的中期电力负荷预测[J].华北电力大学学报,2007,34(4):22-26.

[11] 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其处理策略[J].电网技术,2006,30(7):5-10.

[12] Leng G, Prasad G, Mcginnity T M. An on-line algorithm for creating self-organizing fuzzy neural networks [J]. Neural Networks, 2004, 17(10): 1477-1493.

[13] Lin B Q, Jiang Z J. Designation and influence of household increasing block electricity tariffs in China [J]. Energy Policy, 2012(42): 164-173.

[14] 宋才华,王永才,蓝源娟,等.基于数据挖掘的电力行业客户细分模型分析[J].现代电子技术,2014,37(14):21-24.

(收稿日期:2015-06-29) EA

欢迎订阅 欢迎投稿 欢迎刊登广告