DOI：10.13334/j.0258-8013.pcsee.2016.03.003 文章编号：0258-8013 (2016) 03-0604-11 中图分类号：TM 71

采用互信息与随机森林算法的用户用电关联因素辨识及用电量预测方法

赵腾[1]，王林童[1]，张焰[1]，田世明[2]

(1．电力传输与功率变换控制教育部重点实验室(上海交通大学)，上海市 闵行区 200240；
2．中国电力科学研究院，北京市 海淀区 100192)

Relation Factor Identification of Electricity Consumption Behavior of Users and Electricity Demand Forecasting Based on Mutual Information and Random Forests

ZHAO Teng[1], WANG Lintong[1], ZHANG Yan[1], TIAN Shiming[2]

(1. Key Laboratory of Control of Power Transmission and Conversion, Ministry of Education (Shanghai Jiao Tong University), Minhang District, Shanghai 200240, China; 2. China Electric Power Research Institute, Haidian District, Beijing 100192, China)

ABSTRACT: With the development of smart grid, electric utilities have accumulated vast amounts of data, which provides a foundation for the meticulous forecasting of electricity demand. However, the utilization of big data related to electricity demand forecasting is sophisticated, taking into account the data features of variety, volume, velocity and high-dimension. In view of the above problem, a subspace clustering method applicable to massive users was proposed to extract diverse power consumption modes, based on the assessment indexes of users' electricity consumption characteristics. Then, all users were classified into several groups according to different electricity consumption modes, and mutual information matrix was constructed in each user group to identify the relation factors of users' electricity consumption, from the aspects of regional and industrial economic data, weather conditions, and electricity price, etc. Finally, a big data driven forecasting model based on random forests was established. The method proposed in this paper effectively identified the relation factors of electricity consumption for each user group, and reduced the adverse impacts on electricity demand forecasting caused by the diverse electricity consumption modes. Simulation results show that the suggested method can provide higher prediction accuracy and is suitable for the analytical processing of big data.

KEY WORDS: big data; subspace clustering; mutual information; relation factor identification; random forests; electricity demand forecasting

摘要：随着智能电网的不断发展，电力企业积累的大量数据为用户用电量精细化预测提供了数据基础。针对与用电量预测相关的大数据种类多、体量大、维度高和生成速度快等特点，在研究用户用电特性评价指标的基础上，提出海量用户用电特性子空间聚类分析方法，挖掘用户多种用电模式。根据不同用电模式对用户进行群体划分，并利用互信息矩阵从区域及行业经济数据、气候条件，以及电力价格等方面辨识与用户群体用电量相关联的因素，进而构建基于随机森林算法的用电量大数据预测模型。该文方法可以有效识别不同用户群体的用电关联因素，规避用电模式差异性为用电量预测带来的不利影响。仿真结果表明，该方法具有较高的预测精度，且适用于大数据分析处理。

关键词：大数据；子空间聚类；互信息；关联因素辨识；随机森林；用电量预测

0 引言

准确的用电量预测对电网规划和经济部门的管理决策具有重要的指导意义。在研究不同用户用电特性的基础上开展用电量预测，可以帮助电力企业更好地了解用户个性化服务需求，为未来电网发展及电力需求侧响应政策的制定提供数据支撑[1]。

随着我国社会经济的持续发展以及产业结构的不断调整，电力用户的用电特性正呈现多样化发展趋势：对同一行业的不同用户，其用电行为的差异化日益明显，仅以行业总体特性进行用电模式识

基金项目：国家 863 高技术研究发展计划项目(2015AA050203)；国家电网公司科技项目(520900150037)。

The National High Technology Research and Development of China 863 Program (2015AA050203); State Grid Science & Technology Project (520900150037).

别已无法客观挖掘足够的信息[2]；对不同行业的用户，由于其在社会经济活动中的分工、资源配置和服务对象等具有一定的不确定性，且对上下游行业发展情况存在依赖关系，因此用户用电特性还与除本行业以外的多种社会经济因素存在关联关系，这种关联关系的复杂度也在不断提高。不同区域用户的用电特性呈现与不同行业用户类似的变化趋势。用户用电特性的多样化对传统的用电量预测方法提出了挑战。

与此同时，随着智能电网的建设和发展，电力企业内部逐渐形成了包括生产数据、营销数据，以及相关社会经济数据等在内的智能配用电大数据[3]，为计及用户用电特性的用电量精细化预测提供了数据基础。应用智能配用电大数据分析结果，将用户用电特性进行多维度分解，对隶属于不同用电模式的用户群体采用差异化建模方法，分别建立有较强针对性的预测模型，可以提高用电量预测精度，同时，清晰的用电模式信息有助于电力企业更深刻地认识用户及其群体效应[2-3]。

然而，由于智能配用电大数据种类多、体量大、维度高和生成速度快等特征，使得传统的用电量预测方法在挖掘海量数据信息方面存在一定的局限性，难以准确把握用户的用电量关联因素及变化规律。如何在大数据环境下研究用户的用电特性和用电量关联因素，并对其用电量进行预测，是摆在研究者面前的一个挑战[4]。

目前，多数的用电量预测模型可归为 3 类：基于时间序列的模型、用户端模型，以及计量经济模型[5]。其中：基于时间序列的模型可用于识别用户用电需求在过去与将来之间的时间因果关系，常用的方法包括自回归移动平均模型、支持向量机，以及神经网络等[6-8]，该类方法可作为黑箱工具来精确捕捉电能消耗的时间动态特性，然而，它们无法提供隐藏于电能需求变化之后的深层次关联关系；用户端模型将用户用电量分解为若干个主要组成部分，对各部分分别进行预测建模[9-10]，可以较为方便地解释用户用电量与用电行为之间的关联关系，但该类模型的预测精度高度依赖于可用信息的质量，而且各用电量组成部分的划分尚缺乏科学的理论指导；计量经济模型可以识别影响用户用电行为的主要经济因素，并挖掘关键影响因素和电能消费情况之间的映射关系[11-12]，然而该类方法在确定输入预测模型的影响因素数据集时需进行人为干预。

在智能配用电大数据快速发展的背景下，本文结合现有用电量预测模型的特点，提出一种基于互信息(mutual information，MI)与随机森林(random forests, RF)算法的用户用电关联因素辨识及用电量预测方法。首先从多维度分析不同用户的用电特性，并根据用电特性的差异对用户进行群体划分，然后利用平均互信息自动筛选出与用户群体的用电量存在强关联的因素，最后采用随机森林算法针对不同用户群体分别建立用电量预测模型。以上海某区域 5360 家用户为例进行用电量预测，说明上述方法的科学性和有效性。

1　基于互信息与随机森林算法的用户用电量预测方法原理

基于互信息与随机森林算法的用户用电量预测方法主要内容包括：建立多维评价指标对用户用电特性进行分析，并根据不同用户的用电特性，通过在多个维度进行模糊 C 均值聚类，实现用户的子空间聚类；运用互信息理论对用户用电量数据与潜在关联因素数据进行关联分析，并分别建立对应于不同用户群体的互信息矩阵，辨识与用户用电行为存在强关联关系的因素；基于各类用户的用电量数据及其强关联因素数据，构建训练样本集，经过面向各数据样本的参数寻优和基于随机森林算法的预测建模后，对各用户群体的用电量进行预测。该方法的本质是将所有用户按用电特性进行精细分类，分析与每类用户用电行为相关联的因素，进而针对每类用户分别建立用电量预测模型，实现各类用户以及全体用户的用电量预测。方法的具体实现流程如图 1 所示。

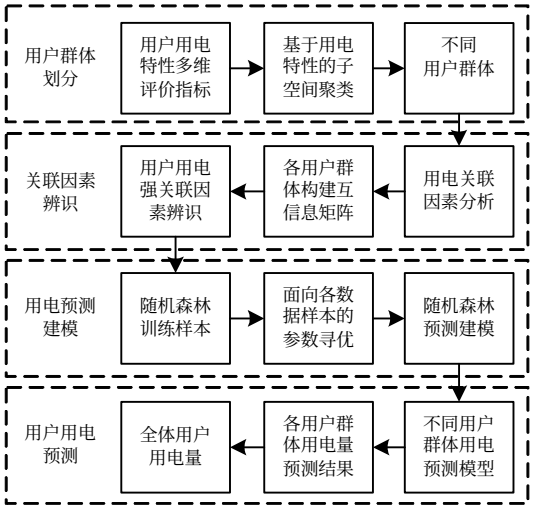


图 1　基于互信息与随机森林算法的用户用电量预测方法流程图

Fig. 1　Flow chart of electricity demand forecasting method based on MI and RF

首先，基于配用电大数据构建表征 m 个用户用电特性的数据集 V_D 和包含 l 种潜在关联因素的数据集 Y_D（由描述多种潜在关联因素的时序数据 $Y_1,Y_2,\ldots,Y_r$ 组成），V_D 中第 j 个用户的用电数据可以表示为 $V_j(\alpha;\beta;\gamma)$（其中 α，β 和 γ 为用户用电特性评价指标），根据 α,β,γ 对所有用户进行子空间聚类，得到 n 个用户群体 $G_k(k=1,\ldots,n)$；其次，分析 $G_k(k=1,\ldots,n)$ 中各用户的用电数据 $V_j(\alpha;\beta;\gamma)$ 与 $Y_1,Y_2,\ldots,Y_r$ 之间的互信息，面向各用户群体形成互信息矩阵 $I_k(k=1,\ldots,n)$，并利用平均互信息辨识与 $G_k(k=1,\ldots,n)$ 中用户的用电特性存在强关联的因素；然后，基于用户用电数据和用电特性强关联因素数据，针对 $G_k(k=1,\ldots,n)$ 形成训练样本 $S_k(k=1,\ldots,n)$，并在参数优化的基础上建立基于随机森林算法的预测模型 $F_k(k=1,\ldots,n)$；最后，将待测用户群体的强关联因素数据输入预测模型 $F_k(k=1,\ldots,n)$ 中，求得各用户群体以及全体用户的用电量预测值。

2 计及多维评价指标的用户用电特性分析

2.1 用户用电特性多维评价指标的选取

根据用电特性的差异，可将不同用户的用电行为划分为多种用电模式，进而基于不同用电模式对用户进行分类，研究各类用户的用电量变化规律。用电模式的判别方法和结果与用电特性指标的选取密切相关，因此，需要定义合理的用电特性评价指标以辅助用电模式识别和用户分类。

文献[13]从时间特征量的角度选取年度用电数据、季节用电数据、日用电类型数据等组成聚类特征向量；文献[14]从负荷特性的角度选取典型日负荷曲线、用户平均负荷率，以及最大负荷率作为电力用户分类的依据；文献[1]从峰谷特性的角度建立了包含峰时耗电率和谷电系数等指标的时间序列特征。本文在综合现有评价指标的基础上，针对不同用户提出了包含时序与非时序数据的用户用电特性多维评价指标：

$$\begin{cases} V_j=\{\alpha_{t1},\alpha_{t2},\cdots,\alpha_{tu};\beta_{t1},\beta_{t2},\cdots,\beta_{tv};\\ \qquad \gamma_1,\gamma_2,\cdots,\gamma_w\}\in V_D \\ j=1,2,\cdots,m \end{cases} \tag{1}$$

式中：$\alpha_{t1},\alpha_{t2},\ldots,\alpha_{tu}$ 和 $\beta_{t1},\beta_{t2},\ldots,\beta_{tv}$ 是时序特征向量，分别代表用户年用电量和月用电量时间序列数据；$\gamma_1,\gamma_2,\ldots,\gamma_w$ 是非时序特征向量，包括年最大负荷利用小时数 γ_1、负荷密度 γ_2、典型日平均负荷率 γ_3、季不均衡系数 γ_4，以及峰谷电量比 γ_5 等负荷特性指标数据。

2.2 用户用电特性子空间聚类分析方法

对描述用户用电特性的多维评价指标进行聚类将有助于不同用电模式的判别。实现聚类算法的一种重要途径是根据特征向量 $\boldsymbol{V}_C$ 的相似性将数据对象进行分类。根据用户用电特性多维评价指标，若将年用电量时间序列 $\boldsymbol{V}_\alpha$、月用电量时间序列 $\boldsymbol{V}_\beta$ 和负荷特性数据 $\boldsymbol{V}_\gamma$ 共同组成描述用户用电特性的特征向量 $\boldsymbol{V}_C=[\boldsymbol{V}_\alpha,\boldsymbol{V}_\beta,\boldsymbol{V}_\gamma]$，则当样本时间跨度为 10 年时，$\boldsymbol{V}_C$ 的维度将高达 10+10×12+5=135 维；即使采用历年平均值作为月用电量数据，$\boldsymbol{V}_C$ 的维度依旧有 10+12+5=27 维。这将给基于用电特性的用户聚类带来很大困难。

在用户聚类过程中，随着 $\boldsymbol{V}_C$ 维度的增加，聚类的时间和空间复杂度将迅速上升，可能导致距离函数失效，带来“维度灾难”[15]。此外，$\boldsymbol{V}_C$ 中包含时序与非时序等多种类型的数据，数据序列长度和数据类型的差异使得距离函数的选择变得困难，低维指标中蕴含的信息有可能被高维指标中的信息“湮没”。虽然以降维为目的属性约简技术可以使 $\boldsymbol{V}_C$ 中具有较少的属性，但是其结果的可解释性较差，且可能会丢失重要的聚类信息，在对高维数据的处理中有一定的局限性[16]。

针对上述问题和现有解决方法中存在的不足，本文采用子空间聚类方法，在用户用电特性数据集 V_D 的不同子空间上寻找簇，进而对不同电力用户进行聚类。根据搜索方向的不同，可将子空间聚类方法分成两大类：“自下而上 (Bottom-up)”搜索策略和“自上而下 (Top-down)”搜索策略[17]。“自下而上”策略是在低维空间中寻找数据密集区域并整合形成簇，但是该策略的本质决定了各个簇之间会有重叠，即一个点可能在零个或多个簇中。“自上而下”策略在初始时将数据集划分为 k 个部分，并为每个部分建立簇，且一个点只能赋给一个簇，这就意味着不会有重复簇产生。

本文基于“自上而下”的子空间聚类搜索策略，建立用电特性多维度解析模型。根据数据类型和指标含义的不同，将用电特性数据集 V_D 划分为 3 个子空间 L_1、L_2 和 L_3，并相应地将原特征向量 $\boldsymbol{V}_C$ 进行拆分。在各子空间中，分别以年用电量时间序列 $\boldsymbol{V}_\alpha$、月用电量时间序列 $\boldsymbol{V}_\beta$，以及负荷特性数据 $\boldsymbol{V}_\gamma$ 作为特征向量，利用模糊 C 均值方法[13]进行聚类。

对用户的年用电量数据序列及月用电量数据序列进行聚类分析时，距离函数选用相关距离；对

负荷特性指标进行聚类分析时，考虑到各指标的量纲不同，距离函数采用标准欧氏距离。相关距离和标准欧氏距离的定义见文献[18]。

在 V_D 的 3 个子空间 L_1、L_2 和 L_3 中，通过模糊 C 均值聚类分别发现 r、s 和 t 个簇，则样本数据点对不同子空间中各个簇的隶属度可以表示为

$$\boldsymbol{U}=\begin{Bmatrix}u_{\alpha,1},\cdots,u_{\alpha,i},\cdots,u_{\alpha,r}\\ u_{\beta,1},\cdots,u_{\beta,j},\cdots,u_{\beta,s}\\ u_{\gamma,1},\cdots,u_{\gamma,k},\cdots,u_{\gamma,t}\end{Bmatrix} \tag{2}$$

式中 $u_{\alpha,i},u_{\beta,j},u_{\gamma,k}\in[0,1]$，且满足 $\sum_{i=1}^{r}u_{\alpha,i}=1$，$\sum_{j=1}^{s}u_{\beta,j}=1$ 和 $\sum_{k=1}^{t}u_{\gamma,k}=1$。

从子空间 L_1、L_2 和 L_3 中分别取出 1 个簇进行融合，所形成的全空间簇可以确定 1 种用户用电模式。在全空间中，根据簇的不同，可以将全体用户的用电特性定义为 $r\times s\times t$ 种用电模式。根据用电模式的不同将用户进行分组，可以分为 $n=r\times s\times t$ 个群体，即 $G_i(i=1,\ldots,n)$。在子空间 L_1、L_2 和 L_3 中，通过最大隶属度筛选样本数据点所在簇 c_1、c_2 和 c_3，则样本数据点对 c_1、c_2 和 c_3 的隶属度构成隶属度矩阵：

$$\boldsymbol{U}_{\max}=\begin{bmatrix}u_{\alpha,\max}\\ u_{\beta,\max}\\ u_{\gamma,\max}\end{bmatrix}=\begin{bmatrix}\max(u_{\alpha,1},\cdots,u_{\alpha,i},\cdots,u_{\alpha,r})\\ \max(u_{\beta,1},\cdots,u_{\beta,j},\cdots,u_{\beta,s})\\ \max(u_{\gamma,1},\cdots,u_{\gamma,k},\cdots,u_{\gamma,t})\end{bmatrix} \tag{3}$$

用户所在群体由 c_1、c_2 和 c_3 融合所组成的全空间簇决定。将用户对其所在群体的隶属度定义为

$$u=\left\|\boldsymbol{U}_{\max}\right\|_2=\sqrt{u_{\alpha,\max}^2+u_{\beta,\max}^2+u_{\gamma,\max}^2} \tag{4}$$

用户对其他群体的隶属度计算方法与式(4)类似。显然，用户对其他群体的隶属度小于 u。

3　面向用户用电关联因素辨识的互信息分析方法

用户用电量与多种社会经济因素存在不同程度的关联关系，而同一种社会经济因素对不同用户群体用电量的影响程度亦存在差异。利用互信息理论[19]对与用户用电相关的社会经济因素进行分析和排序，可以揭示与精确用电量预测最相关的因素，并准确剔除对用电量预测贡献较少的因素，从而降低用电量预测建模的复杂度并提高预测精度。

在用电量关联因素的识别和筛选过程中，将各个用户的用电量数据序列作为解释变量 X，各潜在关联因素数据序列作为条件变量 Y。Y 和 X 之间的互信息大小反映了潜在关联因素与用电量之间的关联程度。为了让各条件变量和解释变量更具有统计学意义，需要对各个变量进行变量域离散化处理，即把各个变量的数值序列转化为概率分布区间[20]。离散化后，解释变量 X 和条件变量 Y 之间的互信息可由下式得出：

$$I(X,Y)=-\sum_{i=1}^{N_i}[\frac{M_i}{M}\log\frac{M_i}{M}]-\{-\sum_{u=1}^{N_j}P(y_u)\cdot[\sum_{v=1}^{N_i}\frac{M_{uv}}{M}\log\frac{M_{uv}}{M}]\} \tag{5}$$

式中：M 为解释变量 X 和条件变量 Y 所有取值的个数和；N_i 为解释变量 X 的区间数量；M_i 为解释变量 X 落在第 i 个区间的数值个数；N_j 为条件变量 Y 的区间数量；$P(y_u)$ 为条件变量 Y 落在第 u 个区间的概率；M_{uv} 为当条件变量 Y 落在第 u 个区间时，解释变量 X 恰好落在第 v 个区间的数值个数。

对于用户群体 $G_k(k=1,\ldots,n)$，假设其中 p 个用户的用电量数据序列构成数据集 $X_D=\{X_1,X_2,\ldots,X_p\}$，$l$ 种潜在关联因素的数据序列构成数据集 $Y_D=\{Y_1,Y_2,\ldots,Y_l\}$，则 $G_k(k=1,\ldots,n)$ 中各用户用电量与各潜在关联因素之间的互信息可表示为

$$\boldsymbol{I}_k=\begin{bmatrix}I(X_1,Y_1)&\cdots&I(X_1,Y_j)&\cdots&I(X_1,Y_l)\\ \vdots&\ddots&\vdots&\ddots&\vdots\\ I(X_i,Y_1)&\cdots&I(X_i,Y_j)&\cdots&I(X_i,Y_l)\\ \vdots&\ddots&\vdots&\ddots&\vdots\\ I(X_p,Y_1)&\cdots&I(X_p,Y_j)&\cdots&I(X_p,Y_l)\end{bmatrix},\quad k=1,\cdots,n \tag{6}$$

式中：$X_i\in X_D$；$Y_j\in Y_D$；Y_j 与 $X_1,X_2,\ldots,X_p$ 之间互信息的平均值，即平均互信息，可以表示为

$$\overline{I(X,Y_j)}=\frac{1}{p}\sum_{i=1}^{p}I(X_i,Y_j),\quad j=1,\cdots,l \tag{7}$$

对用户群体 $G_k(k=1,\ldots,n)$，可利用平均互信息评价潜在关联因素 Y_j 与用户用电量之间的关联关系强弱：平均互信息越大，二者之间的关联性越强[5]。对平均互信息大于 0 的关联因素进行排序，形成关联因素列表，选取列表中排名靠前的强关联因素，与用户用电量数据一起构建训练样本集 $S_k(k=1,\ldots,n)$，用于用户群体 $G_k(k=1,\ldots,n)$ 的用电量预测建模。

4　基于随机森林算法的用电量预测模型

在获得多种关联因素的基础上，利用随机森林算法[21]对各用户群体开展用电量预测建模，建模过

程如图 2 所示。首先，对用户群体 $G_k(k=1,\ldots,n)$，利用 Bootstrap 方法[22]从原始训练样本集 $S_k(k=1,\ldots,n)$中随机抽取多个训练样本子集，对每个子集分别进行决策树建模，然后利用测试集对各决策树进行测试，综合多棵决策树的测试结果，通过投票得出最终的用电量预测模型。

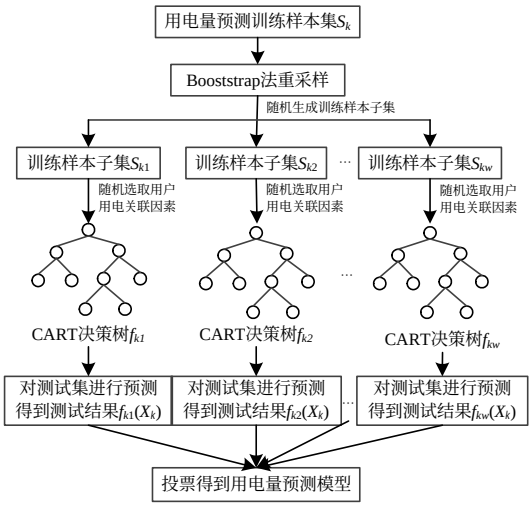


图 2 基于随机森林算法的用电量预测建模过程

Fig. 2 Modeling process of electricity demand forecasting based on RF

4.1 训练样本子集的随机选取

原始训练样本集 S_k 由两类数据构成：一类为 S_k 中用户总用电量时序数据，作为预测模型的输出；另一类为与之对应的 M 种关联因素的时序数据，作为预测模型的输入。利用 Bootstrap 抽样方法从 S_k 中随机选取 w 个训练样本子集 S_{k1}，S_{k2}，…，S_{kw}(各子集都包含上述两类数据)，用于构建 w 棵分类回归树 (classification and regression tree，CART)。由于在训练样本子集抽取过程中采用有放回抽样法进行采样，因此 S_k 中约 37%的样本不会出现在采集的样本集合中，这些数据称为袋外(out of bag，OOB)数据[23]。以 OOB 数据作为测试集，对 CART 决策树的误差进行估计。将 w 棵决策树的误差估计取平均，可以得到随机森林的泛化误差估计值，并以此对用电量预测模型的精度进行量化度量。

4.2 CART 决策树构建

对每一个训练样本子集，以 Gini 系数最小为原则，采用 CART 算法生成一棵决策树，共生成 w 棵决策树，从而形成“森林”[23]。为保证决策树构建时的随机性，避免过拟合问题，在每一棵决策树构建时，从 M 种用户用电量关联因素中随机选取 F 种作为随机特征变量，参与决策树节点分裂过程，其中 F 取小于等于 $\log_2(M+1)$的最大正整数[24]。此外，整个随机森林中决策树的棵数 w 需根据预测结果进行调整。

4.3 用电量预测结果投票

当 w 棵 CART 决策树构建完成后，利用测试集数据进行仿真。将测试集中与用电量 Y_k 相关的关联因素数据 X_k 作为输入，得到各决策树模型的预测结果序列$\{f_{k1}(X_k),\ f_{k2}(X_k),\ldots,f_{kw}(X_k)\}$。基于随机森林算法的预测模型最终输出的用电量预测结果采用投票方式产生[22]：

$$F_k(X_k)=\arg\max_{Y_k}\sum_{i=1}^{w}I(f_{ki}(X_k)=Y_k)\quad k=1,2,\cdots,n \tag{8}$$

式中：F_k 为面向用户群体 G_k 的用电量组合预测模型；f_{ki} 为单棵决策树预测模型；$I(\circ)$为示性函数。将各用户群体的用电量预测模型 F_k 进行线性组合，即可得到全体用户的总用电量预测模型。

4.4 用电量预测模型对大数据环境的适应性

基于随机森林算法的用电量预测模型利用 Bootstrap 方法对原始训练样本集进行随机采样的过程以及大量决策树分裂时随机抽取输入变量的过程，彼此都是相互独立的，这保证了该预测模型能通过并行化计算来提高大数据预测的效率[25]。此外，当存在大量用电量关联因素时，该预测模型可以通过在各决策树分裂过程中随机抽取输入变量的方式，降低模型的计算复杂度和空间复杂度，从而极大地提高模型运行效率，因此具有适应大数据处理环境的优势。

5 算例分析

5.1 数据源

用户用电量数据来自上海某区域 5360 家用户的月度用电信息等，时间跨度为 2005 年 1 月至 2014 年 12 月。对于用户年最大负荷利用小时数γ_1、负荷密度γ_2、典型日平均负荷率γ_3、季不均衡系数γ_4，以及峰谷电量比γ_5等负荷特性指标数据，通过用户用电量、占地面积和典型日负荷曲线等数据计算得到。利用用户用电量时序数据和负荷特性指标数据，构建用户用电特性数据集 V_D。

5360 家用户所属行业涵盖工业，交通运输、仓储和邮政业，商业、住宿和餐饮业等 8 大类行业。将上海市该 8 类行业的行业总产值、行业利润总额、行业固定资产投资和行业景气指数，共 4×8=32 种因素，作为用户用电量潜在关联因素。

在该区域 5360 家用户中制造类企业所占比重较大，根据文献[26]，上述制造类企业共涵盖制造

业下属的 20 个子行业。将这 20 个子行业在上海市的主要产品产量、主要原材料价格指数、主要产品出厂价格指数、主要产品出口价格指数和产品库存量，共 5×20=100 种因素，也作为用户用电量潜在关联因素。

此外，还要考虑与用户所处区域相关的因素，如常住人口数、人均可支配收入、社会消费品零售总额、总 GDP、生产价格指数、居民消费价格指数、第一与第二和第三产业 GDP、进出口贸易额、固定资产投资额、商务楼宇面积、房地产新开工面积、土地交易价格、道路总里程数、电力价格指数、月平均气温和季节指标，共 18 种因素。

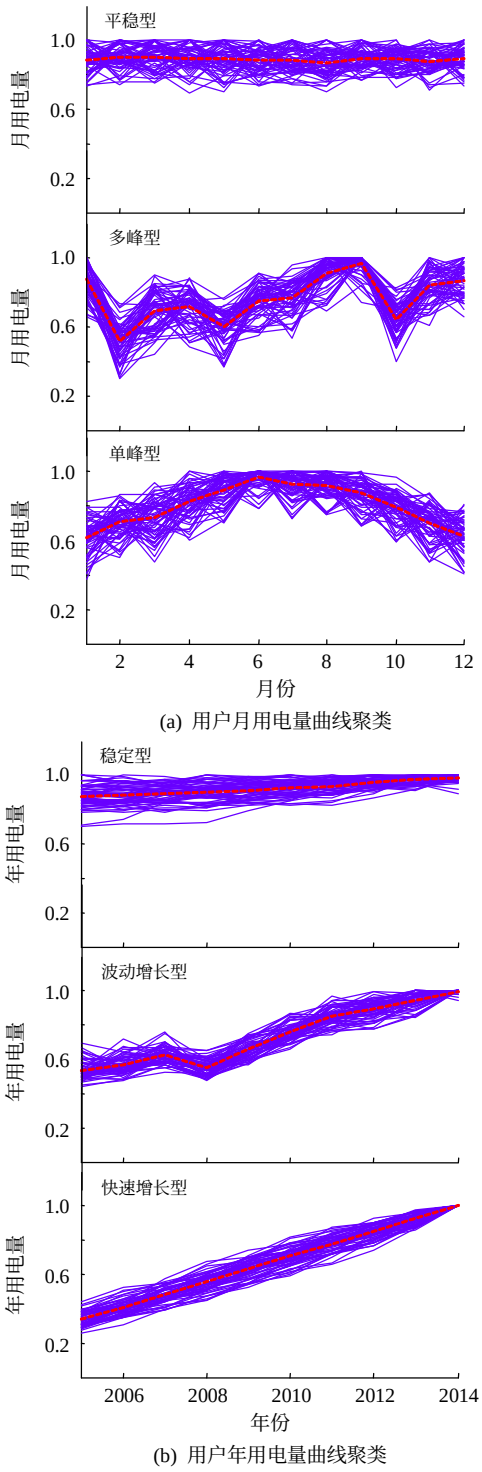
将上述 32+100+18=150 种因素作为用户用电量潜在关联因素，构建潜在关联因素数据集 Y_D。将用户用电量和潜在关联因素的月度数据进行归一化，结果见附表 A1。

5.2　用户用电特性聚类分析

用于聚类的用户用电特性数据集 V_D 包括 3 类数据：1）2005—2014 年年用电量时序数据，作为 10 维用电量趋势指标，用于子空间 L_1 的聚类；2）2005—2014 年各月份月度用电量的同期平均值，作为 12 维用电量波动指标，用于子空间 L_2 的聚类；3）年最大负荷利用小时数 γ_1、负荷密度 γ_2、典型日平均负荷率 γ_3、季不均衡系数 γ_4，以及峰谷电量比 γ_5，作为 5 维负荷特性指标，用于子空间 L_3 的聚类。

采用模糊 C 均值方法，初始化聚类数目 c=3，模糊程度系数 m=2，分别在子空间 L_1、L_2 和 L_3 中展开聚类，得到用电量及负荷特性的聚类结果分别如图 3 和表 1 所示。

在图 3 中，根据子空间 L_1 和 L_2 中的聚类结果，将用户的月用电量及年用电量曲线分别表示为 3 类曲线簇，虚线代表其聚类中心。图 3(a)中各图分别代表平稳型、多峰型和单峰型的月用电量曲线簇：平稳型主要对应化学制品制造业等用电连续



(a) 用户月用电量曲线聚类

(b) 用户年用电量曲线聚类

图 3　用电量曲线聚类簇

Fig. 3　Clusters of electricity consumption curves

表 1　各类负荷特性指标的聚类结果

Tab. 1　Clustering results of different load characteristic indexes

负荷特性指标	指标分段区间		
	第一类	第二类	第三类
γ_1	[4294,6748]	[1514,5120]	[568,2639]
γ_2	[71.65,204.24]	[52.96,305.43]	[9.25,98.46]
γ_3	[0.7524,0.9626]	[0.6152,0.8002]	[0.5171,0.7345]
γ_4	[0.6292,0.8224]	[0.5726,0.7399]	[0.7194,0.9271]
γ_5	[0.2083,1.4629]	[0.9074,1.8420]	[0.8412,1.2174]

性较强的用户；多峰型涵盖受节假日、生产周期和季节特征等多重因素影响的用户；单峰型主要包括在年中的 4—9 月间出现用电高峰的用户。图 3(b)中各图分别表示稳定型、波动增长型和快速增长型的年用电量曲线簇：稳定型主要包括用电量缓慢增长、不增长和逐渐下降的用户；波动增长型涵盖部分出口导向性较强的用户，该类用户的用电量在一定程度上受到 2008 年国际金融危机的影响；快速

增长型主要对应于各行业中发展态势较好的用户。

根据子空间 L_3 中的聚类结果，将不同用户在该子空间中的用电模式分为 3 种类型，各类型用电模式中负荷特性指标所覆盖的数值区间如表 1 所示。

在表 1 中，第一类子空间用电模式具有较高的最大负荷利用小时数和负荷密度，主要对应于重工业、电力和水供应业等高耗能行业的用户；第二类子空间用电模式具有较小的季不均衡系数和较大的峰谷电量比，说明用户用电行为对于季节因素和分时电价较为敏感，该类型用电模式主要对应于轻工业、金融、商务和餐饮住宿等行业的用户；第三类子空间用电模式中，年最大负荷利用小时数、负荷密度和典型日平均负荷率均较小，该类型用电模式主要对应于公共事业单位和管理组织等能耗较低的用户。

在每个子空间中，根据聚类结果将用户在该子空间中的用电模式划分为 3 种类型，用户对每种子空间用电模式的隶属度区间为[0,1]，选取隶属度最大的类型作为用户在该子空间中的用电模式；为便于数据可视化，将用户对每个子空间中 3 种用电模式的隶属度依次变换到[0,1]、[1,2]和[2,3]。在子空间 L_1、L_2 和 L_3 组成的全空间中，可将用户用电模式组合为 3×3×3=27 种，用户对不同的全空间用电模式的隶属度可通过式(4)进行计算，以隶属度最大的类型作为用户在全空间中的用电模式。

以年用电量聚类中各用户对其子空间用电模式的隶属度为 x 轴，以月用电量聚类中各用户对其子空间用电模式的隶属度为 y 轴，以负荷特性聚类中各用户对其子空间用电模式的隶属度为 z 轴，建立电力用户子空间聚类坐标系。以用户对其所属的不同子空间用电模式的隶属度为坐标，可得 5360 家电力用户的子空间聚类效果如图 4 所示。根据所属全空间用电模式的不同，将所有用户划分为 27 个群体，每个用户群体对应 1 种全空间用电模式，

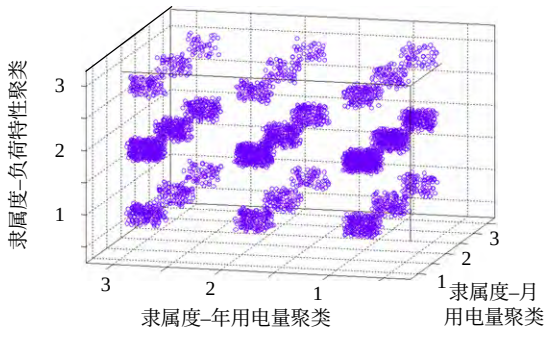


图 4 电力用户子空间聚类效果

Fig. 4 Subspace clustering effect of electricity users

27 个用户群体分别包含的用户数量如图 5 所示。

由图 4、图 5 可知，不同用户群体中的用户数量存在一定差异。其中，第 23 个用户群体(属于年用电量–快速增长型、月用电量–多峰型、负荷特性指标–第二类)的用户数量最多，达到 645 家。此外，在负荷特性聚类中，第二类子空间用电模式(图 4 中 z 轴，隶属度区间[1,2])对应的用户数量明显多于其他用电模式，这表明样本中对季节因素和分时电价较为敏感的用户所占比重较大。

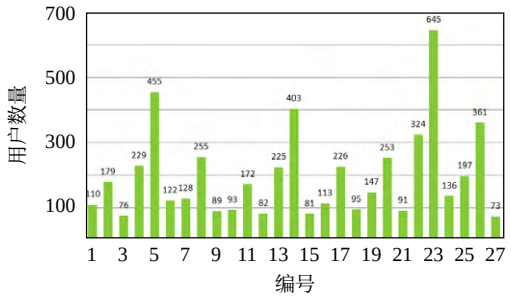


图 5 隶属于各用户群体的用户数量

Fig. 5 Number of users affiliated with different groups

5.3 用户用电关联因素辨识

隶属于不同用电模式的用户，其用电量关联因素会存在差异。为简明起见，本文以图 5 中第 20 个用户群体(属于年用电量–快速增长型、月用电量–单峰型、负荷特性指标–第二类)为例，说明用电关联因素辨识及用电量预测建模的全过程。

将第 20 个用户群体中 253 个用户的月度用电量数据作为解释变量，150 种潜在关联因素的月度数据作为条件变量，分析解释变量与条件变量之间的互信息，部分结果如图 6 所示。

在图 6 中，每行代表一个用户，每列代表一种潜在关联因素，每个色块的颜色表示用户用电量与潜在关联因素的互信息值，互信息值越大，说明用

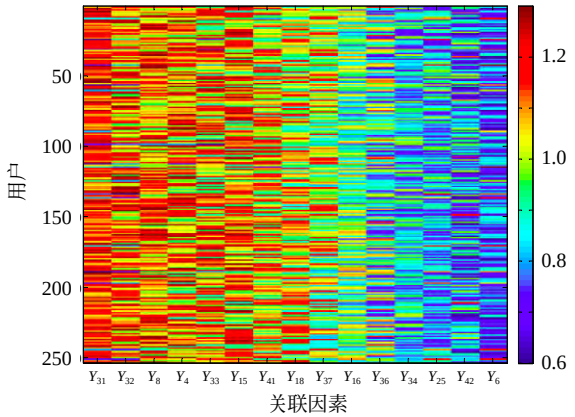


图 6 用户用电量与关联因素的互信息分析结果

Fig. 6 Results of mutual information analysis between users' electricity consumption and relation factors

户用电量与该因素的关联程度越高。如果只分析单个用户的用电量与各潜在关联因素之间的互信息值，可以发现每一行的色块分布都存在个性化差异；若将众多用户的分析结果进行整合，则能从色块图的整体颜色分布中捕获关于用电量与各因素关联关系的共性特征，进而确定影响该用户群体用电量的强关联因素。根据式(7)求得各因素与用户用电量的平均互信息，并选取列表中排名前 15 的强关联因素，结果见表 2。

表 2　关联因素的平均互信息值

Tab. 2　Average mutual information of relation factors

因素	平均互信息	因素	平均互信息	因素	平均互信息
Y_{31}	1.1453	Y_{15}	1.0897	Y_{36}	0.8913
Y_{32}	1.1185	Y_{41}	1.0722	Y_{34}	0.8543
Y_8	1.1078	Y_{18}	1.0448	Y_{25}	0.8228
Y_4	1.1061	Y_{37}	1.0050	Y_{42}	0.8011
Y_{33}	1.0930	Y_{16}	0.9236	Y_6	0.7688

在表 2 中，15 种强关联因素包括：用户所在区域的 GDP(Y_4)、居民消费价格指数(Y_6)、第二产业 GDP(Y_8)、道路总里程数(Y_{15})、电力出厂价格指数(Y_{16})，以及季节指标(Y_{18})；交通运输、仓储和邮政业的行业固定资产投资(Y_{25})；交通运输、电气、电子设备制造业的行业总产值(Y_{31})、行业利润总额(Y_{32})、行业固定资产投资(Y_{33})，以及行业景气指数(Y_{34})；信息传输、计算机服务和软件业的行业利润总额(Y_{36})和行业固定资产投资(Y_{37})；电力、燃气及水的生产和供应业的行业固定资产投资(Y_{41})和行业景气指数(Y_{42})。由表 2 可知，第 20 个用户群体的用电量与交通运输、电气、电子设备制造业的多种因素(Y_{31}、Y_{32}、Y_{33}、Y_{34})关联性较强，且受固定资产投资类因素(Y_{25}、Y_{33}、Y_{37}、Y_{41})的影响较大。

5.4　用户用电量预测

以 15 种强关联因素的月度数据作为输入，以第 20 个用户群体的总用电量月度数据作为输出，形成原始训练样本集 S_{20}，进而建立基于随机森林算法的用电量预测模型。采用随机森林算法进行用电量预测时，若预测值超过了训练样本集的数值范围，预测精度会大打折扣。为保证预测的稳定性，本文将强关联因素和月用电量数据转化为月度同比增长率，作为预测模型的输入和输出。

利用 Bootstrap 方法从 S_{20} 中随机选取 w 个训练样本子集，用于生成 w 棵决策树；剩余的 OOB 数据作为测试集对随机森林预测模型的误差进行估计。每棵决策树生成时随机选取 $\log_2(15+1)=4$ 种强关联因素作为随机特征变量，参与节点分裂过程。

图 7 显示了当 w 取不同值时，基于随机森林算法预测模型的预测值与实际值之间的平均绝对百分误差(mean absolute percentage error，MAPE)。整体而言，预测模型的 MAPE 会随着 w 的增加而减小。但当 w 取值较大时，过细的分类会导致计算量迅速增加。在综合考虑建模速度和预测误差的情况下，本文取 w=150 作为最佳决策树棵数。

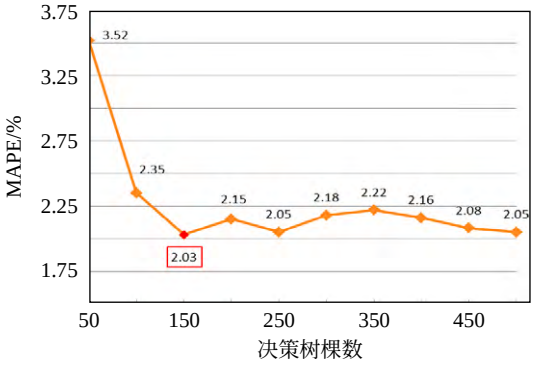


图 7　不同森林规模下的预测误差

Fig. 7　Forecasting error of RF with different sizes

通过平均准确度下降率来度量各种强关联因素对预测模型的重要性。所谓平均准确度下降率，是指当模型去除一个关联因素后预测准确度的下降程度，指标值越大说明该关联因素越重要。15 种关联因素的重要性程度如图 8 所示，竖线左侧部分面积可理解为关联因素对预测精度的累积贡献。由图 8 可知，Y_{31}、Y_{32}、Y_4、Y_{33} 和 Y_8 的重要性较高，这与表 2 中的分析结果基本吻合。

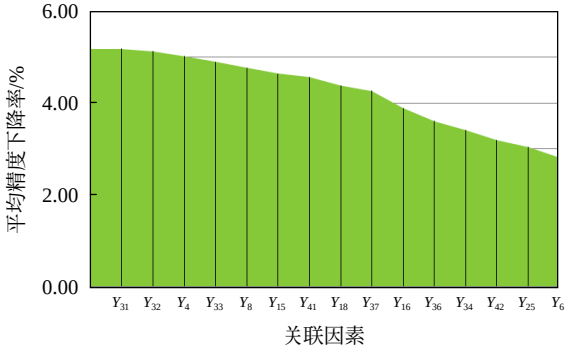


图 8　不同关联因素对预测精度的影响程度

Fig. 8　Impact of different relation factors on forecasting accuracy

利用随机森林模型进行预测，在得到用电量月度同比增长率的基础上，以上一年同期的月用电量为基准得到该月用电量预测值。为比较不同模型的预测能力，基于相同的训练样本集，建立支持向量机(support vector machine，SVM)预测模型，并与随机森林模型的预测结果进行对比，结果见表 3。

表 3 不同模型的预测结果比对

Tab. 3 Forecasting results of different models

组别	实际用电量/(10^6kW·h)	SVM 预测值/(10^6kW·h)	SVM APE/%	RF 预测值/(10^6kW·h)	RF APE/%
1	65.39	63.98	2.16	63.99	2.14
2	116.28	113.38	2.49	113.85	2.09
3	131.18	125.41	4.40	130.97	0.16
4	138.04	133.5	3.29	136.32	1.25
5	125.40	118.39	5.59	122.3	2.47
6	87.15	84.73	2.78	84.59	2.94
MAPE/%		3.45		1.84	

表 3 中列出了第 20 个用户群体在某 6 个月(对应表中的组别)的实际用电量以及 SVM 模型和随机森林模型的预测值。在 6 组数据中，除第 6 组外，随机森林模型的绝对百分误差(absolute percentage error，APE)均小于 SVM 模型；从整体上看，随机森林模型的 MAPE 明显小于 SVM 模型，具有更高的预测精度。

为进一步说明随机森林预测模型的有效性，采用上述方法对 27 个用户群体分别进行预测建模，用电量预测的误差分布如图 9 所示。图 9 显示，对于 27 个用户群体，随机森林预测模型在统计意义上具有更高的预测精度。

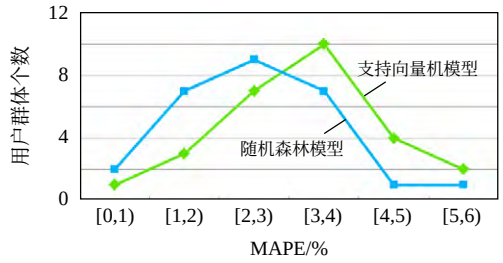


图 9 用电量预测误差分布图

Fig. 9 Distribution map of electricity demand forecasting errors

6 结论

1）提出了一种面向海量用户的用电特性子空间聚类分析方法，结合用电特性多维评价指标及特征向量，对用户用电特性数据集进行子空间划分，采用自上而下的子空间搜索策略提取出多种子空间用电模式，并根据用电模式的差异性对用户进行多维度解析和群体划分，从而拓展了现有的用户用电行为分析方法。

2）运用互信息理论对不同用户群体的用电关联因素进行判别，挖掘出各种潜在关联因素和用户用电量之间的关联关系，以强关联因素数据作为输入建立基于随机森林算法的用户用电量组合预测模型，实现了用电预测全过程的数据驱动，从而为用户用电强关联因素的筛选以及用电预测建模提供了一种新思路。

3）考虑到智能电网大数据环境的逐步完善以及大数据处理技术的不断发展，下一步的工作准备对更大规模数据集进行分析，在算法效率优化及分布式实现的基础上，深入挖掘用电预测相关大数据本身的特性。

参考文献

[1] 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(in Chinese).

[2] 赵岩，李磊，刘俊勇，等. 上海电网需求侧负荷模式的组合识别模型[J]. 电网技术，2010，34(1)：145-151.
Zhao Yan，Li Lei，Liu Junyong，et al. Combinational recognition model for demand side load profile in Shanghai power grid[J]. Power System Technology，2010，34(1)：145-151(in Chinese).

[3] 赵腾，张焰，张东霞. 智能配电网大数据应用技术与前景分析[J]. 电网技术，2014，38(12)：3305-3312.
Zhao Teng，Zhang Yan，Zhang Dongxia. Application technology of big data in smart distribution grid and its prospect analysis[J]. Power System Technology，2014，38(12)：3305-3312(in Chinese).

[4] 王继业，季知祥，史梦洁，等. 智能配用电大数据需求分析与应用研究[J]. 中国电机工程学报，2015，35(8)：1829-1836.
Wang Jiye，Ji Zhixiang，Shi Mengjie，et al. Scenario analysis and application research on big data in smart power distribution and consumption systems [J]. Proceedings of the CSEE，2015，35(8)：1829-1836(in Chinese).

[5] Han Y，Sha X，Grover-Silva E，et al. On the impact of socio-economic factors on power load forecasting [C]//2014 IEEE International Conference on Big Data. Washington，DC，USA：IEEE，2014：742-747.

[6] 朱陶业，李应求，张颖，等. 提高时间序列气象适应性的短期电力负荷预测算法[J]. 中国电机工程学报，2006，26(23)：14-19.
Zhu Taoye，Li Yingqiu，Zhang Ying，et al. A new algorithm of advancing weather adaptability based on arima model for day-ahead power load forecasting [J]. Proceedings of the CSEE，2006，26(23)：14-19(in

Chinese).

[7] 李瑾，刘金朋，王建军．采用支持向量机和模拟退火算法的中长期负荷预测方法[J]. 中国电机工程学报, 2011, 31(16)：63-66.
Li Jin，Liu Jinpeng，Wang Jianjun．Mid-long term load forecasting based on simulated annealing and SVM algorithm[J]．Proceedings of the CSEE，2011，31(16)：63-66(in Chinese).

[8] 牛东晓，陈志业，邢棉，等．具有二重趋势性的季节型电力负荷预测组合优化灰色神经网络模型[J]. 中国电机工程学报，2002，22(1)：29-32.
Niu Dongxiao，Chen Zhiye，Xing Mian，et al．Combined optimum gray neural network model of the seasonal power load forecasting with the double trends [J]．Proceedings of the CSEE，2002，22(1)：29-32(in Chinese).

[9] Kolter J Z，Ferreira Jr J．A large-scale study on predicting and contextualizing building energy usage[C]// Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence．San Francisco，California，USA：Association for the Advancement of Artificial Intelligence，2011：1349-1356.

[10] Aman S，Simmhan Y，Prasanna V K．Improving energy use forecast for campus micro-grids using indirect indicators[C]//2011 11th IEEE International Conference on Data Mining Workshops．Vancouver，BC，Canada：IEEE，2011：389-397.

[11] 王鹏，陈启鑫，夏清，等．应用向量误差修正模型的行业电力需求关联分析与负荷预测方法[J]. 中国电机工程学报，2012，32(4)：100-107.
Wang Peng，Chen Qixin，Xia Qing，et al．Correlation analysis and forecasting method on industrial electricity demand based on vector error correction model [J]．Proceedings of the CSEE，2012，32(4)：100-107(in Chinese).

[12] 马瑞，周谢，彭舟，等．考虑气温因素的负荷特性统计指标关联特征数据挖掘[J]．中国电机工程学报，2015，35(1)：43-51.
Ma Rui，Zhou Xie，Peng Zhou，et al．Data mining on correlation feature of load characteristics statistical indexes considering temperature[J]．Proceedings of the CSEE，2015，35(1)：43-51(in Chinese).

[13] 杨浩，张磊，何潜，等．基于自适应模糊 C 均值算法的电力负荷分类研究[J]．电力系统保护与控制，2010，38(16)：111-115.
Yang Hao, Zhang Lei, He Qian, et al. Study of power load classification based on adaptive fuzzy C means[J]．Power System Protection and Control，2010，38(16)：111-115(in Chinese).

[14] 薛承荣，顾洁，赵建平，等．基于用户用电特性及供电成本分摊的销售侧电价机制研究[J]．华东电力，2014，42(1)：168-173.
Xue Chengrong，Gu Jie，Zhao Jianping，et al．Electricity retail tariff mechanism based on customers' electrical characteristics and cost apportionment[J]．East China Electric Power，2014，42(1)：168-173(in Chinese).

[15] 牛琨，张舒博，陈俊亮．采用属性聚类的高维子空间聚类算法[J]．北京邮电大学学报，2007，30(03)：1-5.
Niu Kun，Zhang Shubo，Chen Junliang．Subspace clustering through attribute clustering[J]．Journal of Beijing University of Posts and Telecommunications，2007，30(03)：1-5(in Chinese).

[16] 甘杨兰．面向高维数据的子空间聚类算法研究[D]．合肥：合肥工业大学，2007.
Gan Yanglan. The research on subspace clustering for high dimensional data[D]．Hefei：Hefei University of Technology，2007(in Chinese).

[17] Parsons L，Haque E，Liu H．Subspace clustering for high dimensional data：a review[J]．ACM SIGKDD Explorations Newsletter，2004，6(1)：90-105.

[18] 薛毅，陈立萍．统计建模与 R 软件 [M]．北京：清华大学出版社，2007： 397-403.
Xue Yi，Chen Liping．Statistical modeling and software R[M]．Beijing：Tsinghua University Press，2007：397-403(in Chinese).

[19] Vahabie A H，Yousefi M M R，Araabi B N，et al．Mutual information based input selection in neuro-fuzzy modeling for short term load forecasting of Iran national power system[C]//2007 IEEE International Conference on Control and Automation. Guangzhou, China：IEEE, 2007：2710-2715.

[20] 原媛，顾洁，黄薇，等．互信息在电力系统中长期负荷预测中的应用[J]．华东电力，2009，37(2)：236-239.
Yuan Yuan，Gu Jie，Huang Wei，et al．Application of mutual information to power system medium and long-term load forecasting[J]．East China Electric Power，2009，37(2)：236-239(in Chinese).

[21] Breiman L．Random forests[J]. Machine Learning, 2001, 45(1)：5-32.

[22] 方匡南，吴见彬，朱建平，等．随机森林方法研究综述[J]．统计与信息论坛，2011，26(3)：32-38.
Fang Kuangnan, Wu Jianbin, Zhu Jianping, et al. A review of technologies on random forests[J]．Statics & Information Forum，2011，26(3)：32-38(in Chinese).

[23] 曹正凤. 随机森林算法优化研究[D]. 北京：首都经济贸易大学，2014.
Cao Zhengfeng. Study on optimization of random forests algorithm[D]. Beijing：Capital University of Economics and Business, 2014(in Chinese).

[24] 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(in Chinese).

[25] 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(03)：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(03)：527-537(in Chinese).

[26] 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. GB/T 4754-2011 国民经济行业分类[S]. 北京：中国标准出版社，2011.
General Administration of Quality Supervision Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. GB/T 4754-2011 Industrial classification for economic activities[S]. Beijing：Standard Press of China, 2011(in Chinese).

附录 A

表 A1 归一化的用户用电量与潜在关联因素月度数据

Tab. A1 Normalized monthly data of users' electricity consumption and potential relation factors

用户	2005.1	2005.2	2005.3	2005.4	2005.5	2005.6	…	2014.12
X_1	0.535	0.510	0.537	0.528	0.525	0.539	…	0.978
X_2	0.865	0.822	0.927	0.913	0.929	0.951	…	0.887
X_3	0.687	0.644	0.675	0.689	0.690	0.687	…	0.895
X_4	0.836	0.820	0.669	0.639	0.739	0.775	…	0.840
X_5	0.826	0.771	0.800	0.717	0.695	0.785	…	0.835
X_6	0.587	0.564	0.534	0.506	0.597	0.601	…	0.935
X_7	0.549	0.512	0.523	0.545	0.557	0.558	…	0.987
X_8	0.613	0.602	0.610	0.618	0.620	0.624	…	0.995
X_9	0.689	0.667	0.694	0.706	0.698	0.701	…	1
…	…	…	…	…	…	…	…	…
X_{5358}	0.535	0.510	0.537	0.528	0.525	0.539	…	0.978
X_{5359}	0.865	0.822	0.927	0.913	0.929	0.951	…	0.887
X_{5360}	0.687	0.644	0.675	0.689	0.690	0.687	…	0.895
潜在关联因素	2005.1	2005.2	2005.3	2005.4	2005.5	2005.6	…	2014.12
Y_1	0.656	0.657	0.657	0.660	0.660	0.662	…	0.980
Y_2	0.534	0.540	0.542	0.538	0.545	0.550	…	0.995
Y_3	0.565	0.583	0.549	0.530	0.572	0.586	…	0.872
Y_4	0.425	0.430	0.432	0.437	0.446	0.460	…	0.975
Y_5	0.874	0.875	0.883	0.875	0.890	0.882	…	0.954
Y_6	0.624	0.602	0.613	0.625	0.630	0.628	…	0.950
Y_7	0.573	0.584	0.582	0.594	0.601	0.595	…	0.998
Y_8	0.592	0.536	0.521	0.578	0.643	0.684	…	0.836
Y_9	0.784	0.795	0.643	0.663	0.678	0.799	…	0.864
…	…	…	…	…	…	…	…	…
Y_{148}	0.765	0.789	0.790	0.802	0.810	0.823	…	1
Y_{149}	0.102	0.154	0.256	0.307	0.578	0.893	…	0.154
Y_{150}	1	1	0.25	0.25	0.25	0.5	…	1

收稿日期：2015-08-17。

作者简介：


赵腾

赵腾(1990)，男，博士研究生，主要研究方向为大数据在智能电网中的应用，zhaoteng@sjtu.edu.cn；

王林童(1992)，男，硕士研究生，主要研究方向为大数据在智能电网中的应用；

张焰(1958)，女，博士，教授，博士生导师，主要研究方向为电力系统规划、电力系统可靠性、大数据分析等；

田世明(1965)，男，教授级高级工程师，主要研究方向为智能电网、需求侧管理、大数据分析等。

(责任编辑 李泽荣)