第 41 卷　第 3 期　2017 年 2 月 10 日
DOI：10.7500/AEPS20160507003
电力系统自动化
Automation of Electric Power Systems
Vol.41 No.3 Feb.10，2017

考虑地域差异的配电网空间负荷聚类及一体化预测方法

刘　思[1]，傅旭华[2]，叶承晋[2]，黄民翔[1]
（1. 浙江大学电气工程学院，浙江省杭州市 310027；2. 国网浙江省电力公司经济技术研究院，浙江省杭州市 310008）

摘要：针对基于智能算法的负荷密度指标法对样本依赖性强且在各地实际应用困难的不足，提出一种考虑地域差异的配电网空间负荷聚类及一体化预测方法。该方法首先通过大量调研得到分布在不同地区、分属不同类型的负荷样本及所处地区信息；然后利用基于日负荷曲线的负荷分类校验及精选方法对所有调研样本进行分类精选；再根据区域分类、负荷分类对精选样本构成的全样本空间进行两级划分，得到分层级子样本空间；最后根据待测地块的属性信息对子样本空间进行匹配，选取与其最相似的子样本空间作为训练样本，构建支持向量机模型预测各地块的负荷密度，进而得到电力负荷的空间分布。工程实例分析表明了该方法的实用性和有效性。

关键词：空间负荷预测；负荷密度指标法；地域差异；聚类分析；支持向量机

0　引言

分布式能源的广泛接入，电动汽车等可控负荷的大量投入，要求配电网具备主动协调、灵活控制的能力[1]，对配电网规划设计、保护控制、优化调度[2-4]等方面提出更高要求。空间负荷预测是指对供电区内未来电力负荷大小和位置的预测，是城市配电网规划的前提和基础[5]。用地仿真法[6]和负荷密度指标法[7]是两种代表性预测方法。前者假定各类用地面积和分类负荷密度已知，核心是用地决策；后者假定用地决策已知，核心是分类负荷密度的确定[8]。

人工智能和数据挖掘等技术在空间负荷预测中的应用，进一步提升了上述方法的性能。文献[9]在采用元胞自动机建模的基础上，利用云理论表现负荷影响因素的随机性和模糊性；文献[10]通过对元胞根据街路距离、馈线负荷、用地信息等属性进行多级聚类，充分利用负荷与元胞属性之间的联系来改善预测效果；文献[11]基于空间异质性对元胞空间进行不同尺度下的不规则区域划分，通过叠加每一尺度下的结果来预测空间负荷的分布，提高了空间负荷预测的准确率；文献[12]建立了基于多 Agent 技术的负荷预测模型，将各元胞与影响因素均看作单个 Agent，通过各 Agent 的并行计算进行负荷预测，提高了预测精度与计算效率；文献[13]采用克里金插值法进行空间负荷预测，避免预测精度受通信故障影响。

基于双层贝叶斯分类模型、模糊理论、自适应神经模糊推理系统（ANFIS）、最小支持向量机（support vector machine，SVM）等智能算法的负荷密度指标法[14-19]的出现，从本质上改善了传统方法依靠规划人员的主观经验确定负荷密度因而预测精度不高[20]的不足。该类方法的主要思路[8]是首先通过大量调研，对负荷按性质精细分类，并按类确定影响负荷密度的主要因素，形成分类的负荷密度标准样本集；然后在此基础上，基于不同智能算法构建分类预测模型；最后根据待测样本的性质和特征对其正确分类，用该类模型预测其负荷密度。

现有研究偏重预测方法的理论创新，缺乏对调研样本数据质量的合理检验。在样本挑选（即负荷分类）过程中，通常根据电力公司给出的用户名称或调研人员的主观判断对用户进行分类，缺乏科学性。首先，电力公司给出的用户名称不一定与该用户的实际行业归属一致；其次，某些样本的用电行为无显著行业特征，缺乏代表性。未经校验和精选的将上述样本作为预测模型的训练样本，将严重影响算法的性能。

用电负荷受当地经济及产业发展情况影响[21]，各地负荷密度指标存在较大的地域差异，同一套标准或方法不存在区域间不经调整直接借鉴移植的可行性。现有研究未考虑地域差异对空间负荷预测的影响，对样本依赖性强，各地分别对当地负荷样本进行研究又极其费时费力，存在明显不足。

针对以上问题，本文提出一种考虑地域差异的

收稿日期：2016-05-07；修回日期：2016-08-24。
上网日期：2016-10-31。
国家电网公司科技项目（5211JY150016）。

配电网空间负荷聚类及一体化预测方法。以浙江电网为例，对分布杭州、宁波等 10 个城市，分属工业、商业、居住、行政办公等 4 种类型的电力用户进行广泛调研，并以湖州市某地块为预测对象，对本文所提方法进行实例检验，结果表明该方法实用、有效，满足精度要求。

1　整体思路及实现流程

考虑地域差异的配电网空间负荷聚类及一体化预测方法的整体思路概述如下：首先对不同地区、不同负荷类型的电力用户进行广泛调研，得到调研样本的用户信息和所处地区信息；其次对调研样本的负荷分类进行校验，并对样本进行精选，以精选样本构成预测的全样本空间；然后根据区域类型和负荷类型对全样本空间进行二级划分，得到分层级子样本空间；最后，运用 SVM 理论构建分类预测模型，经子样本空间匹配和参数寻优后，用于预测待测地区各地块的负荷密度。该方法的实现流程如图 1 所示。

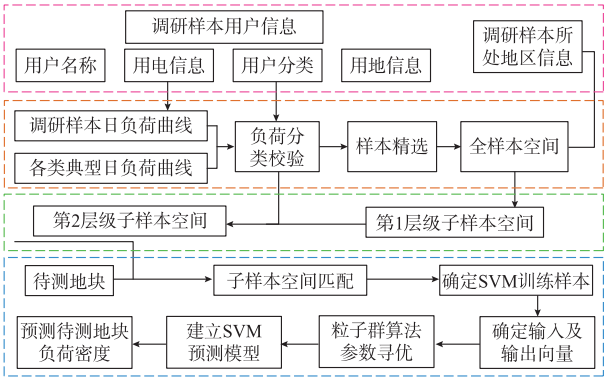

图 1　研究方法流程图

Fig.1　Flow chart of research method

2　负荷分类校验及样本精选

2.1　各类典型日负荷曲线提取

日负荷曲线是反映一日内负荷随时间变化规律的曲线，可直接反映用户的用电行为。用电行为相似的用户，其日负荷曲线形态也高度相似，并呈现明显的行业聚集性，不同类型的用户其负荷曲线形态则差异较大。本文利用改进 k-means 算法对各类用户的日负荷曲线进行聚类分析，以提取其典型日负荷曲线，主要包括以下步骤。

1）从有关部门得到分属工业、居民住宅、商业（即初始分类）等 L 个类别的电力用户的日负荷曲线，设每条日负荷曲线有 q 个量测数据，记第 i 条日负荷曲线为 $\boldsymbol{y}_i$，$\boldsymbol{y}_i=[y_{i1},y_{i2},\cdots,y_{iq}]$。

2）利用极大值标准化方法对每条日负荷曲线进行标准化处理，去除基荷数据的影响。

3）设定聚类数 k（显然 $k=L$），设定初始聚类中心为标准化处理后各类负荷曲线的中心线，以改善原始 k-means 算法因初始聚类中心随机设定易陷入局部最优[22]的不足。

4）以标准化处理后负荷曲线的每个采集点数据作为输入，以负荷曲线间的余弦相似度[23]作为相似性量度判据，将用户分为曲线形态相似的 k 个类别，重新标记该用户分类，记作聚类分类。

5）比较、分析各用户的初始分类与聚类分类结果，剔除分类不正确或用电行为不典型的用户后，求取各类负荷的典型日负荷曲线（标准化后同类日负荷曲线的中心线），记作 $\boldsymbol{y}_l(l=1,2,\cdots,L)$。

2.2　基于典型日负荷曲线的负荷分类校验及精选

负荷分类校验的实质是量度调研样本日负荷曲线与该类典型日负荷曲线间负荷形态的相似性。文献[23]指出，当负荷形态相近但存在细微波动趋势差异时，基于余弦相似度的相似性判据比欧氏距离更能反映负荷曲线之间的相似波动特性。本文以余弦相似度为判据，对调研样本的负荷分类进行校验，并对样本进行精选。具体步骤如下。

1）对所有调研样本的日负荷曲线依次进行极大值标准化处理，记标准化处理后的调研样本的日负荷曲线为 $\boldsymbol{c}_t(t=1,2,\cdots,T)$。

2）依次计算标准化处理后每个调研样本的日负荷曲线 $\boldsymbol{c}_t$ 与各类典型日负荷曲线 $\boldsymbol{y}_l$ 的余弦相似度。

$$\mathrm{Sim}_{\cos}(\boldsymbol{c}_t,\boldsymbol{y}_l)=\cos(\boldsymbol{c}_t,\boldsymbol{y}_l)=\frac{\boldsymbol{c}_t\boldsymbol{y}_l}{|\boldsymbol{c}_t||\boldsymbol{y}_l|}\quad l=1,2,\cdots,L \tag{1}$$

3）找出与 $\boldsymbol{c}_t$ 最相似（即与 $\boldsymbol{c}_t$ 余弦相似度最大）的典型日负荷曲线 y^*，对 $\boldsymbol{c}_t$ 标记 y^* 所属分类，记作校验分类。

4）比较 $\boldsymbol{c}_t$ 的初始分类和校验分类，筛选并复核两次分类不同的样本，修正所有分类错误样本的类标签。

5）对所有校验后的负荷样本按类进行精选，得到负荷分类正确且具备行业典型性的样本，构成预测的全样本空间。具体可参照 2.1 节，以每类样本的日负荷曲线为聚类对象，设定聚类数 $k=2$，利用改进 k-means 算法对每类样本进行再次聚类，把元素较少的一类剔除，把元素较多的一类作为该类负荷的精选样本。

3　分层级子样本空间形成

3.1　基于区域类型划分的第 1 层级子样本空间

负荷水平与当地经济社会发展水平、产业结构、气温气候密切相关。从以上三个角度出发，选取城镇化率、人口密度、人均用电量、人均国内生产总值(GDP)、一产 GDP、二产 GDP、三产 GDP、一产用电量、二产用电量、三产用电量、居民生活用电量、全年平均气温、冬季平均气温、夏季平均气温等 14 项指标，对各地负荷水平进行综合评估。各指标物理定义和选取依据见附录 A 表 A1。

设 $\boldsymbol{X}$ 为由 m 个地区上述 14 项指标构成的 $m\times 14$ 阶评估矩阵，对其无量纲化处理后，利用凝聚的层次聚类(agglomerative nesting，AGNES)[24]算法对各地区进行分类。算法的核心是量度两个簇之间的距离，通常采用欧氏距离。设 $\boldsymbol{R}_a$ 是簇 $\boldsymbol{C}_a$ 中的对象，$\boldsymbol{R}_b$ 是簇 $\boldsymbol{C}_b$ 中的对象，则簇 $\boldsymbol{C}_a$ 与簇 $\boldsymbol{C}_b$ 的距离为：

$$\mathrm{dist}(\boldsymbol{C}_a,\boldsymbol{C}_b)=\min_{\substack{R_a\in C_a\\ R_b\in C_b}}\sqrt{(\boldsymbol{R}_{a1}-\boldsymbol{R}_{b1})^2+(\boldsymbol{R}_{a2}-\boldsymbol{R}_{b2})^2+\cdots+(\boldsymbol{R}_{a14}-\boldsymbol{R}_{b14})^2} \tag{2}$$

若簇 $\boldsymbol{C}_a$ 与簇 $\boldsymbol{C}_b$ 的距离是所有不同簇中距离最小的，则簇 $\boldsymbol{C}_a$ 与簇 $\boldsymbol{C}_b$ 将被合并。

3.2　基于负荷类型划分的第 2 层级子样本空间

以校验后的负荷分类为依据，对第 1 层级的子样本空间进行二级划分，形成第 2 层级子样本空间，则经两级划分的子样本空间可标记为区域Ⅰ型工业负荷、区域Ⅱ型商业负荷等，分层级子样本空间示意图如图 2 所示。

图 2　分层级子样本空间示意图
Fig.2　Diagram of hierarchical sub-sample space

4　基于 SVM 的空间负荷预测建模

SVM[25]是一种基于统计学习理论的学习方法，以结构风险最小化为准则，在处理有限样本、非线性及复杂模式识别问题等方面具有突出优势。利用本文方法构建 SVM 模型进行空间负荷预测的步骤如下。

4.1　子样本空间匹配及训练样本确定

参照城市用地规划将待测地块划分成功能小区，根据待测地块所处地区信息(即附录 A 表 A1 中 14 项指标)匹配第 1 层级子样本空间，根据待测地块所属负荷类型匹配第 2 层级子样本空间，找到与待测地块最相似的子样本空间，并用该子样本空间的负荷样本作为 SVM 模型的训练样本。

4.2　输入及输出向量确定

把负荷密度的影响因素作为输入向量，把负荷密度作为输出向量。文献[19]给出了各类负荷的负荷密度影响因素，比如，把影响商业负荷密度的主要因素定为：地理位置 $\boldsymbol{C}_1$、商场定位 $\boldsymbol{C}_2$、商场面积 $\boldsymbol{C}_3$、当地生活消费水平 $\boldsymbol{C}_4$ 等，把影响居民负荷密度的主要因素定为人均收入 $\boldsymbol{R}_1$、人均用电量 $\boldsymbol{R}_2$、容积率 $\boldsymbol{R}_3$、电煤价格比率 $\boldsymbol{R}_4$ 等。

在此基础上，考虑到本文样本的地域差异，对每个样本增设当地人均用电量 $\boldsymbol{D}_1$、人均 GDP $\boldsymbol{D}_2$、夏季平均气温 $\boldsymbol{D}_3$、冬季平均气温 $\boldsymbol{D}_4$(精确到区县)等 4 个表征地域差异的负荷密度影响因素。

4.3　核函数选取及参数寻优

以径向基核函数作为预测模型中的核函数，并利用粒子群算法[26]对建模过程需要的惩罚参数和核参数进行寻优。将寻优后的参数输入 SVM 模型，可得到待测地块的预测负荷密度值。

5　实例与分析

以浙江电网为例，以附录 A 图 A1 湖州某地区为预测对象，以分布杭州、宁波等其他 10 个城市，分属工业、商业、居民住宅、行政办公等 4 种负荷类型的共 2 386 个样本为数据基础，对本文所提方法进行实例分析。

从上述样本中每类优选 25 个样本总计 100 个样本，用来提取各类负荷的典型日负荷曲线。以 100 个样本 2015 年 1 月某工作日的 24 h 日负荷曲线为对象，利用改进 k-means 算法对其进行聚类分析，聚类结果见附录 A 表 A2。

由附录 A 表 A2 可知，每类用户均存在一定比例的非本类或非典型用户，剔除分类不合理和非典型的用户后，提取得到各类用户的典型日负荷曲线，如图 3 所示。

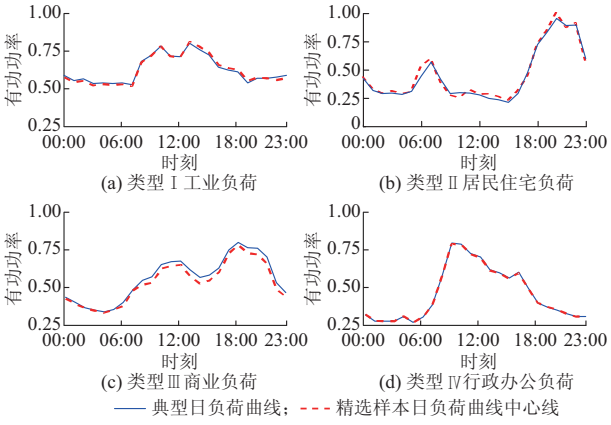

图 3　典型日负荷曲线与精选样本日负荷曲线对比
Fig.3　Comparison between typical load curves and selected sample load curves

由图 3 可知，各类负荷的典型日负荷曲线（蓝色实线）的负荷形态差异较大：工业负荷全天负荷率都较高，这是因为受峰谷分时电价影响，一般为三班制或两班制作业；居民住宅负荷在晨间和晚上呈现双峰段，与人们的工作时间和生活习惯相符；商业负荷在午间、晚间出现双峰段，凌晨 04:00 左右出现负荷低谷，与人们的消费和生活习惯相符；行政办公负荷在 09:00 左右出现全天负荷峰值，17:00 负荷显著减小，与该行业的上班时间相符。因此，从工程角度来看，提取的典型日负荷曲线合理，较好地反映了各类负荷的用电特点和负荷特性。

利用图 3 中 4 类典型日负荷曲线对全样本空间中的 2 386 个样本的负荷分类进行校验，共修正 413 个分类错误样本的类标签，剔除 481 个非典型样本，得到 1 905 个精选样本，精选比例为 79.8%。由图 3 可以看出，精选样本的日负荷曲线中心线（红色虚线）与该类典型日负荷曲线的负荷形态高度相似，验证了精选样本不仅分类正确，且具备典型行业特征。

利用 AGNES 算法对调研样本分布地区（精确到地市）进行聚类，聚类结果见附录 A 图 A2。附录 A 图 A2 兼顾各地区样本数量，将所有地区分为 3 类，分类结果见附录 A 表 A3。由此，根据区域分类、校验后负荷分类对全样本空间进行两级划分，得到 12 个子样本空间，用来存放 SVM 的训练样本。

以预测附录 A 图 A1 划线地块的负荷密度为例，按式(2)计算湖州与其他区域的欧氏距离，结果见附录 A 表 A4。由表 A4 可以看出，湖州与区域Ⅰ型（金华、台州、嘉兴、绍兴、温州）最为相似，且由附录 A 图 A1 可知划线地块为居住用地，故选取区域Ⅰ型居民负荷的子样本空间中的负荷样本作为该地块的训练样本。列举部分训练样本如附录 A 表 A5 所示。

已知该地块 2015 年的实际负荷密度为 8.56 W/m^2，各影响因素值为{6 439，58 477，26.3，6.7，516，59 939，1.77，1.8}，将上述影响因素值输入 SVM 模型，经粒子群算法参数寻优后得该地块负荷密度的预测值为 8.32 W/m^2，相对误差为 2.80%，满足精度要求。再收集 100 个湖州居民负荷为训练样本，利用文献[18-19]所提方法对该地块的负荷密度进行预测，预测结果见表 1。

表 1　不同预测方法下负荷密度预测值及相对误差
Table 1　Load density forecasting value and its relative error with different forecasting methods

预测方法	训练样本所处地区	预测值/(W·m^{-2})	相对误差/%
本文方法	金华、台州、嘉兴、绍兴、温州（即区域Ⅰ型）	8.32	2.80
文献[18]	湖州	7.97	6.89
文献[19]	湖州	8.13	5.02

由表 1 可知，利用本文、文献[18-19]方法预测的该地块预测值相对误差分别为 2.80%，6.89%，5.02%，均在合理范围内，本文所提方法的预测精度较文献[18-19]有明显提升。进一步，再按表 2 所列情况 2 至 6 构建 SVM 模型对其进行预测，通过控制变量的方式检验本文所提方法各环节的合理性。

由表 2 可知，对比情况 1 和情况 2，控制变量为训练样本所处地区，若未将调研样本按地区进行分类和匹配，即用所有区域的居民负荷样本作训练样本时，预测值远高于实际值，误差为 18.22%，对所有居民负荷样本区域构成进行分析可知，约 43.2%居民负荷样本属于区域Ⅲ型，因而可能受杭州、宁波居民负荷样本的影响，导致预测值偏高。

对比情况 1、情况 3 和情况 4 可知，控制变量为分类校验和样本精选，若未对负荷分类进行校验，即用初始分类为居民负荷的样本作训练样本时，误差达到 14.25%，用校验后分类为居民负荷的样本作训练样本时误差缩小到 4.32%，进一步对样本精选后预测误差可达到 2.80%，由此可知对负荷分类进行校验和对样本进行精选，能显著提高预测精度。

表 2　不同情况下负荷密度预测值及相对误差

Table 2　Load density forecasting value and its relative error under different circumstances

情况	训练样本	是否分类校验	是否样本精选	是否参数寻优	地域差异影响因素	预测值/(W·m^{-2})	相对误差/%
1	区域Ⅰ型居民负荷	是	是	是	有	8.32	2.80
2	所有区域居民负荷	是	是	是	有	10.12	18.22
3	区域Ⅰ型居民负荷	否	否	是	有	7.34	14.25
4	区域Ⅰ型居民负荷	是	否	是	有	8.19	4.32
5	区域Ⅰ型居民负荷	是	是	是	无	7.89	7.83
6	区域Ⅰ型居民负荷	是	是	否	有	8.16	4.67

对比情况 1 和情况 5，控制变量为是否增加地域差异影响因素，在没有增加地域差异影响因素作为 SVM 模型输入向量的情况下误差为 7.83%，较情况 1 误差未明显扩大，可见预测地块与训练样本地块的地域差异已经较小，说明区域分类合理，子样本空间匹配正确，增加地域差异影响因素作为输入向量可进一步缩小误差。

对比情况 1 和情况 6，由控制变量为“是否参数寻优”可知，引入粒子群算法对参数寻优同样能提高预测精度。

6　结语

本文在现有基于智能算法的负荷密度指标法的基础上，结合实际应用需求，将地域差异纳入考虑，提出了一种考虑地域的配电网空间负荷聚类及一体化预测方法。在工程实例中得以成功应用得益于以下 4 个方面。

1)在分类样本调研的基础上，增加了分地区的样本调研，对调研样本所处地区的负荷水平合理评估后，利用层次聚类算法将所有样本划分为具有地域差异的若干类，从而为该方法在其他地区的应用提供了样本基础和科学判据。

2)通过提取各类负荷的典型日负荷曲线，对调研样本的负荷分类进行校验，并对样本进行精选，极大提升了调研样本即训练样本的质量，显著提高了预测的精度。

3)应用 SVM 模型建模预测时，在分类负荷密度影响因素的基础上，增加了表征地域差异的负荷密度影响因素共同作为训练样本的输入向量，并运用粒子群算法对参数寻优，使预测精度进一步提高。

4)本文中有多处用到聚类分析方法，均根据实际需求对算法改进后应用，使得聚类结果更为合理。

随着信息社会的发展和大数据时代的到来，各类信息将更容易获取。随着全样本空间的扩大，本文所提方法的预测精度将进一步提高。因而，利用数据挖掘技术挖掘有效信息，构建满足更精细化预测需求的多层级子样本空间是今后该方法的发展方向。

附录见本刊网络版(http://www.aeps-info.com/aeps/ch/index.aspx)。

参考文献

[1] LIU Junyong, GAO Hongjun, MA Zhao, et al. Review and prospect of active distribution system planning[J]. Journal of Modern Power Systems and Clean Energy, 2015, 3(4): 457-467.

[2] GE Shaoyun, WANG Shiju, LU Zhiying, et al. Substation planning method in an active distribution network under low-carbon economy[J]. Journal of Modern Power Systems and Clean Energy, 2015, 3(4): 468-474.

[3] XIAO Jun, HE Qibo, ZU Guoqiang. Distribution management system framework based on security region for future low carbon distribution systems[J]. Journal of Modern Power Systems and Clean Energy, 2015, 3(4): 544-555.

[4] 蒲天骄，陈乃仕，王晓辉，等.主动配电网多源协同优化调度架构分析及应用设计[J].电力系统自动化，2016，40(1)：17-23.DOI：10.7500/AEPS20150520010.
PU Tianjiao, CHEN Naishi, WANG Xiaohui, et al. Application and architecture of multi-source coordinated optimal dispatch for active distribution network[J]. Automation of Electric Power Systems, 2016, 40(1): 17-23. DOI: 10.7500/AEPS20150520010.

[5] WILLIS H L, NORTHCOTE-GREEN J E D. Spatial electric load forecasting: a tutorial review[J]. Proceedings of the IEEE, 1983, 71(2): 232-253.

[6] CHOW M Y, ZHU Jinxiang, TRAM H. Application of fuzzy multi-objective decision making in spatial load forecasting[J]. IEEE Trans on Power Systems, 1998, 13(13): 1185-1190.

[7] 肖白，杨修宇，穆钢，等.基于元胞历史负荷数据的负荷密度指标法[J].电网技术，2014，38(4)：1014-1019.
XIAO Bai, YANG Xiuyu, MU Gang, et al. A load density index method based on historical data of cell load[J]. Power System Technology, 2014, 38(4): 1014-1019.

[8] 肖白，周潮，穆钢.空间电力负荷预测方法综述与展望[J].中国电机工程学报，2013，33(25)：78-92.
XIAO Bai, ZHOU Chao, MU Gang. Review and prospect of the spatial load forecasting methods[J]. Proceedings of the CSEE, 2013, 33(25): 78-92.

[9] 刘自发，庞铖铖，王泽黎，等.基于云理论和元胞自动机理论的城市配电网空间负荷预测[J].中国电机工程学报，2013，33(10)：98-105.
LIU Zifa, PANG Chengcheng, WANG Zeli, et al. Spatial load forecasting for distribution network based on cloud theory and

cellular automata[J]. Proceedings of the CSEE, 2013, 33(10): 98-105.

[10] 肖白，聂鹏，穆钢，等.基于多级聚类分析和支持向量机的空间负荷预测方法[J].电力系统自动化，2015，39(12)：56-61.DOI：10.7500/AEPS20140520001.
XIAO Bai, NIE Peng, MU Gang, et al. A spatial load forecasting method based on multilevel clustering analysis and support vector machine[J]. Automation of Electric Power Systems, 2015, 39(12): 56-61. DOI: 10.7500/AEPS20140520001.

[11] 赵强，景罗，赵光俊，等.顾及空间异质性的多尺度空间负荷预测[J].电力自动化设备，2014，34(2)：91-96.
ZHAO Qiang, JING Luo, ZHAO Guangjun, et al. Multi-scale spatial load forecasting considering spatial heterogeneity[J]. Electric Power Automation Equipment, 2014, 34(2): 91-96.

[12] MELO J D, CARRENO E M, PADILHA-FELTRIN A. Multi-agent simulation of urban social dynamics for spatial load forecasting[J]. IEEE Trans on Power Systems, 2012, 27(4): 1870-1878.

[13] GU C, YANG D, JIRUTITIJAROEN P, et al. Spatial load forecasting with communication failure using time-forward Kriging[J]. IEEE Trans on Power Systems, 2014, 29(6): 2875-2882.

[14] 陶文斌，张粒子，潘弘，等.基于双层贝叶斯分类的空间负荷预测[J].中国电机工程学报，2007，27(7)：13-17.
TAO Wenbin, ZHANG Lizi, PAN Hong, et al. Spatial electric load forecasting based on double-level Bayesian classification[J]. Proceedings of the CSEE, 2007, 27(7): 13-17.

[15] 符杨，曹家麟，谢楠，等.基于模糊综合评判的负荷密度指标选取新方法[J].电网技术，2007，31(18)：19-21.
FU Yang, CAO Jialin, XIE Nan, et al. A novel fuzzy comprehensive evaluation based method to select load density and index[J]. Power System Technology, 2007, 31(18): 19-21.

[16] 符杨，朱兰，曹家麟.基于模糊贴近度理论的负荷密度指标求取新方法[J].电力系统自动化，2007，31(19)：46-49.
FU Yang, ZHU Lan, CAO Jialin. A new method to obtain load density according to the theory of fuzzy approach degree [J]. Automation of Electric Power Systems, 2007, 31(19): 46-49.

[17] 刘自发，庞铖铖，魏建炜，等.基于 IAHP 和 TOPSIS 方法的负荷密度指标计算[J].电力系统自动化，2012，36(13)：56-59.
LIU Zifa, PANG Chengcheng, WEI Jianwei, et al. Index calculation of load density based on IAHP and TOPSIS methods[J]. Automation of Electric Power Systems, 2012, 36(13): 56-59.

[18] 周淥，孙威，张昀，等.基于改进型 ANFIS 的负荷密度指标求取新方法[J].电力系统保护与控制，2011，39(1)：29-34.
ZHOU Quan, SUN Wei, ZHANG Yun, et al. A new method to obtain load density based on improved ANFIS[J]. Power System Protection and Control, 2011, 39(1): 29-34.

[19] 周淥，孙威，任海军，等.基于最小二乘支持向量机和负荷密度指标法的配电网空间负荷预测[J].电网技术，2011，35(1)：66-71.
ZHOU Quan, SUN Wei, REN Haijun, et al. Spatial load forecasting of distribution network based on least squares support vector machine and load density index system[J]. Power System Technology, 2011, 35(1): 66-71.

[20] 王成山，黄纯华，葛少云，等.一个实用的城市电力负荷密度预测系统[J].电网技术，1992，16(6)：41-46.
WANG Chengshan, HUANG Chunhua, GE Shaoyun, et al. A practical forecasting system for urban electric load density[J]. Power System Technology, 1992, 16(6): 41-46.

[21] 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.
KANG Chongqing, XIA Qing, ZHANG Boming. Review of power system load forecasting and its development [J]. Automation of Electric Power Systems, 2004, 28(17): 1-11.

[22] 邓海，覃华，孙欣.一种优化初始中心的 K-means 聚类算法[J].计算机技术与发展，2013，23(11)：42-45.
DENG Hai, QIN Hua, SUN Xin. A K-means clustering algorithm of meliorated initial center[J]. Computer Technology and Development, 2013, 23(11): 42-45.

[23] 黄宇腾，侯芳，周勤，等.一种面向需求侧管理的用户负荷形态组合分析方法[J].电力系统保护与控制，2013，41(13)：20-25.
HUANG Yuteng, HOU Fang, ZHOU Qin, et al. A new combinational electrical load analysis method for demand side management[J]. Power System Protection and Control, 2013, 41(13): 20-25.

[24] 韩家炜.数据挖掘：概念与技术[M].范明，译.北京：机械工业出版社，2012：299-301.

[25] 张学工.关于统计学习理论与支持向量机[J].自动化学报，2000，26(1)：36-46.
ZHANG Xuegong. Introduction to statistical learning theory and support vector machines[J]. Acta Automatica Sinica, 2000, 26(1): 36-46.

[26] 邵信光，杨慧中，陈刚.基于粒子群优化算法的支持向量机参数选择及其应用[J].控制理论与应用，2006，22(5)：740-743.
SHAO Xinguang, YANG Huizhong, CHEN Gang. Parameters selection and application of support vector machines based on particle swarm optimization algorithm[J]. Control Theory and Applications, 2006, 22(5): 740-743.

刘　思(1991—)，女，硕士研究生，主要研究方向：电力系统规划、电力系统需求侧管理。E-mail：candyls@zju.edu.cn

傅旭华(1972—)，男，高级工程师，主要研究方向：电力系统配电网规划。

叶承晋(1987—)，男，工程师，主要研究方向：电力系统优化、电力系统配电网规划。

黄民翔(1955—)，男，通信作者，教授，主要研究方向：电力系统规划、电力市场。E-mail：huangmx@zju.edu.cn

（编辑　孔丽蓓）

（下转第 82 页　continued on page 82）

（上接第 75 页　continued from page 75）

Spatial Load Clustering and Integrated Forecasting Method of Distribution Network Considering Regional Difference

LIU Si[1], *FU Xuhua*[2], *YE Chengjin*[2], *HUANG Minxiang*[1]

(1. College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China;

2. Economic Research Institute of State Grid Zhejiang Electric Power Company, Hangzhou 310008, China)

Abstract: In view of the load density method based on intelligent algorithm heavily dependent on the sample and hardly applicable, a spatial load clustering and integrated forecasting method of distribution network considering regional difference is proposed. Firstly, power users and regional information on different cities and different types are obtained through massive investigation. Secondly, according to an analysis of the inadequacy of the current load classification mode, a method based on typical daily load curves is put forward to check the sample load classification label and select the proper samples. Thirdly, a full sample space is constructed by the selected samples, which is further classified into a hierarchical sub-sample space according to a two-level partition of regional property and checked classification label. Finally, a support vector machine (SVM) is built with the most similar sub-sample space to the plot to be predicted, thus the spatial load distribution is obtained. An actual case shows the practicality and effectiveness of the proposed method, which meets the accuracy requirements.

This work is supported by State Grid Corporation of China (No. 5211JY150016).

Key words: spatial load forecasting; load density method; regional difference; clustering analysis; support vector machine (SVM)