

DOI: 10.7667/PSPC180050

## 融合历史数据和实时影响因素的精细化负荷预测

席雅雯<sup>1</sup>, 吴俊勇<sup>1</sup>, 石琛<sup>1</sup>, 朱孝文<sup>1</sup>, 蔡蓉<sup>2</sup>

(1. 北京交通大学电气工程学院, 北京 100044; 2. ABB 中国研究院, 北京 100015)

**摘要:** 随着智能电网技术的飞速发展, 对负荷预测的精度提出了越来越高的要求。融合负荷、天气等多源数据, 提出了一种基于数据融合的支持向量机精细化负荷预测方法。首先对负荷历史数据进行聚类分析, 将运行日分成六类。然后将负荷数据和温度、湿度等天气数据进行融合, 针对六类聚类结果分别建立基于数据融合的支持向量机精细化负荷预测模型, 并对模型参数进行全局优化。采用不同的预测模型对浙江省某地级市 2013 年的负荷进行预测, 结果表明所提出的负荷预测方法的预测精度明显高于传统的负荷预测方法的预测精度。

**关键词:** 负荷预测; 数据融合; 支持向量机; 预测精度

### A refined load forecasting based on historical data and real-time influencing factors

XI Yawen<sup>1</sup>, WU Junyong<sup>1</sup>, SHI Chen<sup>1</sup>, ZHU Xiaowen<sup>1</sup>, CAI Rong<sup>2</sup>

(1. School of Electrical Engineering, Beijing Jiaotong University, Beijing 100044, China;  
2. ABB China Research Institute, Beijing 100015, China)

**Abstract:** With the rapid development of smart grid technology, increasingly demand on the accuracy of load forecasting is put forward. Integrating load, weather and other multi-sourced data, a refined load forecasting method of Support Vector Machine (SVM) based on data fusion is proposed. Firstly, the historical load data is clustered and the operation days are divided into six categories. Then the weather data such as temperature and humidity are combined with the load data, and the refined load forecasting models of SVM based on data fusion are established respectively for the six clustering results. And the parameters of the model are optimized globally. Different forecasting models are used to predict the load of a prefecture-level city in Zhejiang Province in 2013, the prediction results show that the prediction accuracy of the load forecasting method proposed in this paper is obviously higher than that of the traditional load forecasting method.

This work is supported by National Natural Science Foundation of China (No. 51577009) and ABB China Research Institute (No. ABB20171128REU-CTR).

**Key words:** load forecasting; data fusion; support vector machines (SVM); prediction accuracy

## 0 引言

负荷预测<sup>[1]</sup>是电力系统安全稳定运行的重要依据, 也是电力系统规划、计划和调度的一项基本工作。随着智能电网技术的飞速发展, 电力生产和消费更加市场化, 负荷预测的实时结果数据将成为电力交易中最重要数据之一<sup>[2]</sup>, 这就对负荷预测的准确性和实时性提出了更高的要求。准确的负荷预测对于保证各经济实体的利益具有基础性的作用,

电力系统各环节的运行规划都有赖于准确的负荷预测。其中发电企业需要准确的负荷预测以避免其报价过低或者不能满发; 输配电企业需要准确的负荷预测以保证各条线路输送的容量经济且安全; 配售电企业需要负荷预测以准确地把握用户负荷的需求, 及时反应市场的变化波动以提高自身的竞争力。

无论在国外还是在国内, 许多学者都对负荷预测的方法进行了深入研究, 总结出多种多样有关负荷预测的方法。传统的负荷预测方法有外推法、指数平滑法、相关分析法、回归分析法、时间序列法、卡尔曼滤波法和灰色系统法等。现在常用的负荷预

**基金项目:** 国家自然科学基金项目资助(51577009); ABB 中国研究院项目资助(ABB20171128REU-CTR)

测方法有人工神经网络<sup>[3]</sup>、支持向量机<sup>[4-5]</sup>、小波变换<sup>[6-7]</sup>及模糊回归等,各种方法均有着独特的优势和明显的缺点。由于影响负荷的因素复杂且不同因素对负荷的影响呈现非线性的特点,所以负荷预测仍然存在相当大的困难,实践证明目前没有一种理论和方法能够完全适用于所有情况下的负荷预测,应用数据挖掘技术进行历史数据的筛选并与现有的负荷预测方法相结合是负荷预测发展的新趋势。

文献[8]利用多核函数对影响电力系统负荷波动的诸多因素进行融合,建立支持向量机预测模型,预测结果证明此种方法能够有效提高负荷预测的速度和精度。文献[9]采用负荷相似日及气象相似日作为训练样本,减少输入变量的维数,消除大量冗余数据。文献[10]提出了一种联合灰色模型和支持向量机回归算法的电力系统短期智能组合预测方法。

本文提出了一种融合历史数据和实时影响因素的支持向量机精细化负荷预测方法。首先,利用数据挖掘技术深度挖掘负荷的日属性规律,针对不同属性的日期(即不同的聚类类别)分别建立相应的预测模型,将负荷数据和湿度、温度等天气数据进行数据融合,作为模型的输入样本;其次本文所采用的负荷数据及天气数据均来自浙江省某地级市,由于浙江省地处东南沿海地带,夏季雨热同期,因此要充分考虑当地的地理位置和气候条件,定量分析天气各影响因素与负荷之间的相关程度,突出主要的天气因素,提高预测精度;最后,对支持向量机的参数对  $c$  和  $g$  进行全局搜索寻优,找到最佳参数对。针对实际系统的算例表明,与传统负荷预测方法相比,本文所提方法可将负荷预测精度提高 5% 到 10% 左右。

## 1 数据的采集及预处理

本文所采用的负荷数据和天气数据均来源于浙江省某地级市电网公司,其中负荷数据来自供电区域内 2007 年至 2013 年 15 个 220 kV 的变电站,采样间隔为 15 min,即一天 96 个采样点。天气数据为供电区域内 2007 年至 2013 年的温度、湿度和降雨量等,采样间隔为 1 h,即一天 24 个采样点。

### 1.1 不良负荷数据的处理

本文采用的是基于数据融合的负荷预测方法,需要大量的样本数据作为负荷预测的依据,样本数据的质量对负荷预测的精度有很大影响。因此需要对采集到的负荷数据手动进行不良数据<sup>[11-12]</sup>的处理,不良数据包括空数据点、零数据点、连续恒定值及异常阶跃值等。但为了保证数据样本不失真,应少对数据进行修改,尽量使用原数据,以免造成

不必要的误差。

对于单个不良数据的处理方法:依据电力系统负荷曲线具有平滑性的特点,将异常点的上一个负荷数据与下一个负荷数据取平均值得到该点的负荷数据。对于连续多个不良数据的处理方法:依据电力系统负荷曲线走势大致相同的特点,尽量采用在附近相邻日期找到相似负荷曲线进行置换的方法,若无法找到,则直接剔除。

### 1.2 天气数据的插值处理

原始负荷数据为一天 96 个采样点,原始天气数据为一天 24 个采样点,而实际上支持向量机在建模时需要多点的负荷数据对应多点的天气数据,因此需要对现有的天气数据进行插值预处理,本文采用三次样条插值处理。

### 1.3 星期类型及节假日的量化

在建立负荷预测的模型时所使用的数据必须是量化的数据,而原始的星期类型和节假日类型都是非量化的数据,因此必须将这些原始数据映射为建模能够识别的数值数据。本文将星期类型量化为  $A=\{1,2,3,4,5,6,7\}$ ,1~7 分别对应星期一到星期日。本文的节假日主要指国庆、春节等国家法定节假日,将节假日属性量化为  $B=\{0,1,2\}$ ,其中 0 代表节假日,1 代表节假日后一天,2 代表除这两种类型以外的其他日期。

### 1.4 样本数据的归一化

常用的归一化函数主要有三种:线性函数转换、对数函数转换以及反余切函数转换。本文采用简单的线性归一化,转换公式为

$$y_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

式中:  $y_i$  为经归一化处理后的输入样本数据;  $x_i$  为经预处理后的输入样本数据;  $x_{\min}$  为经预处理后输入样本中的最小值;  $x_{\max}$  为经预处理后输入样本中的最大值。

## 2 负荷的规律特性分析

负荷的规律特性主要体现在时间维度和空间维度上,以下将从负荷的日规律特性、星期规律特性及节假日规律特性三方面对负荷时间维度的规律特性进行分析。从分析结果可以得知,对不同日属性的负荷分别建立预测模型,可以提高预测精度。

### 2.1 负荷的日负荷曲线

典型的日负荷曲线具有“早高峰”和“晚高峰”两个用电高峰。早高峰大约从 06:00—07:00 多负荷达到峰值,随后负荷开始下降,到 09:00 左右

降低低谷；晚高峰大约从 18:00—19:00 左右负荷达到峰值，从 21:00 左右负荷开始下降。早高峰和晚高峰出现的时间与人们的生活规律和工作时间完全相符，而且每日的早高峰和晚高峰出现的时间基本都是固定的，这为分析用户的用电规律提供了便利。典型的日负荷曲线如图 1 所示。

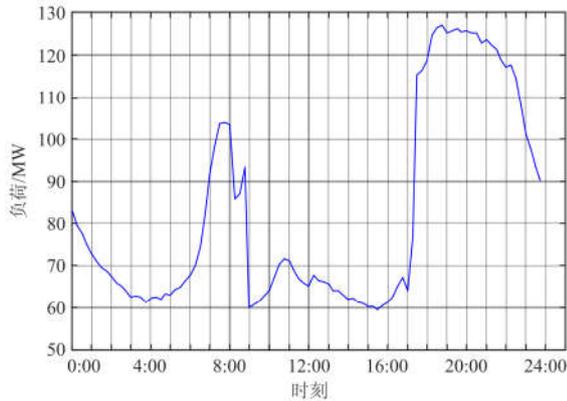


图 1 典型的日负荷曲线

Fig. 1 Typical daily load curve

### 2.2 负荷的星期规律特性

负荷的星期规律特性主要体现在以下两点：(1) 工作日和周末的负荷曲线走势大致相同，每日的负荷曲线都有早高峰和晚高峰两个用电高峰；(2) 周末的负荷明显低于工作日的负荷，原因是由于工业负荷在当地负荷中所占的比例比较大，周末工厂会停工，因此负荷会下降。以 2011 年浙江省某地区第二星期的负荷数据为例，分析负荷的星期规律特性，其中红线表示周一到周五的负荷曲线，蓝线表示周末的负荷曲线，如图 2 所示。

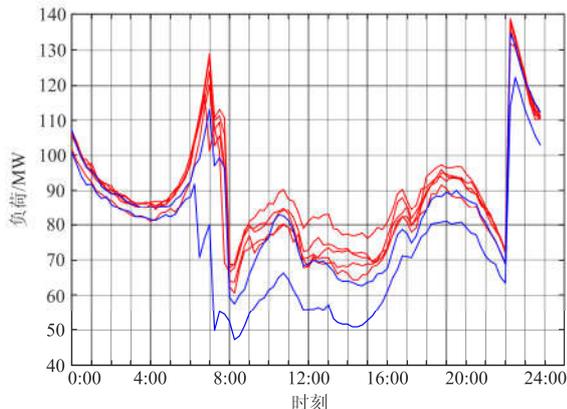


图 2 一星期的负荷曲线

Fig. 2 Load curve in one week

### 2.3 负荷的节假日规律特性

文中的节假日主要指中秋节、国庆节等国家法定节假日，不同节假日的负荷变化规律大致相同，

没必要对所有节假日一一进行分析。以 2010、2011 及 2012 年浙江省某地区国庆节的负荷数据为例，绘制连续三年国庆节的负荷曲线，如图 3 所示，其中黑线表示 2010 年国庆节的负荷曲线，红线表示 2011 年国庆节的负荷曲线，蓝线表示 2012 年国庆节的负荷曲线，从图中可以看出不同年份国庆节负荷曲线的走势大致相同，且每年节前和节后的负荷明显高于国庆节期间的负荷。

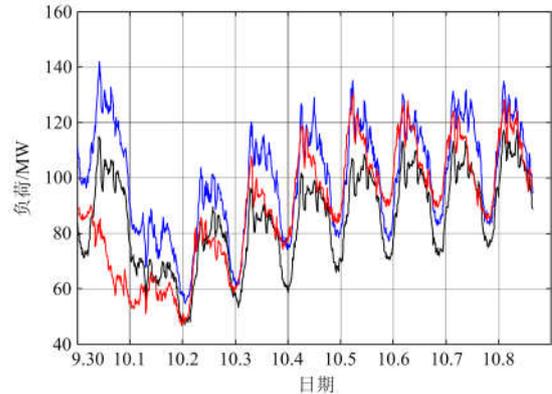


图 3 2010、2011、2012 国庆节负荷曲线

Fig. 3 Load curve during the National Day in 2010, 2011, and 2012

### 3 负荷与各影响因素的相关性分析

除了上述负荷的星期规律特性及节假日规律特性外，温度、湿度等天气因素<sup>[13]</sup>对负荷的影响也不容忽略，定量分析天气各影响因素与负荷之间的相关程度，根据相关程度的大小，保留主要影响因素，剔除次要影响因素，提高预测精度。采用 Pearson 相关分析法，从两个时间维度上进行相关性分析。一是从全年的角度，用 2012 年浙江省某地区全年的负荷及天气数据进行 Pearson 相关性分析，定量分析结果如表 1 所示。从表 1 可以看出，相对于其他影响因素，温度和湿度对负荷的影响比较大。二是从季节的角度，由于浙江省地处东南沿海地区，属于亚热带季风气候，雨热同期，用 2012 年浙江省某地区夏季的负荷及天气数据进行 Pearson 相关性分析，定量分析结果如表 2 所示。从表 2 可以明显地看出：尤其在夏季，温度和湿度与负荷之间的相关性很大，即温度和湿度对负荷的影响非常大。

实际上浙江省初夏天气闷热潮湿多暴雨，盛夏晴热少雨多高温天气，是全年最热的时段，这一时期除台风和局部地区雷阵雨等极端天气外，降水较少，常有伏旱出现。即夏季降雨与负荷之间的相关性也很强，但分析结果并没有很明显。因此，以 2012 年浙江省某地区夏季的负荷及天气数据为例，对天

气情况进行分类, 进一步分析夏季降雨与负荷之间的相关程度。分析结果如表 3 所示, 浙江省某地区夏季的温度基本都在  $20^{\circ}\text{C}$  以上, 当温度在  $23^{\circ}\text{C}$  左右, 湿度在 90% 左右时, 降雨较多; 而当温度在  $27^{\circ}\text{C}$  以上, 湿度在 85% 以下时, 降雨较少。

表 1 2012 年全年负荷与各影响因素的相关性分析

Table 1 Correlation analysis of annual load and its influencing factors in 2012

相关系数	负荷
温度	0.476
湿度	-0.382
降雨量	0.040
2 min 风速	0.147
2 min 风向	0.003

表 2 2012 年夏季负荷与各影响因素的相关性分析

Table 2 Correlation analysis of summer load and its influencing factors in 2012

相关系数	负荷
温度	0.758
湿度	-0.557
降雨量	-0.115
2 min 风速	0.074
2 min 风向	0.083

表 3 2012 年夏季负荷与降雨的相关性分析

Table 3 Correlation analysis of summer load and rainfall in 2012

相关系数	负荷
降雨(温度在 $23^{\circ}\text{C}$ 左右, 湿度在 90% 左右)	-0.627
降雨(温度在 $27^{\circ}\text{C}$ 以上, 湿度在 85% 以下)	-0.045

## 4 建立基于数据融合的支持向量机负荷预测模型

### 4.1 聚类方法及数目的选取

聚类分析是常用的一种多元统计分析方法, 就是把数据对象集划分成多个组或簇的过程, 使得簇内的对象具有很高的相似性, 但与其他簇中的对象很不相似, 相异性和相似性根据描述对象的属性值评估, 并且通常涉及距离度量。

基本的聚类方法有基于划分的聚类方法、基于层次的聚类方法、基于密度的聚类方法以及基于网格的聚类方法。最常采用的聚类方法有 k 均值聚类、分步聚类等, 但由于 k 均值聚类需要事先指定 k 的值, 且当数据量不大时, 初始化分组很大程度上决定了聚类的结果, 分步聚类方法可以通过生成距离矩阵输出任意指定聚类的数目。因此本文选择采用

基于欧式距离的分步聚类方法, 其步骤如下:

(1) 对经归一化处理的样本数据, 用 `pdist` 函数求出各输入变量基于欧氏距离的相似矩阵;

(2) 使用 `linkage` 函数产生基于加权平均距离的聚类树;

(3) 用 `cophenet` 函数评价聚类信息;

(4) 用 `cluster` 函数进行聚类, 返回聚类列。

由于负荷受温度、湿度等天气因素影响的同时, 也受到星期类型、节假日类型等日属性的影响。因此在进行负荷预测之前, 有必要对负荷的样本数据进行聚类, 然后再对不同类别的负荷分别建立预测模型, 这样不仅可以大大提高负荷预测精度, 还可以大大缩短训练时间。

目前对聚类数目的选取还没有公认比较好的方法, 只能通过不断尝试, 经本文尝试, 将 2010 年、2011 年及 2012 年连续三年的日 96 点负荷进行聚类分析, 当选取聚类的数目为 6 时, 基本上可以保证每种聚类结果中都含有一定数量的样本, 这六类分别为: 工作日、周末、国家法定节假日、除了国家法定节假日以外的其余节日(例如圣诞节等)、节假日后的第一天以及特殊日(例如 G20 峰会等)。为了将聚类结果直观地展示出来, 用聚类树表示聚类结果, 如图 4 所示, 其中横坐标表示样本数, 纵坐标表示距离。

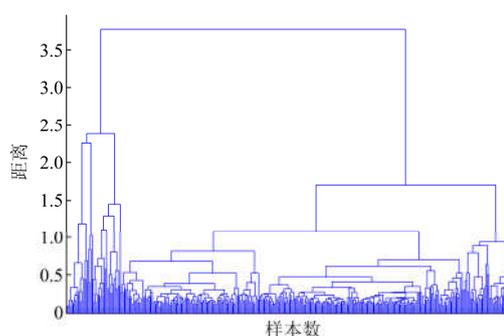


图 4 聚类树

Fig. 4 Clustering tree

### 4.2 支持向量机原理

SVM<sup>[14-16]</sup>是一种常用的机器学习算法, 其在解决小样本、非线性及高维模式识别中表现出许多特有的优势, 并能够推广应用到函数拟合等其他机器学习问题中, 还可用于数据分析、分类和回归分析等。

SVM 以统计学理论作为理论基础, 与传统神经网络学习方法不同的是: SVM 实现了结构风险最小化原理, 同时能够最小化经验风险与 VC 维的界, 具有很好的泛化性能和精度, 减少了对经验的依赖。

支持向量机<sup>[17-19]</sup>用于负荷预测的基本原理是:

对于训练样本集  $\{x_i, y_i\}_{i=1}^l$ ,  $x_i \in R^n$  为输入样本,  $y_i \in R$  为输入样本对应的输出值,  $l$  为训练样本的个数, SVM 的目标就是寻找从输入空间  $R^n$  到输出空间  $R$  的回归函数:

$$y = f(x) = (\omega \cdot x) + b \quad (2)$$

式中:  $\omega \in R^n$  为权重;  $x_i \in R^n$  为输入样本;  $b \in R$  为阈值。

对于线性回归问题, 根据统计学理论, SVM 通过极小化目标来确定回归函数, 即

$$\begin{cases} \text{Min} \left\{ \frac{1}{2} \|\omega\|^2 + C \cdot \left[ v\varepsilon + \frac{1}{l} \sum_{i=1}^l (\zeta_i + \zeta_i^*) \right] \right\} \\ \text{s.t.} \left\{ \begin{aligned} ((\omega \cdot x_i) + b) - y_i &\leq \varepsilon + \zeta_i \\ y_i - ((\omega \cdot x_i) + b) &\leq \varepsilon + \zeta_i^* \\ \zeta_i^* \geq 0, \varepsilon &\geq 0 \end{aligned} \right. \end{cases} \quad (3)$$

式中:  $\varepsilon$  为不敏感损失函数;  $\zeta_i$ 、 $\zeta_i^*$  为松弛因子。对于约束条件,  $\alpha_i^*$ ,  $\alpha_i$ ,  $\eta_i s$ ,  $\eta_i^*$ ,  $\beta \geq 0$ , 根据拉格朗日算法求得 SVM 的回归函数为

$$f(x) = \sum_{i=1}^l (\alpha_i^* - \alpha_i) K(x_i \cdot x) + b \quad (4)$$

对于非线性回归问题, 可以通过非线性变换将原始问题映射到某个高维空间中的线性问题进行求解。在高维特征空间中的核函数为

$$K(x_i, x_j) = \Phi(x_i) \Phi(x_j) \quad (5)$$

常用的核函数有以下三种形式。

(1) 多项式核函数

$$K(x, y) = (x \cdot y + 1)^d, d = 1, 2, \dots \quad (6)$$

(2) 径向基核函数(RBF)

$$K(x, y) = \exp\left[-\frac{\|x - y\|^2}{\sigma^2}\right] \quad (7)$$

(3) sigmoid 核函数

$$K(x, y) = \tanh[b \cdot (x \cdot y) + \theta] \quad (8)$$

本文选用预测效果相对较好的径向基核函数作为 SVM 的核函数。

#### 4.3 交叉验证选择最佳参数 $c$ 和 $g$

关于 SVM 参数的优化<sup>[20-21]</sup>选取, 国际上并没有公认统一的好方法, 目前常用的方法就是让  $c$  和  $g$  在一定的范围内取值, 对于取定的  $c$  和  $g$ , 把训练集作为原始数据集并利用 K-CV 方法得到在此组  $c$  和  $g$  下训练集验证均方误差, 最终取使得训练集均方误差最低的那组  $c$  和  $g$  作为最佳参数。如果有多组  $c$  和  $g$  对应于最低的均方误差, 在这种情况下, 应选取能够达到最低均方误差中的参数  $c$  最小的那组  $c$  和  $g$  作为最佳参数, 如果对应最小的  $c$  有多组  $g$ , 就选取搜索到的第一组  $c$  和  $g$  作为最佳参数。这

样做的理由是: 过高的  $c$  会导致过学习状态发生, 即训练集均方误差很低而测试集均方误差很高, 所以在能够达到最低均方误差中的所有成对的  $c$  和  $g$  中较小的惩罚系数  $c$  是最佳的选择对象。

K-CV 选择的最佳参数  $c$  和  $g$  如图 5 所示。

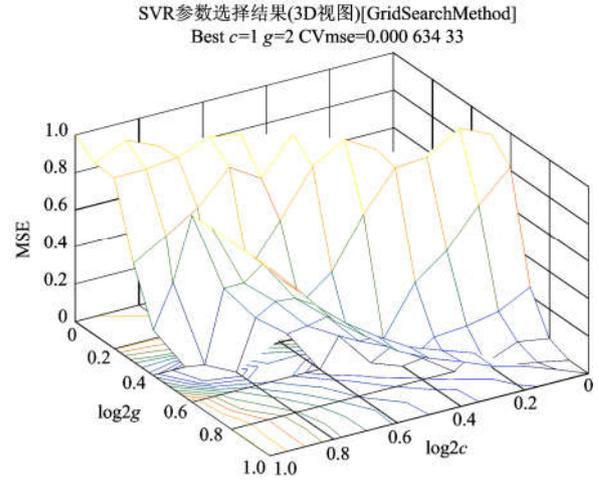


图 5 K-CV 选择最佳参数  $c$  和  $g$

Fig. 5 The best parameters  $c$  and  $g$  selected by K-CV

从图 5 可以看出, 在选择最佳参数  $c$  和  $g$  的过程中, 能够使  $c$  和  $g$  在整个网格内进行搜索, 使均方误差能够在 0 到 1 范围内浮动。经交叉验证选择得到的最佳参数对为:  $c=1$ ,  $g=2$ 。

## 5 算例验证

### 5.1 输入样本数据

本算例选取浙江省某地区 2010 年、2011 年及 2012 年连续三年的负荷数据及温度、湿度等天气数据作为输入样本, 其中负荷数据的采样频率为 15 min, 天气数据的采样频率为 1 h。

### 5.2 预测结果的评价指标

以采样点负荷预测偏差率  $E_i$  (式(9))和日平均负荷预测准确率  $A_d$  (式(10))作为衡量预测效果的评价指标。

$$E_i = \frac{L_{i,f} - L_i}{L_i} \times 100\% \quad (9)$$

式中:  $E_i$  为时刻  $i$  的负荷预测偏差率(%);  $L_{i,f}$  为时刻  $i$  的负荷预测值(MW);  $L_i$  为时刻  $i$  的负荷实际值(MW)。

$$A_d = (1 - \text{RMSE}) \times 100\% \quad (10)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n E_i^2} \quad (11)$$

式中:  $A_d$  为日平均负荷预测准确率(%); RMSE 为

日负荷预测偏差率均方根;  $E_i$  为每 15 min 的负荷预测偏差率(%);  $n$  为日实际考核点数。

### 5.3 结果分析

采用本文提出的融合历史数据和实时影响因素的支持向量机负荷预测方法, 对浙江省某地级市 2013 年的负荷进行预测, 表 4 列举了 3 种典型运行日的负荷预测结果, 即工作日(2013.01.09)、周末(2013.02.16)和节假日(2013.06.12)三大类负荷的日平均负荷预测准确率, 并与经典的时间序列法及传统的支持向量机负荷预测方法的预测结果进行对比。

表 4 各种预测方法的日平均负荷预测准确率

Table 4 Daily average load forecasting rate of various forecasting methods

负荷预测方法	工作日	周末	节假日
ARMA	88.91%	86.72%	87.85%
传统 SVM	93.62%	92.49%	92.47%
本文所提方法	97.46%	97.27%	97.12%

从表 4 可以看出: 相对于经典的时间序列法及传统的支持向量机负荷预测方法<sup>[22-25]</sup>, 本文提出的融合历史数据和实时影响因素的精细化负荷预测方法能够将预测精度提高 5%到 10%左右。

由于文章篇幅有限, 本文仅详细介绍对工作日(2013.01.09)的负荷预测结果。3 种预测方法的预测结果与实际负荷的对比如图 6 所示, 其中横坐标表示 96 个采样点, 纵坐标表示负荷值。

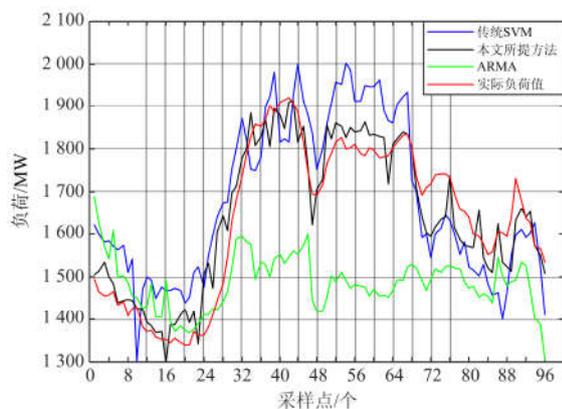


图 6 3 种预测方法的预测结果与实际负荷的对比

Fig. 6 Comparison the prediction results of three kinds of forecasting methods and the actual load

从图 6 中 4 条负荷曲线的对比可知: 无论是整体负荷曲线的平滑度, 还是单个负荷点的预测值, 采用本文所提出的融合历史数据和实时影响因素的支持向量机精细化负荷预测方法都要比传统的支持向量机负荷预测方法和时间序列负荷预测方法的预

测效果好很多。

为了更直观地对比 3 种预测方法的预测精度, 如图 7 所示, 图中曲线表示采用 3 种不同预测方法的 96 个采样点负荷预测偏差率的绝对值, 其中横坐标表示 96 个采样点, 纵坐标表示每个采样点负荷预测偏差率的绝对值。

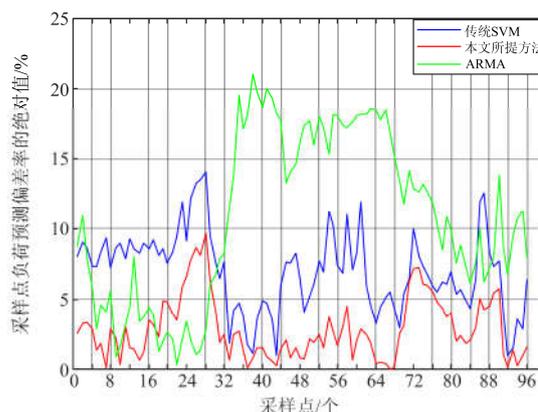


图 7 3 种预测方法预测偏差率绝对值的对比

Fig. 7 Comparison of three prediction methods to predict the absolute deviation rate

从图 7 可以明显看出: 针对 2013 年 1 月 9 日的日 96 点负荷的预测结果, 本文提出的融合历史数据和实时影响因素的支持向量机精细化负荷预测方法的 96 个采样点负荷预测偏差率的绝对值大多都在 5%以下, 集中在 3%附近; 传统的支持向量机负荷预测方法的 96 个采样点负荷预测偏差率的绝对值基本都在 5%以上, 集中在 8%附近, 采样点负荷预测偏差率的绝对值提高了近 5%。而时间序列负荷预测方法的 96 个采样点负荷预测偏差率的绝对值浮动范围较大, 均值在 13%左右。

通过算例分析, 验证了本文所提出的融合历史数据和实时影响因素的支持向量机精细化负荷预测方法的有效性, 使预测精度能够显著提高。

## 6 结论

本文将负荷数据和温度、湿度等天气数据进行融合, 提出了一种基于数据融合的支持向量机精细化负荷预测方法, 得到以下结论:

(1) 根据负荷和天气等多源信息, 采用关联关系分析和聚类算法, 将运行日分成若干大类, 分别建立预测模型, 可有效提高负荷预测的精度;

(2) 对支持向量机的核心参数进行全局优化, 也可提高预测精度;

(3) 针对实际系统的算例结果表明, 与传统的支持向量机和时间序列负荷预测方法相比, 本文所提

方法可将负荷预测精度提高 5%到 10%左右。

由于不同地区的负荷特性、地理位置和天气条件各不相同, 所得到的聚类结果和模型参数也可能不尽相同, 必须因地制宜, 才可以有效提高负荷预测的精度。

### 参考文献

- [1] 雷正新, 韩蓓, 聂萌, 等. 配电网大数据环境下的多点负荷预测算法与应用研究[J]. 电力系统保护与控制, 2016, 44(23): 1-11.  
LEI Zhengxin, HAN Bei, NIE Meng, et al. Multipoint load forecasting algorithm and its application in distribution network big data environment[J]. Power System Protection and Control, 2016, 44(23): 1-11.
- [2] 王新, 孟玲玲. 基于 EEMD-LSSVM 的超短期负荷预测[J]. 电力系统保护与控制, 2015, 43(1): 61-66.  
WANG Xin, MENG Lingling. Ultra-short term load forecasting based on EEMD-LSSVM[J]. Power System Protection and Control, 2015, 43(1): 61-66.
- [3] 苏学能, 刘天琪, 曹鸿谦, 等. 基于 Hadoop 架构的多重分布式 BP 神经网络的短期负荷预测方法[J]. 中国电机工程学报, 2017, 37(17): 4966-4973, 5216.  
SU Xueneng, LIU Tianqi, CAO Hongqian, et al. Short-term load forecasting method for multi-distributed BP neural network based on Hadoop architecture[J]. Proceedings of the CSEE, 2017, 37(17): 4966-4973, 5216.
- [4] 罗楠, 朱业玉, 杜彩月, 等. 支持向量机方法在电力负荷预测中的应用[J]. 电网技术, 2007, 31(增刊 2): 215-218.  
LUO Nan, ZHU Yeyu, DU Caiyue, et al. Application of support vector machine method in electric load forecasting[J]. Power System Technology, 2007, 31(S2): 215-218.
- [5] 耿艳, 韩学山, 韩力. 基于最小二乘支持向量机的短期负荷预测[J]. 电网技术, 2008, 32(18): 72-76.  
GENG Yan, HAN Xueshan, HAN Li. Short-term load forecasting based on least squares support vector machine[J]. Power System Technology, 2008, 32(18): 72-76.
- [6] 张平, 潘学萍, 薛文超. 基于小波分解模糊灰色聚类 and 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, 141.  
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, 141.
- [7] 祖向荣, 田敏, 白焰. 基于模糊聚类与函数小波核回归的短期负荷预测方法[J]. 电力自动化设备, 2016, 36(10): 134-140, 165.  
ZU Xiangrong, TIAN Min, BAI Yan. Short term load forecasting method based on fuzzy clustering and function wavelet kernel regression[J]. Electric Power Automation Equipment, 2016, 36(10): 134-140, 165.
- [8] 吴倩红, 高军, 侯广松, 等. 实现影响因素多源异构融合的短期负荷预测支持向量机算法[J]. 电力系统自动化, 2016, 40(15): 67-72, 92.  
WU Qianhong, GAO Jun, HOU Guangsong, et al. Short-term load forecasting support vector machine algorithm based on multi-source heterogeneous fusion of load factors[J]. Automation of Electric Power Systems, 2016, 40(15): 67-72, 92.
- [9] 李滨, 黄佳, 吴茵, 等. 基于分形特性修正气象相似日的节假日短期负荷预测方法[J]. 电网技术, 2017, 41(6): 1949-1955.  
LI Bin, HUANG Jia, WU Yin, et al. A short-term load forecasting method based on fractal feature correction of meteorological similar days[J]. Power System Technology, 2017, 41(6): 1949-1955.
- [10] 唐杰明, 刘俊勇, 杨可, 等. 基于灰色模型和最小二乘支持向量机的电力短期负荷组合预测[J]. 电网技术, 2009, 33(3): 63-68.  
TANG Jieming, LIU Junyong, YANG Ke, et al. Short-term load combination forecasting by grey model and least square support vector machine[J]. Power System Technology, 2009, 33(3): 63-68.
- [11] 王海云, 田莎莎, 张再驰, 等. 基于负荷预测与关联规则修正的不良数据辨识方法[J]. 电力系统保护与控制, 2017, 45(23): 24-33.  
WANG Haiyun, TIAN Shasha, ZHANG Zaichi, et al. A method of identifying bad data based on load forecasting and association rules[J]. Power System Protection and Control, 2017, 45 (23): 24-33.
- [12] 杨慧霞, 邓迎君, 刘志斌, 等. 含有历史不良数据的电力负荷预测研究[J]. 电力系统保护与控制, 2017, 45(15): 62-68.  
YANG Huixia, DENG Yingjun, LIU Zhibin, et al. Research on power load forecasting with historical bad data[J]. Power System Protection and Control, 2017, 45(15): 62-68.
- [13] 高亚静, 孙永健, 杨文海, 等. 基于新型人体舒适度的气象敏感负荷短期预测研究[J]. 中国电机工程学报, 2017, 37(7): 1946-1955.  
GAO Yajing, SUN Yongjian, YANG Wenhai, et al. Study on short-term forecast of meteorological sensitive load

- based on new human comfort[J]. Proceedings of the CSEE, 2017, 37(7): 1946-1955.
- [14] HUO J, SHI T, CHANG J. Comparison of random forest and SVM for electrical short-term load forecast with different data sources[C] // 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), August 26-28, 2016, Beijing, China: 1077-1080.
- [15] YE N, LIU Y, WANG Y. Short-term power load forecasting based on SVM[C] // World Automation Congress 2012, June 24-28, 2012, Puerto Vallarta, Mexico: 47-51.
- [16] ZHANG M G, LI L R. Short-term load combined forecasting method based on BPNN and LS-SVM[C] // 2011 IEEE Power Engineering and Automation Conference, September 8-9, 2011, Wuhan, China: 319-322.
- [17] SELAKOV A, ILIC S, VUKMIROVIC S, et al. A comparative analysis of SVM and ANN based hybrid model for short term load forecasting[C] // PES T&D 2012, May 7-10, 2012, Orlando, FL, USA: 1-5.
- [18] LIU B, XU G. Short-term power load forecasting based on LS-SVM[C] // 2010 International Conference of Information Science and Management Engineering, August 7-8, 2010, Xi'an, China: 311-314.
- [19] ZHANG Q, LIU T. Application of SVM and wavelet neural network method for short-term power load forecasting[C] // The 2nd International Conference on Computer and Automation Engineering (ICCAE), February 26-28, 2010, Singapore, Singapore: 412-416.
- [20] 牛东晓, 刘达, 陈广娟, 等. 基于遗传优化的支持向量机小时负荷滚动预测[J]. 电工技术学报, 2007, 22(6): 148-153.  
NIU Dongxiao, LIU Da, CHEN Guangjuan, et al. Support vector machine models optimized by genetic algorithm for hourly load rolling forecasting[J]. Transactions of China Electrotechnical Society, 2007, 22(6): 148-153.
- [21] 谢宏, 魏江平, 刘鹤立. 短期负荷预测中支持向量机模型的参数选取和优化方法[J]. 中国电机工程学报, 2006, 26(22): 17-22.  
XIE Hong, WEI Jiangping, LIU Heli. Parameter selection and optimization method of SVM model for short-term load forecasting[J]. Proceedings of the CSEE, 2006, 26(22): 17-22.
- [22] 张玉, 莫寒, 张烈平. 基于模糊支持向量机的光伏发电量预测[J]. 热力发电, 2017, 46(1): 116-120.  
ZHANG Yu, MO Han, ZHANG Lieping. Photovoltaic power generation prediction based on fuzzy support vector machine[J]. Thermal Power Generation, 2017, 46(1): 116-120.
- [23] 张然然, 刘鑫屏. 火电机组超短期负荷预测[J]. 热力发电, 2018, 47(7): 52-57.  
ZHANG Ranran, LIU Xinping. Ultra short term load forecast for thermal power units[J]. Thermal Power Generation, 2018, 47(7): 52-57.
- [24] 李若晨, 朱帆, 朱永利, 等. 结合受限玻尔兹曼机的递归神经网络电力系统短期负荷预测[J]. 电力系统保护与控制, 2018, 46(17): 83-88.  
LI Ruochen, ZHU Fan, ZHU Yongli, et al. Short-term load forecasting of power system based on recurrent neural network combined with restricted Boltzmann machine[J]. Power System Protection and Control, 2018, 46(17): 83-88.
- [25] 方八零, 李龙, 赵家铸, 等. 动态相似与静态相似相结合的短期负荷预测方法[J]. 电力系统保护与控制, 2018, 46(15): 29-35.  
FANG Baling, LI Long, ZHAO Jiazhu, et al. Short-term load forecasting method combining dynamic similarity with static similarity[J]. Power System Protection and Control, 2018, 46(15): 29-35.

收稿日期: 2018-01-10; 修回日期: 2018-03-13

作者简介:

席雅雯(1993—), 女, 通信作者, 硕士研究生, 主要研究方向为人工智能、负荷预测、数据挖掘; E-mail: 16121551@bjtu.edu.cn

吴俊勇(1966—), 男, 教授, 博士生导师, 主要研究方向为人工智能, 智能电网, 能源互联网, 储能技术。E-mail: wujy@bjtu.edu.cn

(编辑 姜新丽)