

DOI: 10.19783/j.cnki.pspc.191594

基于随机森林算法的短期电力负荷预测

李焱¹, 贾雅君², 李磊¹, 郝建姝¹, 张晓英¹

(1. 内蒙古电力(集团)有限责任公司包头供电局, 内蒙古 包头 014030; 2. 上海交通大学, 上海 200240)

摘要: 为了准确预测电力系统的短期负荷变化, 为电力系统安全、经济、高效运行提供指导方向, 提出了一种将模糊聚类以及随机森林回归算法进行组合的电力系统负荷预测方法, 利用粗糙集构建补偿规则, 对预测结果进行修正补偿。首先, 通过对电力系统负荷的周期性、天气相关性等特征进行分析, 利用C均值模糊聚类算法对历史样本进行聚类, 在进行随机森林回归预测时, 使用聚类后同类数据作为训练集样本构建决策树。考虑到随机森林回归预测偏保守、电力系统负荷在峰值处波动大的特征, 在得到预测结果后利用粗糙集理论生成补偿规则, 对负荷预测进行修正。利用所述方法对北爱尔兰地区进行一日24h的负荷预测, 结果跟实际负荷的平均绝对误差百分比为2.09%, 验证了该预测方法的有效性。

关键词: 短期电力负荷预测; 随机森林算法; C均值聚类; 粗糙集理论

Short term power load forecasting based on a stochastic forest algorithm

LI Yan¹, JIA Yajun², LI Lei¹, HAO Jianshu¹, ZHANG Xiaoying¹

(1. Baotou Power Supply Bureau of Inner Mongolia Power (Group) Co., Ltd., Baotou 014030, China;

2. Shanghai Jiao Tong University, Shanghai 200240, China)

Abstract: In order to accurately predict the short-term load change of a power system and provide guidance for safe, economic and efficient operation, a load forecasting method based on fuzzy clustering and random forest regression is proposed. A rough set is used to construct the compensation rules, and the prediction results are modified and compensated for. First, this paper analyzes the periodicity and weather correlation of power system load. Historical samples are clustered using C-mean fuzzy clustering. In the random forest regression prediction, similar data after clustering is used as a training set sample to build a decision tree. Taking into account the conservatism of partial random forest regression prediction and large fluctuations of power system load at the peak, the rough set theory is used to generate compensation rules after the prediction results are obtained, and load forecasting is modified. The 24-hour load forecasting of the Northern Ireland region using the above method shows that the Mean Absolute Percentage Error (MAPE) is 2.09% compared with the actual load, which verifies the effectiveness of the forecasting method.

This work is supported by National Natural Science Foundation of China (No. 51877136) "Key Technologies of New-type Electrical Energy Router Adapt to Distribution Network Flexible Interconnection".

Key words: short-term power load forecast; random forest algorithm; C-means clustering; rough set theory

0 引言

电力系统负荷预测研究是指以系统负荷本身的波动变化以及负荷所在地的外界因素变化为依据, 采用特定的数学方法或者建立数学模型的方式为手段, 通过对地区的电力负荷历史资料进行分析和研

究, 对电力系统的需求做出估计以及研究相关因素对电力负荷的影响^[1-2]。在建立了具体的预测系统后, 根据未来经济、人文、气象因素的发展趋势, 对电力需求进行了初步估计和推测。准确的预测, 特别是精准的短期负荷预测, 对提高电力运营商的运行效率有着直接的影响, 对电力系统的控制、运行和规划具有重要的指导意义^[3-4]。

短期负荷预测技术已经发展了几十年。随着数理统计理论和人工智能技术的发展, 国内外电力工

基金项目: 国家自然科学基金面上项目(51877136)“适用于配电网柔性互联的新型电能路由器关键技术”

作者提出了多种预测方法,以提高短期电力负荷预测的准确率和速度。目前,短期负荷预测方法大致可分为两类:传统预测方法和人工智能方法^[5-12]。属于传统预测方法的有时间序列法、回归分析法和指数平滑法等。传统预测方法的原理比较简单,理论比较成熟,至今仍在使用。但是一方面,这些方法采用的数学模型太过简易,以至于其参数难以及时准确地进行估计和调整,并且不能反映负载的突然变化。因此,它们不能适应使用更精确的数学模型来描述负载变化的规律。属于智能原理的有神经网络法、专家系统法、模糊逻辑法、模糊神经网络法、混合型方法等。基于神经网络等算法的现代智能算法具有自学习和自适应功能,近年来被广泛应用于非线性预测。然而,考虑短期负荷预测时,往往需要考虑气象、假期因素,需要输入网络模型中的参数过多,致使网络的结构复杂,网络训练费时。此外,也有学者提出了一些混合算法,但是都具有一定的局限性。鉴于此,本文提出了随机森林算法^[13-17]对短期电力系统负荷预测,结合模糊、聚类等大数据处理方法,通过对混合预测方法进行分析、验证,达到对一般的随机森林预测方法存在问题进行完善的目的,进一步提高预测的精度和效率。

本文首先对影响电力系统负荷的因素进行分析,获得了对负荷预测影响较大的几类因素,为后续数据处理、数据挖掘提供了方向;随后,对C均值模糊聚类进行分析,获得了相似日数据集并且建立了随机森林回归预测的模型。从理论上分析了利用粗糙集对随机森林或其他回归算法进行回归预测时引起的预测结果误差较大进行修正;最后以北爱尔兰地区为例建立了该地区的负荷预测模型,并利用粗糙集理论对预测结果进行了补偿,计算结果表明利用随机森林算法模型直接得到的预测结果的平均绝对误差百分比为2.32%,而在进行粗糙集补偿后的平均绝对误差百分比为2.09%。

1 电力系统负荷预测

负荷预测中通常会按时间期限来分类,在我国现行系统及实际需求中,电力负荷预测以时间尺度分为超短期、短期、中期、长期电力负荷预测。

(1) 超短期负荷预测

以分钟和小时为最小单位的负荷预测。主要作用是电力系统的潮流控制、电力系统暂稳态参数计算。

(2) 短期负荷预测

以天和周为最小单位的负荷预测。主要作用是用于电力系统功率分配方案的设计,以及负荷实际

的最优发电安排、发电量设计。本文进行的是以天(24 h)为单位的电力系统短期负荷预测^[12]。

(3) 中期负荷预测

以月甚至年为最小单位的负荷预测,主要是用于电力系统发电设备储备计划的制定以及对电力系统发电单位运行状态的确定。

(4) 长期负荷预测

以多年甚至结合当地工农业建设规划周期为最小单位的负荷预测。主要作用是用于电力系统的变电站等建设用地的规划,以及区域性电网、并网规划布局的确定等。

电力系统负荷预测具有时效长度的区别,针对以上的特点,分析可以得出影响电力系统负荷的变化具有规律性,但是影响电力系统负荷变化的因素是多而复杂且不确定的,尽管如此,对于一般的电力系统短期负荷预测而言,抓住其主要矛盾分析得到的精度即可满足需求。通过对影响短期负荷变化因素进行分析,经济、人口、气候等均对负荷变化有重大作用^[18-21],因此本文在后续进行短期负荷预测时着重考虑这些因素的影响。

2 基于随机森林的负荷预测模型

2.1 利用模糊聚类得到相似日数据集

2.1.1 C均值模糊聚类原理

对于给定的数据集 $X = \{x_1, x_2, \dots, x_n\}$, 要确定其C均值模糊聚类,需要输入类别数目 c , 每个聚类的中心 $m_j (j=1, 2, \dots, c)$ 。对于每一个样本 x_k , 相应的有隶属函数 $\mu_{jk} = \mu_{x_k}(x_k)$, 即第 k 个样本对于 X_i 类的隶属程度^[22]。则基于隶属函数的聚类损失指标函数可以表示为式(1)。

$$J = \sum_{j=1}^c \sum_{i=1}^n [\mu_j(x_i)]^b \|x_i - m_j\|^2 \quad (1)$$

式中, b 代表加权指数, 也称之为平滑因子, 用以表征一个样本在模糊类之间的分享程度。关于 b 的最优取值学术界仍然存在着争议, 考虑到计算量以及计算精度的原因, 通常情况下取加权指数为 2。

令 J 对 m_j 和 $\mu_j(x_i)$ 的偏导为 0, 可以求得 J 取得最小值的必要条件为式(2)、式(3)。

$$m_j = \frac{\sum_{i=1}^n [\mu_j(x_i)]^b x_i}{\sum_{i=1}^n [\mu_j(x_i)]^b} \quad (2)$$

$$\mu_j(x_i) = \frac{\|x_i - m_j\|^{-\frac{2}{b-1}}}{\sum_{s=1}^k \|x_i - m_s\|^{-\frac{2}{b-1}}} \quad (3)$$

因此,生成C均值模糊聚类的步骤为输入类别数目 c , 初始化每个聚类的中心 $m_j (j=1, 2, \dots, c)$ 。对于样本 x_k , 相应的隶属函数 $\mu_{ik} = \mu_{x_i}(x_k)$, 而后反复利用式(2)、式(3)求得 $m_j (j=1, 2, \dots, c)$ 、 $\mu_{ik} = \mu_{x_i}(x_k)$, 直到满足精度要求, 即可以确定聚类中心与隶属函数。

2.1.2 最优聚类数的确定

利用C均值模糊聚类进行聚类分析时, 必须先设置类别数目 C , C 的取值对聚类的影响深刻, 如果聚类数目过多, 会使原本应该是同类的样本被分到不同的类中。而如果聚类数目过少, 不同类的数据就可能被分到同一类中。错误聚类数目的设置, 会导致错误的聚类结果甚至使迭代无法收敛。因此, 需要通过设置优化准则对聚类数目进行推算。

对聚类数目进行推算的基本想法是通过引入结果评判指标, 通过对聚类数目进行递增测试, 判断评判指标的变化, 从最优指标结果就可以推算出最优的聚类数目。对于结果评判指标, 考虑到评价聚类结果主要从两个方面进行评价: 相同类的相似度、不同类的差异度, 因此可以确定评判指标为类内相似度 I_a 、类间相似度 I_r [23]。

$$I_{ai} = \frac{1}{n_i} \sum_{x \in X_i} \|x - c_i\| \quad (4)$$

式中: n_i 为类内数据的数目; x 为样本对象; c_i 为 X_i 类的中心。 I_{ai} 的值越小, 说明数据点越集中于类中心, 数据样本越相似。

$$I_{rij} = \min(\|c_i - c_j\|), i \neq j \quad (5)$$

式中: c_i 为 X_i 类的中心; c_j 为 X_j 类的中心。 I_{rij} 值越大, 说明相邻类的中心距离越远, 类间的相似程度越小。

因此, 可以得到综合评价指标:

$$J_i = \sum_{i=1}^c \frac{n}{n_i} I_{ai} + \sum_{i=1}^c \sum_{j=i+1}^c \frac{1}{I_{rij}} \quad (6)$$

因此, 最优聚类数目推算确定过程如下:

(1) 确定分类数目的范围, 通常设置分类数目 c 的范围为 $c \in [2, \sqrt{n}]$ [7], n 代表所有样本数据的数目。

(2) 对于每一个 c 值都进行一次聚类操作。

(3) 根据结果评判指标对每一次聚类得到的数据进行分析, 通常通过寻找评判指标中的拐点、差

分极小点等, 得到数据间的联系关系, 从而确定最优聚类数目。

2.1.3 相似日数据集形成

对于本文所涉及到的问题, 需要对输入的气象、日照数据进行聚类处理, 以求得到与被预测日相近的历史数据, 输入的气象数据包括降水量、最高气温、最低气温、风速、湿度、日照时长、气压数据等, 由于输入数据的量纲不一致, 而且预测模型中输入输出的数据单位不一致, 因此对输入的数据需要进行归一化处理, 归一化后数据不存在量纲, 且取值范围在 $[0, 1]$ 之间, 归一化处理的公式为

$$x_{nor} = \frac{x_{source} - x_{min}}{x_{max} - x_{min}} \quad (7)$$

在得到归一化的数据之后, 便可以利用C均值模糊聚类算法进行聚类处理, 通过最优聚类数目的判定, 最终得到历史数据各类的类中心, 记录分类结果以及类中心数据, 为后续预测做相似日判断。

2.2 随机森林回归预测模型的建立

随机森林预测的结果是由每颗决策树预测的结果组合而成, 因此随机森林预测的关键步骤即是决策树以及森林的形成, 如图1所示。

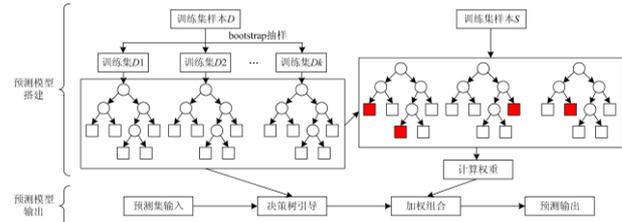


图1 随机森林预测流程图

Fig. 1 Random forest prediction flowchart

本文在对电力系统负荷预测时, 通过 bootstrap 算法抽取得到训练集及其特征后, 选取 CART 算法作为回归决策树的生成算法, 因此节点的分裂依据为最小均方差, 其计算方法如式(8)所示。

$$m = \min_{A,s} [\min_{c_1} \sum_{x_i \in D_1(A,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in D_2(A,s)} (y_i - c_2)^2] \quad (8)$$

式中: s 代表当前节点的所有训练集; A 代表当前节点被抽取出的特征集。训练集 s 根据特征 A 划分为子集 D_1, D_2 , 通过遍历 A 的取值, 算得子集 D_1, D_2 输出值 y_i 的最小均方差之和, 最后遍历所有的属性得到最小均方差的取值 m , 以及对应的属性、对应的属性的值, 得到该节点的生长信息。而后对生成的每一个子节点重复上述过程, 直到达到终止条件。终止条件设置为:

达到设置的树深度 d ;

节点上的样本少于最少样本数 s ;

最小均方差达到阈值 m_0 ;

在生成决策树后, 利用类似的方法, 对抽样得到的所有训练集进行训练, 便形成随机森林。

随机森林预测模型的搭建步骤如下:

(1) 首先利用 C 均值模糊聚类方法对原始数据处理, 得到相似日数据后, 将其分为两个部分: 训练集 D 与测试集 S , 记 D 的样本容量为 N , 记 D 中样本的属性容量为 M 。

(2) 从训练集 D 中利用 bootstrap 重抽样的方法抽取样本容量也为 N 的训练集, 重复 K 次, 便可以得到 K 个训练集 $\theta_1, \theta_2, \dots, \theta_K$, 每个训练集都可以生成与之对应的决策树 $\{T_1(\theta_1), T_2(\theta_2), \dots, T_K(\theta_K)\}$, 因此 K 为随机森林的决策树数。

(3) 利用训练集 $\theta_1, \theta_2, \dots, \theta_K$ 生成相应的 CART 决策树, 在生成决策树过程中, 对于每个分裂节点, 从 M 维属性特征中不放回地抽出 m 维属性特征, 遍历这些属性以及属性对应的值, 计算最小均方差, 得到最优的分割判据。

(4) 训练每一棵决策树, 直到达到终止条件。

利用建立好的随机森林模型, 代入测试集数据:

(1) 输入的数据经过归一化处理, 与相似日数据进行匹配, 代入与之相对应的相似日所形成的随机森林预测模型 $\{T_1(\theta_1), T_2(\theta_2), \dots, T_K(\theta_K)\}$ 。

(2) 分别代入模型的每一棵决策树中进行生长, 培养过程的分裂判据与建立模型类似。直至结束。之后统计每个非 0 叶子节点的权重, 权重计算公式如式(9)所示。

$$w_i = \frac{n(i \in I(\theta))}{N\{i \in S\}} \quad (9)$$

(3) 得到叶子节点的权重后, 根据式(10)可以计算决策树的权重。

$$w_i = \frac{\sum w_i Y_i}{\sum Y_i} \quad (10)$$

(4) 保存决策树的根与节点、权重等信息后, 随机森林预测模型建立完成。

随机森林预测模型建立完成后, 进行预测时的步骤如下:

(1) 输入的数据经过归一化处理, 与相似日数据进行匹配, 代入与之相对应的相似日所形成的随机森林预测模型 $\{T_1(\theta_1), T_2(\theta_2), \dots, T_K(\theta_K)\}$ 。

(2) 培养每棵决策树, 培养过程的分裂判据与建立模型类似。得到每棵树的预测结果后, 利用由测试集生成的权重数据, 进行加权平均便可以得到最终的预测结果。

$$y = \sum_{i=1}^K w_i Y_i \quad (11)$$

2.3 预测结果评判与粗糙集修正

2.3.1 预测结果评判标准

在利用随机森林回归模型进行预测得到结果后, 本文采用平均绝对误差、平均相对误差、 R^2 (此处用 R^2 来衡量预测结果与真实曲线的拟合度, 即相关系数)合格率对预测结果进行评判。

2.3.2 粗糙集理论

利用随机森林或其他回归算法进行回归预测时, 由于回归理论的特定, 会使预测结果趋于保守趋于平滑, 当被预测曲线存在多个峰值或者差方的变化较大时, 会使得预测结果在峰值附近的误差较大。

而粗糙集可以克服这种缺点, 其是由 Z.Pawlak 提出的一种处理大数据中模糊问题、不确定性问题的数学工具^[17,24], 能够有效地分析和处理不正确、需要误差补偿、不一致和有数据丢失的不完备信息, 从中挖掘得到隐含的知识, 揭示原始数据中潜在的规律。

2.3.3 粗糙集补偿系统构建

电力系统负荷补偿模型可以用式(12)表示。

$$\begin{cases} y'_{t+1} = y_{t+1} + s|k_{t+1} - k_t| \\ k_{t+1} = y_{t+2} - y_{t+1} \\ k_t = y_{t+1} - y_t \end{cases} \quad (12)$$

式中: y'_{t+1} 代表 $t+1$ 时刻补偿后的数据, 在 $t+1$ 时刻以前, 该值就代表实际的电力系统负荷; y_{t+1} 表示 $t+1$ 时刻随机森林预测的结果; k_{t+1} 、 k_t 代表预测函数 $t+1$ 时刻两侧的差分; s 是尺度因子, 是负荷预测修正的关键, 下文将用粗糙集理论求算。

为利用粗糙集理论对尺度因子 s 进行求算, 根据粗糙集理论的要求, 首先要构建一个信息系统。此处假设粗糙集理论所依的信息系统为 $K = (U, A)$, 其中集合 U 为随机森林预测输出的预测值集合。 $A = CUS$ 为属性集, 条件属性 C 代表一致的规则, 即从预测值集合中提炼的数据特征, $S = \{s\}$ 代表决策属性, 即需要求得的尺度因子。对于 C 中的每个属性, 都可以用 U 中含有的信息表示, 最终经过规则删减, 尝试获取到属性 S 的表示规则。

关于 C 所代表的条件属性, 结合前人研究结果, 此处定义条件属性 $C = \{a, b, c\}$, 其中:

$$a = \frac{|k_{t+1} - k_t|}{y_t} \quad (13)$$

$$b = \text{sgn}(k_{t+1} - k_t) \quad (14)$$

$$c = \left\lfloor \frac{y_t}{\max(y_t)} \right\rfloor \quad (15)$$

对于 D 所代表的决策属性, 在 t 之前的属性可以根据式(16)算出, $t+1$ 以后的值便是所要确定的决策目标。

$$s_t = \frac{y'_t - y_t}{|k_t - k_{t-1}|} \quad (16)$$

3 基于随机森林的负荷预测实例分析

3.1 实例数据处理

依据上文分析所建立的电力系统负荷预测模型, 本文选取了北爱尔兰地区的气象信息、历史负荷数据作为原始数据, 继续模型建立以及负荷预测分析^[24]。

影响电力系统负荷变化的因素很多, 但由第二章的分析中可以得到, 气象因素、人类活动因素是影响电力系统负荷的较为重要的因素^[7], 由此本文考虑选择的日气象因素、实时气象因素参数。其中日气象信息主要包含最高气温、最低气温、平均气温、平均气压、降水量、平均风速、光照时间、湿度; 实时气象信息主要包含实时温度、实时湿度、实时气压、实时降雨量、实时能见度、实时云量、实时风速、实时日照。

除了气象因素, 影响电力系统负荷变化的还有人类活动因素, 主要有星期信息、节假日信息。

考虑到发电机的惯性特性以及社会经济运动的规律, 电力系统负荷在短时间内的变化较小, 也就是说, 近期的电力系统负荷对短期电力系统负荷的预测具有较大的参考价值, 因此这里也将近期电力系统的负荷作为输入参数用于随机森林的预测之中, 如表1所示。

表1 输入近期负荷参数信息汇总

Table 1 Enter summary of recent load parameter information

类别	参数信息
日负荷信息	最大日负荷
	最小日负荷
实时负荷信息	相似日同时段负荷均值
	前日同时段负荷

由于输入的参数类型较多, 而作为输出的是电力系统负荷, 输入与输出之间的量纲和代表物理意义的单位不同, 因此, 在进行模型建立之前, 需要将上述输入信息按照式(7)进行归一化处理。

3.2 实例预测模型构建

3.2.1 相似日数据形成

根据C均值聚类算法原理, 对历史数据进行聚

类。聚类的基本思想是通过将日天气信息相似的日子进行归档, 对天气差异较大的历史日子进行分类, 以降低环境因素对电力系统负荷预测的影响。

聚类时取日天气信息作为特征, 包括日最高气温、日最低气温、日平均气压、日降水量、日平均风速、日光照时间、日湿度, 同样地输入的数据需要按照式(7)进行归一化处理。设定分类数目 $c \in [2, 30]$, 分别依据每一个 c 值进行一次聚类操作, 根据聚类结果计算综合评价指标 J_t , 绘出其变化趋势图如图2所示。

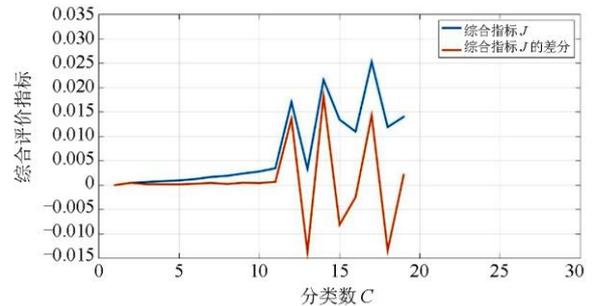


图2 综合指标与聚类数关系图

Fig. 2 Relationship between comprehensive index and cluster number

由结果可以看出, 当聚类数目 c 等于13, 即分为13类时, 综合指标的一阶差分最小, 这意味着, 聚类数目 c 等于13是最优聚类数目。确定了聚类数目后, 便可以将历史天数做相似日归档, 同时得到各个类的聚类中心, 为相似日的判断提供依据。表2提供了各类参数归一化后的聚类中心。

表2 各类的聚类中心(归一化后)

Table 2 Various cluster centers (after normalization)

类别	最高温	最低温	降水量	压强	风速	日照	湿度
1	0.391	0.441	0.067	0.528	0.500	0.108	0.163
2	0.696	0.685	0.020	0.664	0.312	0.437	0.599
3	0.387	0.420	0.041	0.632	0.332	0.156	0.167
4	0.390	0.424	0.042	0.633	0.332	0.155	0.168
5	0.389	0.423	0.042	0.633	0.332	0.156	0.168
6	0.720	0.627	0.007	0.715	0.246	0.716	0.705
7	0.379	0.415	0.044	0.613	0.349	0.155	0.164
8	0.620	0.677	0.038	0.660	0.284	0.133	0.351
9	0.317	0.328	0.024	0.679	0.278	0.193	0.126
10	0.581	0.504	0.014	0.701	0.248	0.519	0.484
11	0.492	0.545	0.040	0.658	0.320	0.144	0.234
12	0.672	0.689	0.026	0.664	0.289	0.257	0.487
13	0.545	0.603	0.040	0.655	0.311	0.134	0.274

3.2.2 随机森林负荷预测

本文以 20140101-20180131 的数据作为历史样本,对 20180201 的电力系统负荷进行预测。在进行相似日聚类后,利用编程工具,依据第 2 节所述方法建立随机森林回归预测模型,获得决策树数目,在利用训练集、测试集对随机森林模型进行搭建后,得到 500 个决策树模型^[25],同时计算得出节点的权重信息。再利用建立好的模型进行回归预测,可以得到 2018-02-01 当天的预测结果如图 3 所示。

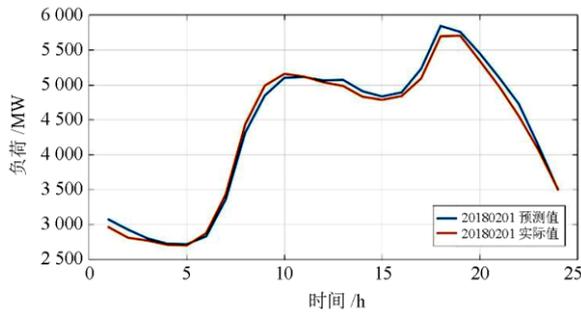


图 3 20180201 负荷结果与实际负荷图

Fig. 3 Load result and actual load graph of 20180201

3.3 预测结果分析与归纳

3.3.1 预测结果评价

根据常用的预测结果评价指标对预测结果进行分析,主要计算平均误差ME(MW)、平均绝对误差MAE(MW)、平均绝对误差百分比MAPE、 R^2 以及合格率通过各个指标的计算方法可以获得各个指标值,如表3所示。

表 3 预测结果评价指标

Table 3 Forecast result evaluation index

指标名称	指标值
ME/MW	85.587 8
MAE/MW	92.039 5
MAPE	2.32%
R^2	0.988 3
合格率3%	58.33%
合格率5%	95.83%

3.3.2 粗糙集修正

从上述预测结果可以看出,对于电力系统短期负荷预测而言,的确存在峰值、突变值误差。因此,可以利用2.3.2引入的粗糙集理论对负荷预测结果进行修正,从而提高电力系统负荷预测的精度。

依据式(13)一式(15)分别计算条件属性 $C = \{a, b, c\}$, 以及在 t 之前的决策属性 D , 从而得到粗糙集信息系统。

鉴于粗糙集理论处理数据的要求,需要对条件

属性、决策属性进行等频离散化。在进行等频离散化时,需要设定编码对应规则,此处设定条件属性 $C = \{a, b, c\}$ 对应的编码规则为 $C = \{a \in [1, 7], b \in [1, 3], c \in [1, 7] | a, b, c \in Z\}$ 以及 $D = \{d \in [1, 106], d \in Z\}$ 。编码后便可以得到决策表。由于决策表包含了所有的决策数据,篇幅较长,此处摘录前6条决策表信息如表4所示。

表 4 决策表信息(节选)

Table 4 Decision table information (excerpt)

a	b	c	d
2	1	1	97
5	3	1	90
3	3	1	94
4	3	1	27
5	3	1	29
5	3	1	28

依据决策表,对条件属性、决策规则进行约简,即推算得到决策属性由条件属性决定的最简关系。约简过程是一个 NP 问题,文章通过遗传算法进行求解。得到结果后,可以计算出各个时刻的决策属性,从而得到补偿后的负荷预测数据如图 4 所示。

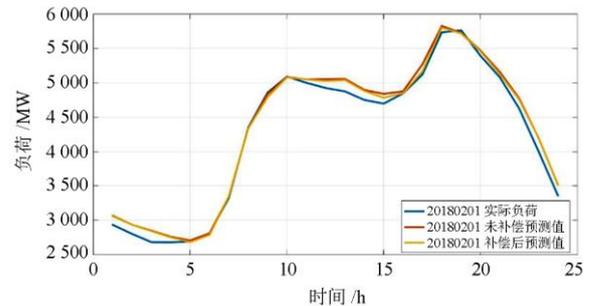


图 4 利用粗糙集进行预测修正

Fig. 4 Predictive correction using rough sets

对于补偿后的数据,根据再次计算几个对补偿结果评价的指标,如表5所示。

表 5 补偿后的预测指标

Table 5 Compensated forecast index

指标名称	未补偿	补偿后
ME/MW	85.587 8	67.144 5
MAE/MW	92.039 5	80.659 6
MAPE	2.32%	2.09%
R^2	0.988 3	0.990 6
合格率3%	58.33%	75.00%
合格率5%	95.83%	95.83%

由表 5 可以看出,进行补偿后,模型的输出结果精度更高,预测的结果更加接近实际值。

4 结语

综上所述, 本文提出了一种将模糊聚类以及随机森林回归算法进行组合的电力系统负荷预测方法。文章首先通过分析归纳获得影响电力系统负荷变化的因素, 为建立随机森林网络负荷预测模型提供基础。其次利用聚类算法对历史数据中的相似日进行了划分, 使得预测的样本数据更有针对性并讨论了预测参数的选取问题, 建立基于随机森林的负荷预测模型采用粗糙集算法对预测结果进行修正。最后, 通过引入北爱尔兰地区的数据作为实例进行了计算, 利用随机森林算法模型直接得到的预测结果与实际值的平均绝对误差百分比为2.32%, 而在进行粗糙集补偿后的平均绝对误差百分比降低为2.09%。总之, 通过实例算法验证了本文提出的算法的有效性, 并且采用粗糙集补偿算法后预测结果更加准确。

参考文献

- [1] 黄青平. 基于随机森林的电力系统短期负荷预测研究[D]. 北京: 华北电力大学, 2018.
HUANG Qingping. Study on power system short-term load forecast based on random forest[D]. Beijing: North China Electric Power University, 2018.
- [2] 康重庆, 夏清, 刘梅. 电力系统负荷预测[M]. 北京: 中国电力出版社, 2017.
- [3] 黎祚, 周步祥, 林楠. 基于模糊聚类与改进 BP 算法的日负荷特性曲线分类与短期负荷预测[J]. 电力系统保护与控制, 2012, 40(3): 56-60.
LI Zuo, ZHOU Buxiang, LIN Nan. Classification of daily load characteristics curve and forecasting of short-term load based on fuzzy[J]. Power System Protection and Control, 2012, 40(3): 56-60.
- [4] 廖旋焕, 胡智宏, 马莹莹, 等. 电力系统短期负荷预测方法综述[J]. 电力系统保护与控制, 2011, 39(1): 147-152.
LIAO Nihuan, HU Zhihong, MA Yingying, et al. Review of the short-term load forecasting methods of electric power system[J]. Power System Protection and Control, 2011, 39(1): 147-152.
- [5] 刘春霞, 张雪艳. 改进人工智能神经网络的短期电力负荷预测[J]. 电气应用, 2013(4): 74-77.
LIU Chunxia, ZHANG Xueyan. Short-term power load prediction based on improved artificial intelligence neural network[J]. Electrotechnical Application, 2013(4): 74-77.
- [6] 张乔榆, 蔡秋娜, 刘思捷, 等. 基于样本扩展和特征标记的节假日短期负荷预测[J]. 广东电力, 2019, 32(7): 67-74.
ZHANG Qiaoyu, CAI Qiuna, LIU Sijie, et al. Holiday short-term load forecasting based on sample expansion and feature extraction[J]. Guangdong Electric Power, 2019, 32(7): 67-74.
- [7] 邢书豪, 高广玲, 张智晟. 基于双层随机森林算法的短期负荷预测模型[J]. 广东电力, 2019, 32(9): 160-166.
XING Shuhao, GAO Guangling, ZHANG Zhisheng. Short-term load forecasting model based on double-layer random forest algorithm[J]. Guangdong Electric Power, 2019, 32(9): 160-166.
- [8] 王志勇, 郭创新, 曹一家. 基于模糊粗糙集和神经网络的短期负荷预测方法[J]. 中国电机工程学报, 2005, 25(19): 7-11.
WANG Zhiyong, GUO Chuangxin, CAO Yijia. A method short term load forecasting integrating fuzzy-rough with artificial neural network[J]. Proceedings of the CSEE, 2005, 25(19): 7-11.
- [9] 张丽, 张涛, 王福忠, 等. 基于柔性负荷响应特性的超短期预测方法[J]. 电力系统保护与控制, 2019, 47(9): 27-34.
ZHANG Li, ZHANG Tao, WANG Fuzhong, et al. Ultra-short-term forecasting method based on response characteristics of flexible load[J]. Power System Protection and Control, 2019, 47(9): 27-34.
- [10] HUANG L, YANG Y, ZHAO H, et al. Time series modeling and filtering method of electric power load stochastic noise[J]. Protection and Control of Modern Power Systems, 2017, 2(3): 269-275. DOI: 10.1186/s41601-017-0059-8.
- [11] PAPPAS S S, EKONOMOU L, KARAMELAS P, et al. Electricity demand load forecasting of the Hellenic power system using an ARMA model[J]. Electric Power Systems Research, 2010, 80(3): 256-264.
- [12] 彭显刚, 郑伟钦, 林利祥, 等. 考虑负荷自适应检测和修复的鲁棒极限学习机短期负荷预测方法[J]. 中国电机工程学报, 2016, 36(23): 6409-6417.
PENG Xiangang, ZHENG Weiqin, LIN Lixiang, et al. Short-term load forecasting method based on outlier robust extreme learning machine considering adaptive load detection and repair[J]. Proceedings of the CSEE, 2016, 36(23): 6409-6417.
- [13] 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.
- [14] 骆小满, 皇甫成, 阮江军, 等. 基于神经网络的热电联产机组热负荷和电负荷预测[J]. 热力发电, 2019, 48(9): 46-50.
LUO Xiaoman, HUANGFU Cheng, RUAN Jiangjun, et al. Prediction of heat and electric load of cogeneration unit based on neural network[J]. Thermal Power Generation, 2019, 48(9): 46-50.
- [15] 刘翊枫, 周国鹏, 刘昕, 等. 基于智能相似日识别及偏差校正的短期负荷预测方法[J]. 电力系统保护与控制, 2019, 47(12): 138-145.
LIU Yifeng, ZHOU Guopeng, LIU Xin, et al. A short-term load forecasting method based on intelligent similar day recognition and deviation correction[J]. Power System Protection and Control, 2019, 47(12): 138-145.
- [16] SELVAKUMAR J, LAKSHMI A, ARIVOLI T. Brain tumor segmentation and its area calculation in brain MR images using K-mean clustering and Fuzzy C-mean algorithm[C] // IEEE International Conference on Advances in Engineering, Science and Management, March 30-31, 2012, Nagapattinam, Tamil Nadu, India: 186-190.
- [17] LAHOUAR A, SLAMA J B H. Day-ahead load forecast using random forest and expert input selection[J]. Energy Conversion & Management, 2015, 103: 1040-1051.
- [18] CHENG Y Y, CHAN P P K, QIU Z W. Random forest based ensemble system for short term load forecasting[C] // International Conference on Machine Learning and Cybernetics, July 15-17, 2012, Xi'an, China: 52-56.
- [19] ZATSEPIN I, SV ÍTEK M. Night earth observation for smart cities[C] // 2015 Smart Cities Symposium Prague (SCSP), June 24-25, 2015, Prague, Czech Republic: 1-4.
- [20] 杨国田, 张涛, 王英男, 等. 基于长短期记忆神经网络的火电厂 NO_x 排放预测模型[J]. 热力发电, 2018, 47(10): 12-17.
YANG Guotian, ZHANG Tao, WANG Yingnan, et al. Prediction model for NO_x emissions from thermal power plants based on long-short-term memory neural network[J]. Thermal Power Generation, 2018, 47(10): 12-17.
- [21] 周勇. 电力系统短期负荷预测的研究[D]. 西安: 西安理工大学, 2001.
ZHOU Yong. Research on power system short-term load forecast[D]. Xi'an: Xi'an University of Technology, 2001.
- [22] 高新华, 严正. 基于主成分聚类分析的智能电网建设综合评价[J]. 电网技术, 2013, 37(8): 2238-2243.
GAO Xinhua, YAN Zheng. Comprehensive assessment of smart grid construction based on principal component analysis and cluster analysis[J]. Power System Technology, 2013, 37(8): 2238-2243.
- [23] 宋清昆, 郝敏. 一种改进的模糊 C 均值聚类算法[J]. 哈尔滨理工大学学报, 2007, 12(4): 8-10.
SONG Qingkun, HAO Min. Improved fuzzy C-means clustering algorithm[J]. Journal of Harbin University of Science and Technology, 2007, 12(4): 8-10.
- [24] 焦润海, 苏辰隽, 林碧英, 等. 基于气象信息因素修正的灰色短期负荷预测模型[J]. 电网技术, 2013, 37(3): 720-725.
JIAO Runhai, SU Chenjun, LIN Biying, et al. Short-term load forecasting by grey model with weather factor-based correction[J]. Power System Technology, 2013, 37(3): 720-725.
- [25] LAHOUAR A, SLAMA J B H. Hour-ahead wind power forecast based on random forests[J]. Renewable Energy, 2017, 109: 529-541.

收稿日期: 2019-12-24; 修回日期: 2020-03-16

作者简介:

李焱(1977—), 女, 硕士, 高级工程师, 主要从事用电营销方向; E-mail: 10755106@qq.com.cn

贾雅君(1973—), 男, 博士, 副研究员, 从事电力系统设备状态监测、配网自动化、继电保护;

李磊(1979—), 男, 本科, 工程师, 主要从事生产管理方向。

(编辑 张爱琴)