

DOI:10.11918/201911084

蚁群算法分配权重的燃气日负荷组合预测模型

周 洲^{1,2}, 焦文玲^{1,2}, 任乐梅^{1,2}, 田兴浩^{1,2}

(1. 哈尔滨工业大学 建筑学院, 哈尔滨 150090; 2. 寒地城乡人居环境科学与技术工业和信息化部重点实验室
(哈尔滨工业大学), 哈尔滨 150090)

摘 要: 为适应城镇燃气日负荷随机性和多变性的特点, 克服特定时刻单一负荷预测模型存在实际应用局限性的问题, 将5种评价准则用于组合预测前剔除冗余模型, 提出了一种建立变全重组预测模型的方法, 通过蚁群算法确定分配权重的组合预测模型, 使得在一个时段上的燃气日负荷预测精度好于各单一模型. 首先对包含诸多随机和模糊等不确定因素的城镇燃气日负荷时变系统和各预测模型特点进行分析; 然后确定岭回归分析(Ridge)、差分自回归积分滑动平均模型(ARIMA)、支持向量机回归(SVR)、极端梯度提升树(XGB)共4类单项目日负荷预测模型, 结合城镇燃气日负荷和模型的特点, 分别给出每个模型各项参数的设置和模型的输入向量; 用平均相对误差、均方根误差、灰色关联度、相关系数、Theil不等系数为评价准则计算出的综合评价指标剔除冗余模型, 最后建立了蚁群算法权重分配的组合预测模型. 预测实例表明, 蚁群算法分配权重的燃气日负荷组合预测模型长期的综合预测效果要优于任意单项模型, 相比于单一模型而言, 组合预测模型的稳定性和容错率更高, 具备较强的泛化能力.

关键词: 城镇燃气日负荷; 组合预测; 岭回归; 差分自回归积分滑动平均; 支持向量机回归; 极端梯度提升树; 蚁群算法

中图分类号: TU996

文献标志码: A

文章编号: 0367-6234(2021)06-0177-07

Combined forecasting model of gas daily load based on weight distribution of ant colony algorithm

ZHOU Zhou^{1,2}, JIAO Wenling^{1,2}, REN Lemei^{1,2}, TIAN Xinghao^{1,2}

(1. School of Architecture, Harbin Institute of Technology, Harbin 150090, China;

2. Key Laboratory of Cold Region Urban and Rural Human Settlement Environment Science and Technology
(Harbin Institute of Technology), Ministry of Industry and Information Technology, Harbin 150090, China)

Abstract: Urban gas daily load has the characteristics of randomness and variability, and single load forecasting models have certain limitations in practical applications, especially for specific time periods. To tackle such problem, five evaluation criteria were used to eliminate redundant model before combination forecasting, and a method of constructing variable weight combined forecasting model was proposed. The distribution weights of the combined forecasting model were determined by ant colony algorithm, so that the precision of the combined forecasting model was better than single models in a certain time period. First, the time-varying system of urban gas daily load and the characteristics of each forecasting model were analyzed, which contains many random and fuzzy uncertainties. Then, four kinds of single daily load forecasting models were determined, including ridge regression (Ridge), differential autoregressive integral moving average model (ARIMA), support vector machine regression (SVR), and extreme gradient lifting tree (XGB). According to the characteristics of urban gas daily load model, the parameters of each model and the input vector of the model were given. The redundant model was eliminated by using the comprehensive evaluation indexes calculated by the average relative error, root mean square error, grey correlation degree, correlation coefficient, and Theil unequal coefficient as the evaluation criteria. Finally, based on the ant colony algorithm, a weight distribution combination forecasting model was developed. Study results show that the long-term comprehensive prediction effect of the proposed model was superior to that of any single model. Compared with a single model, the combined forecasting model had higher stability and fault tolerance rate, and stronger generalization ability.

Keywords: urban gas daily load; combined forecasting; ridge regression; differential autoregressive integral moving average; support vector machine regression; extreme gradient lifting tree; ant colony algorithm

收稿日期: 2019-11-14

作者简介: 周 洲(1996—), 男, 硕士;

焦文玲(1962—), 女, 教授, 博士生导师

通信作者: 焦文玲, wljiao@163.com

能源清洁化是全球城市现代化发展的一个重要标志, 天然气作为最清洁的化石能源和可持续发展的重要能源, 已成为应对气候变化和能源转型的必

然选择^[1],城市燃气日负荷的准确预测关乎于居民生活的用气保障、燃气公司的经济效益、城市智慧燃气的发展等多方面问题.

国内外学者对城镇燃气日负荷组合预测展开了相关研究,Beyzanur 等^[2]采用自回归移动平均法和遗传算法相结合的方法对土耳其伊斯坦布尔的燃气日负荷进行预测,结果表明该组合预测方法具有较强的鲁棒性.Ervural 等^[3]用于预测的递归神经网络模型的输入向量的选取来源于灰色理论模型计算得出的级比残差值序列和模糊理论得到的输出向量,预测结果表明该组合方法优于传统的递归神经网络模型.祖广磊等^[4]在对原始数据进行离群点预处理的基础上,将燃气负荷分解成为趋势性和季节性两种序列,分别利用 ARIMA 模型和 BP 神经网络进行预测,最后采用信息熵理论,将两种模型产生的预测结果进行加权组合.陈庆明^[5]分别建立了回归分析、支持向量机、小波神经网络等模型,并以平均相对误差为目标函数,建立了组合预测模型,实验结果表明组合预测模型的预测精度更高.在上述研究中建立的组合预测模型多基于不变权理论,参与组合预测的单项预测模型往往缺少前期的模型筛选过程,权重计算的目标函数多基于单一准则.

为此,本文建立了岭回归分析(Ridge)、差分自回归积分滑动平均模型(ARIMA)、支持向量机回归(SVR)、极端梯度提升树(XGB)共4类单项预测模型,在组合预测前,利用多个评价准则计算出的综合评价指标对模型进行筛选,最后基于蚁群算法对组合预测模型的权重系数进行计算,通过预测实例,对比了组合预测模型与各单项预测模型的预测效果.

1 单项预测模型的建立

城镇燃气日负荷预测是企业调度部门日常工作的重点和难点.日负荷影响因素多,当地气候因素、各类燃气用户用气规律都不同程度的决定着日负荷的变化和趋势;另外周末、节假日和平常工作日的用气规律也有很大不同.上述各种因素对负荷的影响有些具有一定的相关性和规律性,如负荷与气温成反相关关系,同时存在以年为周期的变化,如图1所示.但是燃气负荷系统存在更多的不确定性和随机性,因此预测对象复杂、难度大,选择多样性的燃气负荷预测模型并组合优化可以改善在一个时段上的预测效果.

现有的众多预测模型,在建模机理和建模采用数据方面都有差异,比如:针对燃气负荷序列的ARIMA模型,要求时间序列大样本,通过差分处理

后能够满足准平稳性要求.负荷与气温的相关性表明基于回归原理的岭回归分析适于用气象和历史负荷建模.人工智能等算法近年来被推广应用于预测领域,基于统计学习理论的支持向量机(SVM)模型和极端梯度提升树模型等,通过将输入向量映射到更高维空间,可以按任意精度逼近非线性函数,类似于模式识别的算法同样适于探讨复杂多因素变量的燃气负荷预测系统.因此本文选择下面4种方法建立单项预测模型.

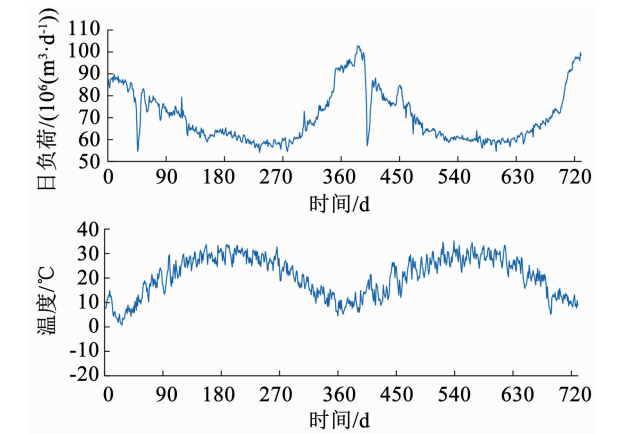


图 1 燃气日负荷及日平均温度变化曲线
Fig. 1 Variation of daily gas load and average temperature with time

1.1 岭回归分析

典型线性回归基于最小二乘法的代价函数假设残差服从均值为零的高斯分布^[6],然而实际燃气日负荷的数据所服从的概率分布很难用一个或多个分布的混合去拟合,则需要引入正则化因子,而岭回归则是结构风险最小化的正则化因子上使用模型参数向量的二范数形式,它是一种用于共线性数据分析的有偏估计回归方法,通过放弃无偏性去获得更符合实际的回归系数^[7].

本文基于 Python 的机器学习 Sklearn 库中的岭回归分析模块进行建模,模型各项参数设置见表1,岭回归模型的具体输入向量见表2.

表 1 岭回归模型参数设置表		
Tab. 1 Parameters in ridge regression model		
参数名称	参数说明	参数值
alpha	正则化强度	10 000
fit_intercept	是否计算模型截距	True
max_iter	共轭梯度求解器的最大迭代次数	1 000
normalize	是否进行归一化处理	True
solver	求解器选择	Sag(随机平均梯度下降)

表 2 岭回归模型输入向量表

输入特征	说明
$X_{(t-n-5)} \sim X_{(t-1)}$	预测起始日前 $n+5$ d 的各日负荷
$X_{(t-n-370)} \sim X_{(t-n-360)}$	预测起始日的上一年阳历对应日的前后 $n+5$ d 的各日负荷
$T_{(t-5)} \sim T_{(t-1)}$	预测起始日前 5 d 的平均温度、最高温度、最低温度
$W_{(t-5)} \sim W_{(t-1)}$	预测起始日前 5 d 的气象数据
$D_{(t)} \sim D_{(t+n)}$	预测日的星期类型

输入向量近似按 365 d 为负荷周期选择,这里忽略农历的周期性是因为闰年等变化可能仅对部分季节或时段负荷的影响较大,本文作为预测方法运用的研究目的,简化为负荷以 365 d 为周期.

1.2 ARIMA 时间序列模型

ARIMA(p,d,q)模型为差分自回归积分滑动平均模型,参数 p 为自回归项,参数 q 为滑动平均项,参数 d 为数据差分化处理的次数^[8].

本文建模时间序列为预测起始日的前 180 d 的负荷数据,建模流程如下:

1)连接数据库,获取需要的日负荷时间序列数据,依次绘制时序散点图、ACF(自相关系数)图和 PACF(偏自相关系数)图.

2)如果序列 ADF 值大于检验水平的临界值,则表明该负荷数据序列是非平稳的,则对序列进行一阶差分处理,处理后再进行差分检验,直至满足平稳性要求.

3)采取“网格搜索”的方式来确定 ARIMA(p,d,q)时间序列模型的最优参数值.由定义参数三元组的不同组合分别计算 AIC(Akaike 信息标准)值来对模型进行训练和评估.

4)对模型进行检验,通过假设检验诊断残差序列是否具有白噪声,通过参数估计检验模型的统计意义.

5)调用模型预测,并对产生的预测值进行差分还原处理.

1.3 支持向量机回归

支持向量机(SVM)是基于统计学习理论的 VC 维理论和结构风险最小化理论的机器学习理论,它是一种二分类模型,通过间隔最大化的分割原理来寻找一个超平面来分类样本^[9].由于其具有全局极小值点和收敛速度快的优点,且可以按任意精度逼近非线性函数,近年来被推广应用于预测领域,即基于支持向量机理论的 SVR 模型. SVR 的建模过程是

通过选取的核函数(特征转换函数)将输入向量映射(分线性映射)到更高维空间并确定超平面.

本文基于 Python 中的 Sklearn 库中的 SVM 类中的 SVR 模块建立预测模型.以均方误差为目标准则,通过 K 折交叉验证法,动态选取模型的核函数以及包括惩罚因子 C 、核系数 γ 、不敏感损失系数 ϵ .参与交叉验证试算的各参数取值范围见表 3,SVR 模型的输入向量见表 4.

表 3 SVR 模型参数设置表

参数名称	参数取值范围
核函数	[Linear、poly、RBF、Sigmod]
惩罚因子 C	[10,40,70,100,300,1 000]
不敏感损失系数 ϵ	[0.01,0.05,0.10,0.50,1.00]
核系数 γ	[0.01,0.05,0.10,0.50,1.00]

表 4 SVR 模型输入向量表

输入特征	说明
$X_{(t-n-5)} \sim X_{(t-1)}$	预测起始日前 $n+5$ d 的各日负荷
$X_{(t-n-370)} \sim X_{(t-n-360)}$	预测起始日的上一年阳历对应日的前后 $n+5$ d 的各日负荷
$T_{(t-5)} \sim T_{(t-1)}$	预测起始日前 5 d 的平均温度、最高温度、最低温度
$W_{(t-3)} \sim W_{(t-1)}$	预测起始日前 3 d 的气象数据
$D_{(t)} \sim D_{(t+n)}$	预测日的星期类型

1.4 XGB 模型

XGB 的全称是 eXtreme Gradient Boosting,它是梯度提升机器算法(GBDT)的扩展,即极端梯度提升树^[10]. Boosting 分类器是集成学习模型的一种,其基本思路是把大量的分类精度较低的单个树模型进行组合,构成一个稳定性和精确度较高的模型. XGB 相比于传统的 GBDT 算法,具有速度快、效果好、能处理大规模数据、支持多种语言、支持自定义损失函数等多种优点.

本文基于 Python 中的 xgboost 库建立预测模型,建立 XGB 模型需要设置 3 类参数,第 1 类是控制提升过程中决定 booster 使用类型的通用参数,第 2 类是选取的 booster 的内置参数,第 3 类是控制学习场景的学习目标参数.本文选取树模型作为梯度提升算法,通过 k 折交叉验证试算法确定其他各项参数,参与交叉验证的各参数取值范围见表 5, XGB 模型的输入向量见表 6.

表 5 XGB 模型参数设置表
Tab.5 Parameters in XGB model

参数名称	参数说明	参数值
eta	防止过拟合,更新过程的收缩步长	[0.1,0.3,1.0]
gamma	节点分裂所需的最小损失函数下降值	[0,1,10,100]
max_depth	树的最大深度	[1,3,6,10]
max_delta_step	树的权重被估计的值	[1,3,6,10]
lambda	L2 正则的惩罚系数	[1,10,100,1 000]
alpha	L1 正则的惩罚系数	[1,10,100,1 000]
objective	学习目标函数	logistic
max_iter	最大迭代次数	1 000

表 6 XGB 模型输入向量表
Tab.6 Vector parameters in XGB model

输入特征	说明
$X_{(t-n-3)} \sim X_{(t-1)}$	预测起始日前 $n+3$ d 的各日负荷
$X_{(t-n-368)} \sim X_{(t-n-362)}$	预测起始日的上一年阳历对应日的前后 $n+3$ d 的各日负荷
$T_{(t-3)} \sim T_{(t-1)}$	预测起始日前 3 d 的平均温度、最高温度、最低温度
$W_{(t-3)} \sim W_{(t-1)}$	预测起始日前 3 d 的气象数据
$D_{(t)} \sim D_{(t+n)}$	预测日的星期类型

2 组合预测模型的建立

2.1 基于多准则评价指标的模型筛选

基于权重分配方式的组合预测实际上是单项预测模型的信息进行选择利用的过程^[11],在过去多数燃气日负荷组合预测的相关文献中,参与组合预测的单项预测模型大都是通过预测研究人员逐一试算各预测模型效果后,主观地根据已有的经验选择单项预测模型直接建立组合预测模型.但是实际预测中的某些特定时间段,并非所有的预测模型都对组合预测的精准度提升有帮助,也不是其提供的信息一定具有价值.因此,有必要在组合预测前剔除这些冗余模型,否则不仅会影响组合预测模型的预测精度,而且还会增加计算时间.

评价预测模型的预测性能以及衡量模型泛化能力需要通过对预测结果进行度量分析,构建这种度量分析的损失函数需要基于不同的评价指标,不同的指标的侧重点和适应范围均不一样.本文选取以下 5 种常见的评价准则来度量预测值与真实值之间的误差.

1)平均相对误差(MAPE).

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{x^{(0)}(t) - \hat{x}^{(0)}(t)}{x^{(0)}(t)} \right|. \tag{1}$$

式中:MAPE($x^{(0)},\hat{x}^{(0)}$)为预测负荷序列与真值负荷序列的平均相对误差; $\hat{x}^{(0)}(t)$ 为第 t 日的预测燃气负荷, m^3 ; $x^{(0)}(t)$ 为第 t 日的真值燃气负荷, m^3 ; n 为预测序列中的预测日数.

2)均方根误差(RMSE).

$$RMSE = \frac{1}{n} \sqrt{\sum_{t=1}^n \left[\frac{x^{(0)}(t) - \hat{x}^{(0)}(t)}{x^{(0)}(t)} \right]^2}, \tag{2}$$

式中,RMSE($x^{(0)},\hat{x}^{(0)}$)为预测负荷序列与真值负荷序列的均方根误差.

3)灰色关联度. 预测值负荷序列与真值负荷序列的关联程度可以用两条曲线几何形状的相似度来表示,即两序列间的差值可作为其关联程度的衡量标准^[12]. 预测值数列与真值数列在各个时刻(即曲线中的各点)的关联系数 ξ_x 可由下式计算得出:

$$\xi_x = \frac{\Delta(\min) + \rho\Delta(\max)}{\Delta_x(k) + \rho\Delta(\max)}. \tag{3}$$

式中: ξ_x 为预测负荷序列与真值负荷序列上各点的关联系数; $\Delta_x(k)$ 为预测负荷序列与真值负荷序列上第 k 点的绝对值差值, m^3 ; ρ 为分辨系数,一般在 0~1 之间,通常取 0.5.

因为关联系数是预测值负荷序列与真值负荷序列在各个时刻的关联程度值,所以为了便于进行整体性比较,需求各个时刻的关联系数的平均值,即为两序列的灰色关联度,当二序列的灰色关联度越大,则表明关联程度越高.

$$\gamma_i = \frac{1}{N} \sum_{k=1}^N \xi_i(k). \tag{4}$$

式中: γ_i 为预测负荷序列与真值负荷序列的灰色关联度; $\xi_i(k)$ 为预测负荷序列与真值负荷序列上第 k 点的关联系数; N 为预测日长, d.

4)相关系数. 相关系数常用于度量预测的精准度,当预测值序列越接近真值序列时,两者的相关系数就越大^[13].

$$P(x^{(0)},\hat{x}^{(0)}) = \frac{Cov(X,\hat{X})}{\sqrt{Var[X]Var[\hat{X}]}}. \tag{5}$$

式中: $P(x^{(0)},\hat{x}^{(0)})$ 为预测负荷序列与真值负荷序列的相关系数; $Cov(X,\hat{X})$ 为预测负荷序列与真值负荷序列上的协方差; $Var[X]$ 为预测负荷序列的方差; $Var[\hat{x}]$ 为真值负荷序列的方差.

5)Theil 不等系数. Theil 不等系数用于计算两个序列的一致性指标,当 Theil 不等系数越接近 0 时,二序列越一致,当越接近 1 时,二序列越不一致^[14].

$$T(x^{(0)}, \hat{x}^{(0)}) = \frac{\sqrt{\frac{1}{n} \sum_{t=1}^n (x^{(0)}(t) - \hat{x}^{(0)}(t))^2}}{\sqrt{\frac{1}{n} \sum_{t=1}^n (x^{(0)}(t))^2 + \frac{1}{n} \sum_{t=1}^n (\hat{x}^{(0)}(t))^2}} \quad (6)$$

式中: $T(x^{(0)}, \hat{x}^{(0)})$ 为预测负荷序列与真值负荷序列的 Theil 不等系数; $x^{(0)}(t)$ 为第 t 日的预测燃气负荷, m^3 ; $\hat{x}^{(0)}(t)$ 为第 t 日的真值燃气负荷, m^3 。

本文采取的模型预测效果综合评价值的计算方法为各单一准则结果数据标准化处理后, 利用加权平均的方式计算得出, 权重系数由操作预测软件的人员在用户界面端口进行修改, 初始权重为等权重。

2.2 蚁群算法计算组合预测权重

蚁群算法是一种模拟蚂蚁觅食行为的仿生式迭代搜索算法^[15], 蚁群算法在解决组合优化问题和最优解问题时的主要优势在于它具有较强的自组织性和自适应性、易于分布式运行、搜索高效、算法优化融合特性较好等^[16]。

本文应用蚁群算法确定组合预测各单项模型权重系数的具体实现步骤如下:

1) 初始化数据. 权重系数的取值范围为 $[0, 1]$, 将权重取值区间平均分为 100 份, 即为 $[0, 0.01], [0.01, 0.02], \dots, [0.99, 1.00]$. 设共有 N 种单项方法参与到组合预测中, 则每个取值区间可以表示为 $[x_{ij}, x_{i(j+1)}], i = 1, 2, \dots, N, j = 0, 1, \dots, 99$, 同时初始化各取值区间的信息素浓度均为 1。

2) 初始参数设置. 算法需要确定的各项参数的定义和取值见表 7。

表 7 蚁群算法参数设置表

Tab. 7 Parameters in ant colony algorithm

参数定义	取值影响	取值
α (信息启发因子)	α 越大, 则新一只蚂蚁选择之前蚂蚁经过的权重区间的概率越大; α 越小, 则蚁群搜索范围减小, 易陷入局部最优解。	1.0
β (期望启发因子)	β 越大, 蚁群选择局部较优的区间概率越大, 迭代收敛速度也越快, 但是也容易陷入局部的相对最优解。	0.5
m (蚁群数量)	m 越大, 算法得到的最优解越精确, 但随着算法接近最优解的收敛, 信息正反馈的作用降低, 增加了大量的重复计算工作。	400
ρ (信息挥发因子)	ρ 越小, 各区间残留信息素浓度越大, 蚁群搜索范围越大, 算法收敛越慢。 ρ 越大, 残留信息素浓度越小, 算法易陷入局部最优解。	0.5

3) 蚁群搜索. 蚁群依次搜索各单项预测方法的各权重区间. 单只蚂蚁的具体搜索过程的算法实现主要分为以下 4 步。

①通过 Python 中随机数生成器确定蚂蚁的出生位置, 并设置禁忌表, 保证蚂蚁搜索各单项预测方法有且仅有一次。

②通过下式计算转移概率, 即计算蚂蚁选择该单项预测方法各权重取值区间的概率:

$$p_j^k = \begin{cases} \frac{[\tau_j]^\alpha \cdot [n_j]^\beta}{\sum_{s \in J_k} [\tau_s]^\alpha \cdot [n_s]^\beta}, & j \in J_k; \\ 0, & \text{其他.} \end{cases} \quad (7)$$

式中: $p_j^k(t)$ 为蚂蚁 k 转移到权重区间 j 的概率, $i, j = 0, 1 \cdots, 99$; τ_j 为权重区间 j 的信息素量; n_j 为权重区间 j 的信息期望启发参数; J_k 为蚂蚁 k 下一步可选择的权重区间的集合; α 为信息启发因子; β 为期望启发因子。

③通过轮盘赌的方式确定蚂蚁选择的权重区间, 本文轮盘赌算法的具体实现是先将该预测方法的各区间取值概率 $p_j(j = 0, 1, \dots, 99)$ 求和得到总概率 P , 再在概率 $[0, P]$ 之间生成一个随机概率, 再依次循环遍历所有区间, 做轮次相减, 首先出现概率小于 0 的区间即为蚂蚁最终选择该方法的权重区间。

④蚂蚁依次遍历其他参与到组合预测的单项预测方法的权重区间的搜索。

4) 信息素更新. 每次迭代需要对各权重区间的信息素浓度更新, 本文选取基于全局信息更新的蚁周模型对各权重区间信息素计算并更新, 即蚁群所有蚂蚁完成一次搜索后, 计算各蚂蚁选择的权重区间下的组合预测模型综合评价指标, 获取最优蚂蚁的路径, 对其选择的权重区间利用式 (8) 进行信息素更新, 对其他权重区间利用式 (9) 进行信息素更新。

$$\tau_i = (1 - \rho) \tau_{i-1} + L, \quad (8)$$

$$\tau_i = (1 - \rho) \tau_{i-1}. \quad (9)$$

式中: τ_i 为第 i 次迭代权重区间信息素增量; ρ 为信息挥发因子; L 为最优蚂蚁选择的权重区间下的组合预测模型综合指标值。

3 预测实例分析

选取 2015—2016 年的各日负荷数据为训练集, 2017 年的各日负荷数据为测试集, 预测天数为 7 d, 以 2017 年 1 月 10 日为预测起始日, 得到各负荷预测模型的预测结果如图 2 所示, 基于多准则评价指标的模型预测效果评价结果见表 8。

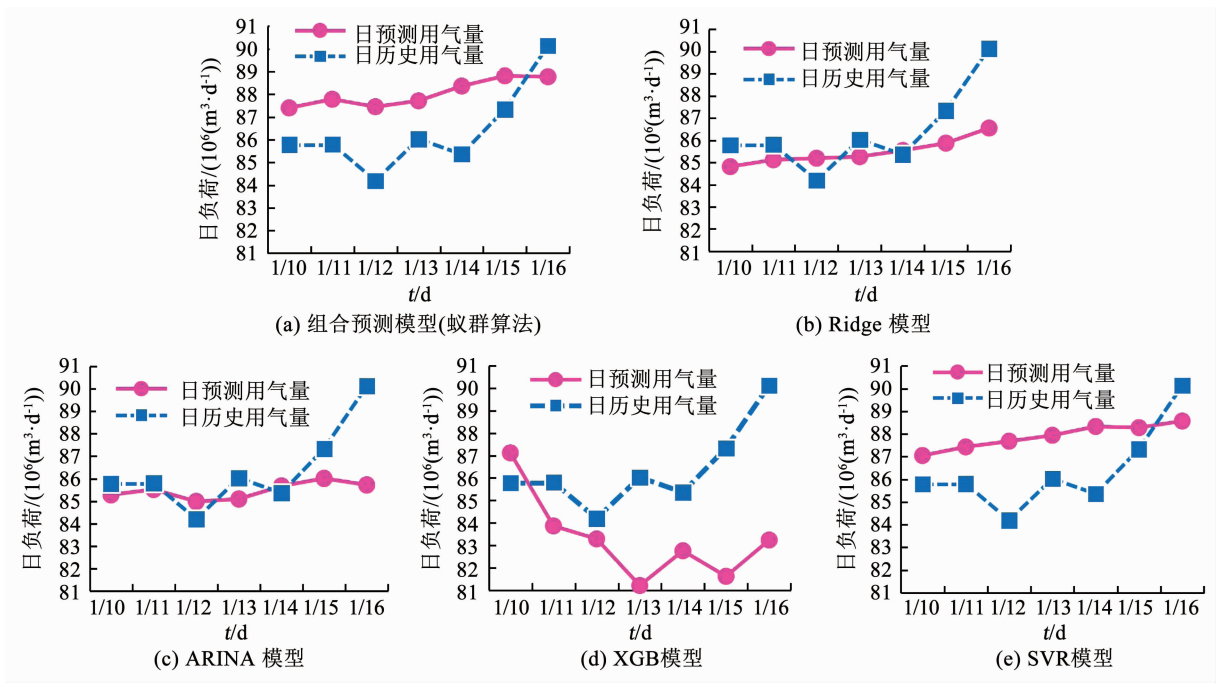


图 2 燃气日用气量预测值与实测值对比

Fig. 2 Comparison between predicted and measured results of daily gas load

表 8 2017 年 1 月 10 日各模型评价指标计算结果

Tab. 8 Calculation results of model evaluation indexes on January 10th, 2017

预测模型	1-MAPE	1-RMSE	ξ_x	γ_i	1-T	综合评价 指标
岭回归	0.985 9	0.993 2	0.706 3	0.857 2	0.990 7	0.906 7
ARIMA	0.986 2	0.992 4	0.797 6	0.605 7	0.989 5	0.874 3
SVR	0.977 1	0.990 5	0.759 3	0.590 5	0.987 7	0.857 5
XGB	0.960 8	0.982 6	0.699 8	-0.151 9	0.976 2	0.693 5
组合预测- 蚁群算法	0.980 5	0.991 7	0.761 2	0.770 1	0.989 3	0.898 6

同时,以 2017 年 1 月份为例,计算得到各模型综合预测效果评价表见表 9. 本文将反向指标,包括平均相对误差 MAPE、均方根误差 RMSE、Theil 不等系数 T 的值,改用它们的等价值表示:1-MAPE,1-RMSE,1- T 的值作为评价指标值,得到的综合评价指标值越高模型效果越好.

表 9 给出的数据表明组合预测-蚁群算法的综合评价指标最高,其次是岭回归和 ARIMA 模型,这一结果表明具有解释性原理的岭回归和 ARIMA 模型具有更好的预测效果. 表 8 数据显示某一日的岭回归预测好于组合预测,但在一个时间段上的综合预测模型好于单一模型,基于权重分配方式的组合预测模型的泛化能力较强,能够提高预测模型容错率和稳定性.

表 9 2017 年 1 月各模型综合预测效果评价表

Tab. 9 Comprehensive prediction evaluation of each model in January, 2017

预测模型	1-MAPE	1-RMSE	ξ_x	γ_i	1-T	综合评价 指标
岭回归	0.962 1	0.984 2	0.623 9	0.767 5	0.985 1	0.864 6
ARIMA	0.953 5	0.980 4	0.751 2	0.581 3	0.975 3	0.848 3
SVR	0.956 7	0.981 3	0.634 8	0.547 3	0.019 7	0.820 0
XGB	0.960 5	0.982 5	0.601 3	0.596 3	0.971 1	0.822 3
组合预测- 蚁群算法	0.967 4	0.986 5	0.751 9	0.720 7	0.982 3	0.881 8

4 结 论

- 1) 单项目日负荷模型为:岭回归分析(Ridge)、差分自回归积分滑动平均模型(ARIMA)、支持向量机回归(SVR)、极端梯度提升树(XGB). 各单项负荷预测模型在不同时间段上的预测效果存在较为显著的差异,其中具有可解释性原理的岭回归和 ARIMA 模型具有较好的预测效果.
- 2) 以 5 个单一评价准则计算的综合评价指标为目标函数的权重系数计算方法所构建的组合预测模型,对各单一评价准则的精度均有改善. 组合预测前的模型筛选可以有效避免冗余模型的干扰,提升组合预测模型的精度.
- 3) 组合预测-蚁群算法的综合评价指标最高,权重分配式组合预测模型的泛化能力较强. 虽然组

合预测模型在具体某一天的预测精度要低于单一模型,但是能够提高预测模型在一个综合时间段上预测的容错率和稳定性.

4) 蚁群算法应用于求解组合预测模型权重系数问题上,具有较强的自组织性和自适应性、易于分布式运行、搜索高效、算法优化融合特性好等优点.

参考文献

[1] 刘金源. 基于组合模型的燃气负荷预测技术研究[D]. 西安: 西安石油大学, 2018
LIU Jinyuan. Research on gas load forecasting technology based on combination model[D]. Xi'an:Xi'an Shiyou University, 2018

[2] ERVURAL B C, BEYCA O F, ZAIM S. Model estimation of ARMA using genetic algorithms: A case study of forecasting natural gas consumption[J]. Procedia-Social and Behavioral Sciences, 2016, 235: 537. DOI:10.1016/j.sbspro.2016.11.066

[3] TASPINAR F, CELEBI N, TUTKUN N. Forecasting of daily natural gas consumption on regional basis in turkey using various computational methods[J]. Energy and Buildings, 2013, 56(1): 23. DOI: 10.1016/j.enbuild.2012.10.023.

[4] 祖广磊, 陆黎明, 徐晓钟. 短期燃气预测的信息熵组合模型研究[J]. 计算机应用与软件, 2013, 30(5): 129
ZU Guanglei, LU Liming, XU Xiaozhong. Study on information entropy combination model for short-term gas load prediction[J]. Computer Applications and Software, 2013, 30(5): 129. DOI:10.3969/j.issn.1000-386x.2013.05.037

[5] 陈庆明. 基于组合方法的短期燃气负荷预测的研究[D]. 上海: 上海师范大学, 2014; 30
CHEN Qingming. Research on short-term gas-load forecasting based on combined method[D]. Shanghai: Shanghai Normal University, 2014; 30. DOI:10.7666/d.Y2526865

[6] 张杨竣, 高文学, 王启, 等. 天然气大气式灶具的 CO 排放特性预测[J]. 天然气工业, 2016, 36(11): 114
ZHANG Yangjun, GAO Wenxue, WANG Qi, et al. Prediction on CO emission performance of atmospheric gas cookers[J]. Natural Gas Industry, 2016, 36(11): 114. DOI: 10.3787/j.issn.1000-0976.2016.11.015

[7] NAIK J, BISOI R, DASH P K. Prediction interval forecasting of wind speed and wind power using modes decomposition based low rank multi-kernel ridge regression[J]. Renewable Energy, 2018, 129: 378. DOI:10.1016/j.renene.2018.05.031

[8] 曾鸣, 李树雷, 王良, 等. 基于 ARMA 模型和 BP 神经网络组合优化算法的风电预测模型[J]. 华东电力, 2013, 41(2): 349
ZENG Ming, LI Shulei, WANG Liang, et al. Wind power prediction model based on the combined optimization algorithm of ARMA model

and BP neural networks[J]. East China Electric Power, 2013, 41(2): 349. DOI: CNKI:SUN:HDDL.0.2013-02-030

[9] 简献忠, 顾祎婷. 一种基于全过程优化支持向量机的短期电力负荷预测方法[J]. 电力科学与工程, 2018, 34(11): 46
JIAN Xianzhong, GU Yiting. Short-term electric load forecasting method based on global optimization support vector machine[J]. Electric Power Science and Engineering, 2018, 34(11): 46. DOI: 10.3969/j.issn.1672-0792.2018.11.007

[10] 许裕栗, 杨晶, 李柠, 等. Xgboost 算法在区域用电预测中的应用[J]. 自动化仪表, 2018, 39(7): 5
XU Yuli, YANG Jing, LI Ning, et al. Application of Xgboost algorithm in the regional electricity consumption forecast[J]. Process Automation Instrumentation, 2018, 39(7): 5. DOI: 10.16086/j.cnki.issn1000-0380.2017120007

[11] 张超. 基于组合模型的我国能源消费预测研究[D]. 长沙: 湖南大学, 2017; 28
ZHANG Chao. Forecast of China's energy consumption based on combination model[D]. Changsha: Hunan University, 2017; 28

[12] 王彬. 基于灰色回归组合模型的铁路客运量预测研究[J]. 重庆理工大学学报(自然科学), 2017, 31(11): 233
WANG Bin. Forecast of railway passenger traffic based on grey linear regression combined model[J]. Journal of Chongqing University of Technology (Natural Science), 2017, 31(11): 233. DOI: 10.3969/j.issn.1674-8425(z).2017.11.035

[13] JADEVICIUS A. The use of combination forecasting approach and its application to regional market analysis[J]. Region, 2014, 1(1): Y6

[14] 孙浩, 杨桂元. 基于 Theil 不等系数的 IGOWLA 算子最优组合预测模型[J]. 西华师范大学学报(自然科学版), 2017, 38(3): 294
SUN Hao, YANG Guiyuan. The combination forecasting model based on Theil coefficient and Induced Geometric Ordered Weighted Logarithm Averaging (IGOWLA) operator[J]. Journal of China West Normal University (Natural Science), 2017, 38(3): 294. DOI: 10.16246/j.issn.1673-5072.2017.03.011

[15] 李智, 谢兆鸿, 秦建华. 蚁群算法在天然气输送管道优化设计中的应用[J]. 天然气工业, 2005, 25(9): 104
LI Zhi, XIE Zhaohong, QIN Jianhua. Application of ant colony algorithms in optimized design of gas transmission pipelines[J]. Natural Gas Industry, 2005, 25(9): 104. DOI: 10.3321/j.issn:1000-0976.2005.09.034

[16] XU mingle, YOU xiaoming, LIU sheng. A novel heuristic communication heterogeneous dual population ant colony optimization algorithm[J]. IEEE Access, 2017, 5: 18510. DOI: 10.1109/ACCESS.2017.2746569

(编辑 张 红)