

基于 Markov 链优化灰色 GM(1,1) 模型 预测城市居民用电量

陈志恒^a, 王宇凡^b, 熊 昕^{*b}

(江汉大学 a. 智能制造学院; b. 人工智能学院, 人工智能研究院, 湖北 武汉 430056)

摘要: 针对数据列波动对 GM(1,1) 模型影响的问题, 通过引入 Markov 概率矩阵来减小波动, 以实现其预测精度的提高。以苏州市 2015 年 1 月—2020 年 10 月之间每月的居民用电量作为模型的训练数据, 以 2020 年 11 月—2021 年 3 月每月居民用电量的数据作为测试数据, 计算模型的预测精度, 并预测 2021 年 4 月的居民用电量。实证结果得出: GM(1,1) 模型预测的平均相对误差为 24.70%, 而通过 Markov 链进行优化后其平均相对误差为 11.62%, 通过 Markov 链进行优化后的 GM(1,1) 模型预测效果要优于传统 GM(1,1) 模型。

关键词: 居民用电量预测; 灰色 GM(1,1) 模型; Markov 链

中图分类号: O29 文献标志码: A 文章编号: 1673-0143(2022)03-0012-09

DOI: 10.16389/j.cnki.cn42-1737/n.2022.03.002

Electricity Consumption Prediction of Urban Residents Based on Grey GM(1,1) Model Optimized by Markov Chain

CHEN Zhiheng^a, WANG Yufan^b, XIONG Xin^{*b}

(a. School of Intelligent Manufacturing; b. School of Artificial Intelligence, Artificial Intelligence Institute, Jianghan University, Wuhan 430056, Hubei, China)

Abstract: In this paper, the Markov probability matrix was introduced to reduce the influence of the high fluctuation data columns on the GM(1,1) model to improve its prediction accuracy. Moreover, the monthly data of the residential electricity consumption in Suzhou from January 2015 to October 2020 were selected as the model's training data. The monthly data of the residential electricity consumption from November 2020 to March 2021 were used as the test data to measure the model's prediction accuracy and forecast the

收稿日期: 2021-09-27

基金项目: 国家自然科学基金资助项目(11901245); 江汉大学校级科研项目(2021yb057)

作者简介: 陈志恒(2000—), 男, 研究方向: 灰色模型、人工智能与智能信息处理。

*通信作者: 熊 昕(1987—), 女, 讲师, 博士, 研究方向: 灰色系统、系统控制与工程。E-mail: scysxiongxin@hotmail.com

residential electricity consumption in April 2021. The empirical results showed that the average relative error of the GM(1,1) model was 24.70%, while the average relative error of the GM(1,1) model optimized by the Markov chain was 11.62%. The prediction performance superiority of the GM(1,1) model optimized by the Markov chain over the traditional GM(1,1) model is demonstrated in this paper.

Key words: prediction on the residential electricity consumption; grey GM(1,1) model; Markov chain

0 引言

自新中国成立以来,我国已逐步建成较为完整的能源供给体系。随着信息化时代不断进步,保障经济高速发展的同时兼顾环境保护工作开展势在必行。党的十九大报告中习总书记提出了“金山银山,不如绿水青山”的环保理念^[1],将我国环境保护、资源节约提上工作日程。随着信息化时代不断发展,我国居民的生活质量日益提升,电子设备等高科技产品使用率也逐年增加,用电水平随之提高。因而,电力资源短缺的问题也随之而来。电能作为一种重要的二次能源,由于其不能大规模储存,故电力系统的发电量需随用户用电量的变化而变化。准确预测未来城市对电力的需求量,对城市供电网的完善和规划有重要的指导意义。

近些年,针对预测问题的方法研究,国内外学者提出了大量的预测方法^[2-3],如灰色预测模型、ARIMA预测模型^[4]等。其中,邓聚龙^[5]1982年提出了GM(1,1)预测模型,该模型属于灰色系统理论中的一种预测模型,能够有效处理一些数据较少的序列,在短期预测上有着极为重要的作用。为解决灰色模型在中长期预测中预测精度不高的问题,孙建梅等^[6]于2019年提出利用滑动平均法对原始数据序列进行处理及优化,增强了灰色预测对波动数据序列的抗干扰性,但该方法在处理一些波动较大的数据序列时存在精度不足等问题。为解决由于数据序列波动而导致预测精度不高的问题,陈友华等^[7]从信息互补的角度提出组合预测模型。何忠华^[8]在2017年将智能优化算法与灰色模型进行结合,并与传统GM(1,1)模型进行了效果比较,验证了模型具有较好的精度和适用性,但此算法针对数据序列波动较大时,效果不佳且有着较大误差。

本文拟对中国苏州城市居民用电量的历史数据进行分析与利用,建立城市居民用电量的预测模型,并对未来居民用电量进行预测。为避免传统灰色模型在中长期预测上的劣势,针对数据序列波动对GM(1,1)模型影响的问题,本文通过引入Markov概率矩阵来减小波动,建立GM(1,1)-Markov预测模型以提高模型预测精度,并基于该预测模型对中国苏州市的用电量进行预测,最后与传统灰色模型进行比较。

1 GM(1,1)模型构建

设序列 $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$, 其中 $x^{(0)}(k) \geq 0 (k = 1, 2, \dots, n)$; 对 $x^{(0)}$ 做一次累加, 得到生成数列为 $X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n))$, 其中 $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i) (k = 1, 2, \dots, n)$, 称

$$x^{(0)}(k) + ax^{(1)}(k) = b \quad (1)$$

为GM(1,1)模型的原始形式。

$X^{(1)}$ 的紧邻均值生成序列 $Z^{(1)} = \{z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n)\}$, 其中

$$z^{(1)}(k) = \frac{1}{2} (x^{(1)}(k) + x^{(1)}(k-1)), k = 2, 3, \dots, n \quad (2)$$

称为一阶线性微分方程。

灰色微分方程为 $x^0(k) + ax^1(k) = b$, 由式(2)可得

$$\frac{dx^1}{dt} + ax^1 = b \quad (3)$$

为灰色微分方程的白化方程。

由式(3)的解可得

$$\hat{x}^1(k+1) = (x^0(1) - \frac{b}{a})e^{-ak} + \frac{b}{a}, k = 0, 1, \dots, n-1, \quad (4)$$

用式(4)来描述 $x^{(1)}(k)$ 的近似值。式(3)中的系数 a 和常数项 b 由式(4)表示, 即

$$\Phi = [a, b]^T = (B^T B)^{-1} B^T Y, \quad (5)$$

$$\text{式中, } Y = \begin{bmatrix} x^0(2) \\ x^0(3) \\ \vdots \\ x^0(n) \end{bmatrix}, B = \begin{bmatrix} -Z^1(2) & 1 \\ -Z^1(3) & 1 \\ \vdots & \vdots \\ -Z^1(n) & 1 \end{bmatrix}。$$

再对式(4)和式(5)做累减还原, 得到原始数列 $x^{(0)}(n)$ 的灰色预测模型为

$$\hat{x}^0(k) = \hat{x}^1(k+1) - \hat{x}^1(k) = (1 - e^a)(x^0(1) - \frac{b}{a})e^{-a(k-1)}, k = 1, 2, \dots, n_0. \quad (6)$$

2 Markov 链模型

2.1 Markov 过程和 Markov 链

Markov 过程是一种随机过程, 其中时间参数是离散时间的集合, 时间所对应的 $x^{(0)}(k)$ 也是离散的, 且 $k+1$ 时刻的概率分布只与 k 时刻的状态有关, 与 k 时刻之前的状态无关, 即为 Markov 链的无后效性^[8], 其表示式为

$$p\{(x(k+1) = j | x(k) = i, x(k-1) = i_{k-1}, \dots, x(0) = i_0)\} = p\{x(k+1) = j | x(k) = i\}. \quad (7)$$

基于 Markov 链模型^[9-11], 将数据划分成 m 个状态 E_1, E_2, \dots, E_m 。通常分类根据实际数据和 GM(1,1) 模型预测数据进行划分。本文通过 GM(1,1) 模型得出预测数据与实际数据之间差值的表达式为

$$N^0(k) = \hat{x}^0(k) - x^0(k). \quad (8)$$

将 $N^0(k)$ 中大于 0 的数据令为一个新的数组 $N^1(k)$, 反之为 $N^2(k)$, 则

$$A = \frac{\sum_{i=1}^k N^1(i)}{k}, B = \frac{\sum_{i=1}^k N^2(i)}{k}, C = \max(N^1(k)), D = \max(N^2(k)). \quad (9)$$

则可将状态指标值的变化区间表示为

$$\begin{aligned}
 E_1 &= [\hat{x}^0(k) - D, \hat{x}^0(k) - B], \\
 E_2 &= [\hat{x}^0(k) - B, \hat{x}^0(k)], \\
 E_3 &= [\hat{x}^0(k), \hat{x}^0(k) + A], \\
 E_4 &= [\hat{x}^0(k) + A, \hat{x}^0(k) + C]。
 \end{aligned} \tag{10}$$

2.2 概率转移矩阵

在Markov链中,系统由一种状态转移到某一种具体状态的可能性用概率来表示,即 $p_{nm} = p(x_n = j | x_m = i)$; $n, m \in N^*$ 称其为转移概率,即

$$p_{nm} = \frac{f_{nm}}{\sum_{i=1}^m f_{ni}}, \tag{11}$$

则称其为一步状态转移概率,即 i 步转移概率为 $p^{(i)}_{nm}$,则有

$$p^{(i)}_{nm} = \begin{bmatrix} p^{(i)}_{11} & p^{(i)}_{12} & \cdots & p^{(i)}_{1m} \\ p^{(i)}_{21} & p^{(i)}_{22} & \cdots & p^{(i)}_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p^{(i)}_{n1} & p^{(i)}_{n2} & \cdots & p^{(i)}_{nm} \end{bmatrix}。 \tag{12}$$

2.3 计算各阶自相关系数

用各阶自相关系数来表示各时间段的用电量之间的相关关系及其强弱,阶数是为了预测系统变量而选取的状态转移的步数,即各阶的自相关系数为

$$r_k = \frac{\sum_{i=1}^{n-k} (\hat{x}^0(i) - \overline{\hat{x}^0(i)}) (\hat{x}^0(i+k) - \overline{\hat{x}^0(i+k)})}{\sum_{i=1}^n (\hat{x}^0(i) - \overline{\hat{x}^0(i)})^2}, \tag{13}$$

式中, n 表示数据列长度; k 表示阶数; $\hat{x}^0(i)$ 表示灰色模型的预测值; $\overline{\hat{x}^0(i)}$ 表示数据预测列的均值。

对各阶自相关系数进行归一化,即

$$\omega_k = \frac{|r_k|}{\sum_{i=1}^k |r_i|}, \tag{14}$$

并将 ω_k 作为各种步长的Markov链的权重。

2.4 计算预测变量处于各状态的概率

将同一状态的各预测概率加权和作为其预测值处于各状态的概率(p_i),即

$$p_i = \sum_{k=1}^m \omega_k p_{nm}^{(k)}. \tag{15}$$

则预测值所对应的状态为 $\max(p_i, i \in m)$,根据区间得出其预测值为

$$y^0(k) = \frac{\otimes_- + \otimes_+}{2} = \hat{x}^0(k) + \frac{1}{2}(u + d), \tag{16}$$

式中, \otimes 表示其区间范围, $u, d \in [A, B, C, D]$ 。

3 预测苏州城市居民用电量

苏州地区属于亚热带季风海洋性气候,四季分明,作为非省会城市其GDP在国内居前10名,对用电量需求很大。本文选取苏州市2015年1月至2020年10月期间每月居民用电量数据^[12]作为模型的训练数据(图1),并以2020年11月至2021年3月中每个月居民用电量($\times 10^3$ GWh)数据作为测试数据测算模型的预测精度。模型达到精度要求后,再用本模型对2021年4月的居民用电量进行预测。

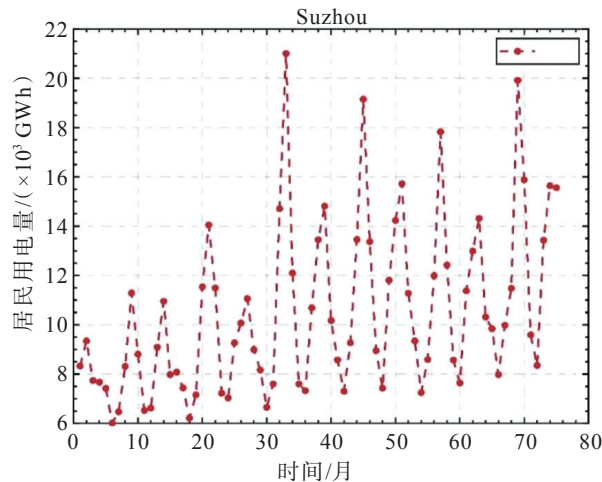


图1 苏州市2015年1月–2021年3月期间每月的居民用电量数据图

Fig. 1 The data graph of residential electricity consumption by months in Suzhou from January 2015 to March 2021

使用图1所示的数据作为灰色模型的原始序列,进行累加生成新的序列 $X^{(1)}$,通过表达式(2)~(6),可计算出GM(1,1)模型的时间响应式,即

$$\hat{x}^0(k) = \hat{x}^1(k+1) - \hat{x}^1(k) = (x^0(1) + 1235.6)e^{-0.0066(k-1)} + 1235.6, k = 1, 2, \dots, n. \quad (17)$$

由式(17)可得2020年11月–2021年3月居民用电量的预测结果,与实际结果的对比见表1。

表1 GM(1,1)模型测试数据的预测结果与实际结果的对比

Tab. 1 The comparisons between predicted results and actual results of the test data of the GM(1,1) model

年/月	实际值/($\times 10^3$ GWh)	预测值/($\times 10^3$ GWh)	残差	相对误差/%
2020/11	9.59	12.915 508 1	3.325 508 141	34.68
2020/12	8.35	13.000 704 4	4.650 704 438	55.70
2021/01	13.43	13.086 462 7	-0.343 537 273	2.56
2021/02	15.64	13.172 786 7	-2.467 213 285	15.78
2021/03	15.56	13.259 680 1	-2.300 319 867	14.78
平均相对误差				24.70

通过表1中测试数据的实际值和预测值,使用式(8)和式(9)可计算得出

$$A = 2.7034, B = 2.1368, C = 10.9398, D = 4.6507,$$

则状态指标值的变化区间表示为

$$E_1 = [\hat{x}^0(k) - 4.6507, \hat{x}^0(k) - 2.1368],$$

$$E_2 = [\hat{x}^0(k) - 2.1368, \hat{x}^0(k)],$$

$$E_3 = [\hat{x}^0(k), \hat{x}^0(k) + 2.7038],$$

$$E_4 = [\hat{x}^0(k) + 2.7034, \hat{x}^0(k) + 10.9398]。$$

其状态区间划分图如图2所示。

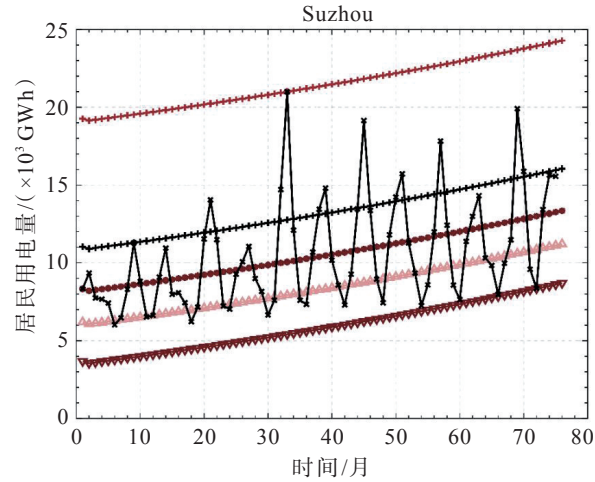


图2 Markov链系统状态区间划分图

Fig. 2 The interval partition of the system states on the Markov chain

根据居民用电量对应的状态序列,由式(11)得到Markov链中状态转移概率矩阵的 $P_{44}^{(1)}, P_{44}^{(2)}, P_{44}^{(3)}, P_{44}^{(4)}$ 可表示为

$$P_{44}^{(1)} = \begin{bmatrix} 0.4211 & 0.3158 & 0.2105 & 0.0526 \\ 0.3330 & 0.3810 & 0.2381 & 0.0476 \\ 0.2105 & 0.2632 & 0.2632 & 0.2632 \\ 0 & 0.2000 & 0.4000 & 0.4000 \end{bmatrix},$$

$$P_{44}^{(2)} = \begin{bmatrix} 0.0526 & 0.2105 & 0.4211 & 0.3518 \\ 0.3810 & 0.2381 & 0.2381 & 0.1429 \\ 0.3684 & 0.3684 & 0.1579 & 0.1053 \\ 0.3000 & 0.5000 & 0.1000 & 0 \end{bmatrix},$$

$$P_{44}^{(3)} = \begin{bmatrix} 0 & 0.1053 & 0.4211 & 0.4737 \\ 0.3000 & 0.3000 & 0.3000 & 0.1000 \\ 0.3158 & 0.5263 & 0.1579 & 0 \\ 0.7000 & 0.2000 & 0 & 0 \end{bmatrix},$$

$$P_{44}^{(4)} = \begin{bmatrix} 0.1111 & 0.3333 & 0.3889 & 0.1667 \\ 0.2000 & 0.3000 & 0.2000 & 0.3000 \\ 0.4737 & 0.2105 & 0.2105 & 0.1053 \\ 0.4000 & 0.3000 & 0.2000 & 0 \end{bmatrix}。$$

由式(13)和式(14)可得测试数据的各阶自相关系数及其权重,分别见表2和表3。

表2 Markov链模型计算测试数据的各阶自相关系数

Tab. 2 The order autocorrelation coefficients of the test data based on the Markov chain model

自相关系数	2020/11	2020/12	2021/01	2021/02	2021/03
r_{k1}	0.998 7	0.998 8	0.998 8	0.998 8	0.998 9
r_{k2}	0.995 0	0.995 1	0.995 3	0.995 4	0.995 5
r_{k3}	0.988 5	0.988 8	0.989 1	0.989 4	0.989 7
r_{k4}	0.978 9	0.979 5	0.980 1	0.980 7	0.981 2

表3 Markov链模型计算测试数据的权重

Tab. 3 The weights of the test data based on the Markov chain model

权重	2020/11	2020/12	2021/01	2021/02	2021/03
w_{11}	0.252 1	0.252 1	0.252 0	0.252 2	0.251 9
w_{22}	0.252 1	0.251 2	0.251 1	0.251 1	0.251 1
w_{33}	0.249 5	0.249 6	0.249 6	0.249 6	0.249 6
w_{44}	0.247 1	0.247 2	0.247 3	0.247 4	0.247 4

通过式(15),并根据权重和状态转移矩阵,可得出拟测试月份居民用电量的转移概率,如表4所示。

表4 基于Markov链模型测试月份居民用电量计算测试数据的转移概率

Tab. 4 The transfer probability of the test data of the residential electricity consumption monthly based on the Markov chain model

转移概率	2020/11	2020/12	2021/01	2021/02	2021/03
p_{11}	0.293 1	0.475 4	0.434 9	0.172 4	0.082 2
p_{22}	0.281 5	0.263 5	0.334 3	0.254 4	0.188 1
p_{33}	0.283 1	0.162 3	0.169 4	0.292 0	0.388 6
p_{44}	0.117 5	0.049 1	0.049 1	0.233 9	0.305 6

最后,通过GM(1,1)-Markov模型拟合出各测试月份的居民用电量,计算得到相对误差,如表5所示。

如图3所示,GM(1,1)模型在面对数据波动较大时,无法适应实际数据波动。然而,由真实数据曲线可知2021年1月相比2020年12月的实际数据有较大波动,2021年1月的预测数据与实际数据之间有偏差。而2021年2月的预测数据通过GM(1,1)-Markov模型调节后,其预测数据随着实际数据的波动而波动,因此更加证实其实验结果满足了预期要求。

表5 基于GM(1,1)-Markov链模型测试数据的预测结果与实际结果对比

Tab. 5 The comparisons of the test data between prediction results and actual results based on the GM(1,1)-Markov chain model

年/月	实际值/($\times 10^3$ GWh)	预测值/($\times 10^3$ GWh)	残差	相对误差/%
2020/11	9.59	9.564 367 834	-0.025 632 166	0.27
2020/12	8.35	9.606 965 982	1.256 965 982	15.05
2021/01	13.43	9.649 845 127	-3.780 154 873	28.15
2021/02	15.64	14.438 444 350	-1.201 555 649	7.68
2021/03	15.56	14.481 891 060	-1.078 108 940	6.93
平均相对误差				11.62

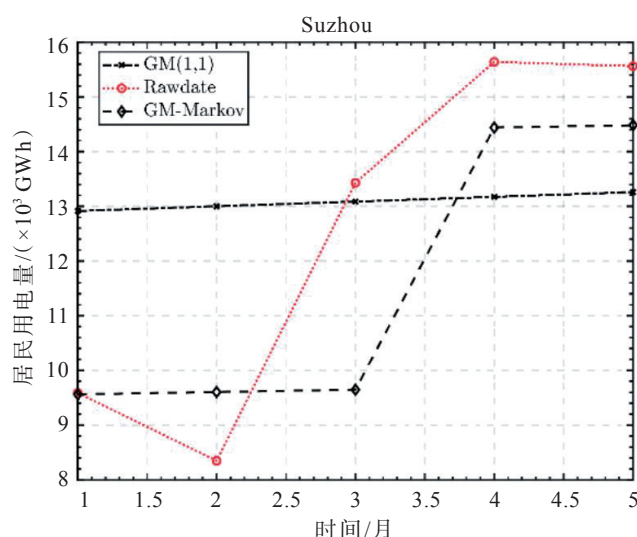


图3 GM(1,1)-Markov模型和GM(1,1)模型测试数据的结果对比

Fig. 3 The comparison of the test data results between the GM(1,1)-Markov model and the GM(1,1) model

4 结论

本文通过建立GM(1,1)模型和GM(1,1)-Markov模型,以苏州市2015年1月至2020年10月期间每月居民用电量数据作为模型的训练数据,取2020年11月至2021年3月期间每月居民用电量的数据作为测试数据,从两个模型的拟合对比发现,GM(1,1)模型预测的平均相对误差为24.70%,而通过Markov链进行优化后其平均相对误差为11.62%。结果表明Markov链进行优化后的GM(1,1)模型预测效果优于传统的GM(1,1)模型。

GM(1,1)-Markov模型预测2021年4月的苏州市居民用电量为: $14.611\ 379 \times 10^3$ GWh。苏州市统计局发布2021年4月实际居民用电量为 10.43×10^3 GWh,GM(1,1)-Markov模型的相对误差为16.36%,与测试数据的平均相对误差相比偏高。实际值与预测值之间有差距,可能是由于苏州市在2021年3月份的社会消费品零售额远高于2021年4月份的社会消费品零售额,导致电子设备等高科技产品使用率增加,用电量则随之提高。但由于本文模型没有考虑到社会

消费品零售额对用电量的影响,后续将引入多种因素约束模型,以达到更好的预测效果,为相关部门的供电调控提供有力依据,减少不必要的能源损耗。

参考文献 (References)

- [1] 高姗. 中国经济发展中的生态环境保护:“金山银山不如绿水青山”的辩证思考[J]. 中国国际财经(中英文), 2018(9):274—276.
- [2] 刘英翰. 电力系统短期电力负荷预测模型研究[D]. 天津:天津理工大学, 2020.
- [3] 张萌. 基于深度学习的电力系统短期电压稳定评估[D]. 吉林:东北电力大学, 2021.
- [4] 于军琪, 聂己开, 赵安军, 等. 基于特征挖掘的 ARIMA—GRU 短期电力负荷预测[J/OL]. 电力系统及其自动化学报, 2022, 34(3):91—99.
- [5] 邓聚龙. 灰色控制系统[M]. 2版. 武汉:华中理工大学出版社, 1985.
- [6] 孙建梅, 钱秀婷, 王永晴. 基于改进灰色模型的中长期电力负荷预测[J]. 电工技术, 2019(19):28—31.
- [7] 陈华友, 朱家明, 丁珍妮. 组合预测模型与方法研究综述[J]. 大学数学, 2017, 33(4):1—10.
- [8] 何忠华. 基于蚁狮优化 GM(1,1)模型的北京市电力需求预测[J]. 现代工业经济和信息化, 2017, 7(22):5—9.
- [9] GOVINDAN K, RAMALINGAM S, BROUMI S. Traffic volume prediction using intuitionistic fuzzy Grey—Markov model [J]. Neural Computing and Applications, 2021, 33(19):1—16.
- [10] HUANG S S, QIAN W Y. The model of Grey—Markov forecasting based on volatility series and its application [C]//2011 IEEE International Conference on Grey Systems and Intelligent Services, WOSC International Congress on Cybernetics and Systems. 2011:298—302.
- [11] JIA Z Q, ZHOU Z F, ZHANG H J, et al. Forecast of coal consumption in Gansu Province based on Grey—Markov chain model [J]. Energy, 2020, 199:117444.
- [12] 苏州市统计局. SUZHOU. GOV. CN [EB/OL]. (2020—11 ~ 2021—3) [2021—04—01]. http://tjj.suzhou.gov.cn/sztjj/tjyx/nav_list_2.shtml.

(责任编辑:胡燕梅)