基于 GPR 模型的气象因素对经济高质量发展的预测——以重庆市为例

李　勇[1],陈栏灵[1],李禹锋[2]
1. 成都信息工程大学 统计学院,成都 610225
2. 广西师范大学 数学与统计学院,广西 桂林 541004

摘　要:目的 针对经济社会与气象变化之间的联系越来越密切的现象,以及气象数据、经济高质量发展数据的复杂特征和传统模型的预测精度不足问题,提出从气象和经济高质量发展关联的视角出发,以统计学方法进行气象因素对经济高质量发展的预测。**方法** 鉴于高斯过程回归模型对于高度非线性回归问题有很强的适应性,同时还能自适应获取最优超参数,并给出具有概率意义的预测结果,故将高斯过程回归模型引入气象对经济高质量发展的预测,采用 7 种不同核函数,并分别训练出最优超参数,通过均方误差比较择出预测效果最好的模型核函数及相应参数。**结果** 对重庆市气象与经济高质量发展历史观测数据构建高斯过程回归(GPR)模型,进行 GPR 建模,并进行预测误差分析,得到的结果表明:选用参数为 8.091 的常值核与缩放参数为 9.454 5 的 RBF 核组合而成的混合核作为最佳核函数的 GPR 模型,相较于 K 邻近回归模型、支持向量回归模型,误差更低,GPR 模型预测点的 y 值绝对误差最大为 0.548,最小为 0.094,较为准确;模型真实值与预测值对比显示拟合效果较为良好。**结论** GPR 模型运用于气象因素对经济高质量发展的预测分析具有优良性,并针对气象与经济高质量发展指数的关系特征,提出了加强气象预报、提高利用效率和精准化预测的有效建议。

关键词:气象因子;经济高质量发展;高斯过程回归(GPR)
中图分类号:F222　　**文献标识码:**A　　doi:10.16055/j.issn.1672-058X.2024.0005.014

Prediction of Meteorological Factors on High-quality Economic Development Based on GPR Model: Taking Chongqing Municipality as an Example

LI Yong[1], CHEN Lanling [1], LI Yufeng[2]
1. School of Statistics, Chengdu University of Information Engineering, Chengdu 610225, China
2. School of Mathematics and Statistics, Guangxi Normal University, Guangxi Guilin 541004, China

Abstract: Objective Aiming at the phenomenon of the increasingly close connection between economy and society and meteorological changes, the complex characteristics of meteorological data and high-quality economic development data, and the problem of insufficient prediction accuracy of the traditional model, this study proposed to carry out the prediction of meteorological factors on the high-quality development of the economy from the perspective of the correlation between meteorology and high-quality development of the economy by statistical methods. **Methods** Gaussian process regression

收稿日期:2023-03-21　**修回日期:**2023-04-26　**文章编号:**1672-058X(2024)05-0110-09
基金项目:国家社科基金重大项目(21&ZD153);重庆社科规划项目(2020ZDTJ08,2020QNJY59);重庆市教育科学规划课题(2020-GX-294);重庆市教改重大项目(201022);四川省高教教改项目(JG2021-393).
作者简介:李勇(1970—),男,重庆人,教授,从事智能大数据统计与气象统计研究.
通讯作者:李禹锋(1995—),男,重庆人,硕士,从事工业大数据与深度学习研究. Email:1284659096@qq.com.
引用格式:李勇,陈栏灵,李禹锋. 基于 GPR 模型的气象因素对经济高质量发展的预测——以重庆市为例[J]. 重庆工商大学学报(自然科学版),2024,41(5):110—118.
LI Yong, CHEN Lanling, LI Yufeng. Prediction of meteorological factors on high-quality economic development based on GPR model: Taking Chongqing Municipality as an example[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2024, 41(5): 110—118.

model has strong adaptability to highly nonlinear regression problems, and it can also adaptively obtain the optimal hyperparameters and give probabilistic prediction results. Therefore, the Gaussian process regression model was introduced into the prediction of meteorological factors on high-quality economic development. Seven different kernel functions were used, the optimal hyperparameters were trained, and the best model kernel function and corresponding parameters were selected by comparing the mean square errors. **Results** A Gaussian process regression (GPR) model was constructed based on the historical observation data of meteorology and high-quality economic development in Chongqing. The results obtained from the analysis of prediction errors showed that compared with *K*-proximity regression model and support vector regression model, the GPR model, which used the hybrid kernel formed by the combination of constant kernel with a parameter of 8.091 and RBF kernel with scaling parameter of 9.454 5 as the best kernel function, had a lower error. The absolute error of the *y*-value predicted by the GPR model was 0.548 at the maximum and 0.094 at the minimum, which was more accurate. The comparison between the real values of the model and the predicted values showed a relatively good fitting effect. **Conclusion** The GPR model is excellent for the forecast analysis of meteorological factors on high-quality economic development. For the characteristics of the relationship between meteorology and high-quality economic development index, effective suggestions, including strengthening the meteorological forecast, improving the utilization efficiency, and making accurate predictions, are put forward.

Keywords: meteorological factor; high-quality economic development; Gaussian process regression (GPR)

1 引　言

近年来,经济社会与气象变化之间的联系越来越密切。气象主要包括温度、空气湿度、风向、风速、降水等各种天气现象[1]。这些气象因素的变化对经济社会发展产生了显著的影响。2022 年发布的《气象服务经济高质量发展纲要(2022—2035 年)》明确提出要提高气象服务经济高质量发展水平,实施气象为农服务提质增效、海洋强国气象保障、交通强国气象保障、“气象+”赋能以及气象助力区域协调发展行动。同年 10 月,党的二十大报告对气象工作发挥基础性、先导性作用更是提出了新要求。气象事业在推动经济高质量发展过程中占据重要一环。分析历年气象的变化特征,可为重庆地区协调发展、绿色生态可持续、社会成果共享等目标的实现及时提供参考,助力经济高质量发展,推动现代化进程。

高质量发展是中国共产党第十九次全国代表大会首次提出的新表述,表明中国经济由高速增长阶段转向高质量发展阶段。在高质量发展测度方面,众多学者主要是多维度地从平衡充分发展视角[2-3]以及五大发展理念[4-5]进行指标体系构建,也有学者从其他角度建立评价体系[6-7]。评价方法多采用熵权法、投入产出法以及空间统计方法等。相关实证大多研究与生态环境保护、金融发展等的关系[8-10]。

而在气象与经济发展关系方面,申丹娜(2018)等[11]认为气象科技创新的总体水平和创新能力是气象强国的关键,气象科技的发展水平和国家经济发展紧密相关,并通过灰色关联法对气象科技竞争力与经济发展的相关性进行了分析,得出相关性较高的结论;丁锋(2021)等[12]从计量经济学的角度,采用柯布-道格拉斯(Cobb-Douglas,C-D)生产函数模型,分析了不同气象变量对青岛地区滨海旅游的经济产出弹性影响,发现 21 世纪以来,青岛地区降水和雷暴天气呈减少的趋势,高温天气则呈增多的趋势,均有利于滨海旅游产值的增加;在气象服务效益评估方面,邓玲(2021)、顾小丽(2022)等[13-14]基于逆推法建立气象服务效益评估模型,均得出服务效益显著的结论;田心如(2021)等[15]运用改进的逆推算法,结合层次分析法分析得出由于台风气象服务,取得综合效益 14.48 亿元,有效避免了 40.9%的经济损失。

综上,现有研究文献对经济高质量发展评价体系较为成熟,评价方法多样,气象与经济发展的相关性证明较为完善。存在的不足:一是以往大多从经济效益角度研究气象与经济关系,缺少气象因素对经济高质量发展预测的探讨;二是预测模型应用方面,前人大多通过计量模型等分析气象对经济发展的影响,精度可以更优化。

基于此,从气象和经济高质量发展关联的视角出发,以统计学方法进行气象因素对经济高质量发展的预测,较之以往研究气象与经济关系大多从单一因素分析,且仅限于局部区域研究,具有视角方面的创新。其次,创新了预测模型应用:前人大多通过计量模型、

资料分析气象对经济发展的影响,而研究气象对经济高质量发展影响因素的极少。通过系统梳理高斯过程理论,丰富了国内对此理论的研究,且鉴于高斯过程回归模型对于高度非线性回归问题有很强的适应性,同时还能自适应地获取最优超参数,并给出具有概率意义的预测结果,故采用 GPR 模型进行预测。通过预测可以更好地应用气象服务促进经济高质量发展,从而进一步推进重庆市现代化发展的进程。

2 高斯过程回归模型

2.1 高斯过程回归原理

高斯过程回归(GPR)是面向自回归问题的高斯过程方法,下面将从权重空间视角进行原理解释。

假设由 n 个训练样本组成的样本集为 $D=\{(\boldsymbol{x}_i,y_i)\mid i=1,2,\cdots,n\}=\{\boldsymbol{X},\boldsymbol{y}\}$。其中,$\boldsymbol{x}_i$ 为第 i 个样本的输入向量;y_i 为第 i 个样本的输出标量;$\boldsymbol{X}$ 为所有训练样本输入向量组成的矩阵;$\boldsymbol{y}$ 为所有样本输出标量组成的向量,样本集 D 的随机扰动服从高斯分布,即 $\varepsilon\sim N(0,\sigma_n^2)$。

第一步,建立线性回归模型:

$$\boldsymbol{y}=f(\boldsymbol{x})+\boldsymbol{\varepsilon}=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\omega}+\boldsymbol{\varepsilon}$$

此处权重 $\boldsymbol{\omega}$ 的先验分布服从高斯分布,即 $\boldsymbol{\omega}\sim N(0,\sum_p)$。则根据贝叶斯公式,权重 $\boldsymbol{\omega}$ 的后验分布为

$$p(\boldsymbol{\omega}\mid\boldsymbol{y},\boldsymbol{X})=\frac{p(\boldsymbol{\omega})p(\boldsymbol{y}\mid\boldsymbol{X},\boldsymbol{\omega})}{p(\boldsymbol{y}\mid\boldsymbol{X})}$$

其中,分母部分 $p(\boldsymbol{y}\mid\boldsymbol{X})$ 的边缘似然为

$$p(\boldsymbol{y}\mid\boldsymbol{X})=\int p(\boldsymbol{y}\mid\boldsymbol{X},\boldsymbol{\omega})p(\boldsymbol{\omega})\mathrm{d}\boldsymbol{\omega}$$

而 $p(\boldsymbol{y}\mid\boldsymbol{X})$ 与 $\boldsymbol{\omega}$ 无关,故有

$$p(\boldsymbol{\omega}\mid\boldsymbol{X},\boldsymbol{y})\propto p(\boldsymbol{y}\mid\boldsymbol{X},\boldsymbol{\omega})p(\boldsymbol{\omega})$$

其中似然 $p(\boldsymbol{y}\mid\boldsymbol{X},\boldsymbol{\omega})$ 可按如下计算

$$p(\boldsymbol{y}\mid\boldsymbol{X},\boldsymbol{\omega})=\prod_{i=1}^{n}p(y_i\mid\boldsymbol{x}_i,\boldsymbol{\omega})=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma_n}\exp\left(-\frac{(y_i-\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\omega})^2}{2\sigma_n^2}\right)=\frac{1}{(2\pi\sigma_n^2)^{n/2}}\exp\left(-\frac{1}{2\sigma_n^2}\|\boldsymbol{y}-\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\omega}\|^2\right)=N(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\omega},\sigma_n^2\boldsymbol{I})$$

所以可推导出 $\boldsymbol{\omega}$ 的后验分布:

$$p(\boldsymbol{\omega}\mid\boldsymbol{X},\boldsymbol{y})\propto\exp\left(-\frac{1}{2\sigma_n^2}\|\boldsymbol{y}-\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\omega}\|^2\right)\exp\left(-\frac{1}{2}\boldsymbol{\omega}^{\mathrm{T}}\sum_p^{-1}\boldsymbol{\omega}\right)$$

也即 $\boldsymbol{\omega}$ 的后验分布服从高斯分布:

$$p(\boldsymbol{\omega}\mid\boldsymbol{X},\boldsymbol{y})\sim N\left(\frac{1}{\sigma_n^2}\left(\frac{1}{\sigma_n^2}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}+\sum_p^{-1}\right)^{-1}\boldsymbol{X}\boldsymbol{y},\left(\frac{1}{\sigma_n^2}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}+\sum_p^{-1}\right)^{-1}\right)$$

这时,若以后验概率 $p(\boldsymbol{\omega}\mid\boldsymbol{X},\boldsymbol{y})$ 为权重,就可求出与任意新输入样本 $\boldsymbol{x}^*$ 对应的函数值 f^* 的后验分布:

$$p(f^*\mid\boldsymbol{x}^*,\boldsymbol{X},\boldsymbol{y})=\int p(f^*\mid\boldsymbol{x}^*,\boldsymbol{\omega})p(\boldsymbol{\omega}\mid\boldsymbol{X},\boldsymbol{y})\mathrm{d}\boldsymbol{\omega}$$

$$p(f^*\mid\boldsymbol{x}^*,\boldsymbol{X},\boldsymbol{y})\sim N\begin{pmatrix}\frac{1}{\sigma_n^2}\boldsymbol{x}^{*\mathrm{T}}\left(\frac{1}{\sigma_n^2}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}+\sum_p^{-1}\right)^{-1}\boldsymbol{X}\boldsymbol{y},\\ \boldsymbol{x}^{*\mathrm{T}}\left(\frac{1}{\sigma_n^2}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}+\sum_p^{-1}\right)^{-1}\boldsymbol{x}^*\end{pmatrix}$$

以上为线性回归问题的 GPR 模型,对于非线性回归问题,可用映射函数组 $\boldsymbol{\Phi}(\boldsymbol{x})=(\Phi_1(\boldsymbol{x}),\Phi_2(\boldsymbol{x}),\cdots,\Phi_N(\boldsymbol{x}))$ 将 d 维输入向量 $\boldsymbol{x}$ 映射到 N 维特征空间中,得到的基函数线性回归方程为

$$\boldsymbol{f}(\boldsymbol{x})=\boldsymbol{\Phi}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{\omega}$$

其分布为

$$(f^*\mid\boldsymbol{x}^*,\boldsymbol{X},\boldsymbol{y})\sim N(\boldsymbol{\Phi}(\boldsymbol{x}^*)^{\mathrm{T}}\sum_p\boldsymbol{\Phi}(\boldsymbol{X})(\boldsymbol{\Phi}(\boldsymbol{X})^{\mathrm{T}}\sum_p\boldsymbol{\Phi}(\boldsymbol{X})+\sigma_n^2\boldsymbol{I})^{-1}\boldsymbol{y},\boldsymbol{B})$$

其中,$\boldsymbol{B}=\boldsymbol{\Phi}(\boldsymbol{x}^*)^{\mathrm{T}}\sum_p\boldsymbol{\Phi}(\boldsymbol{x}^*)-\boldsymbol{\Phi}(\boldsymbol{x}^*)^{\mathrm{T}}\sum_p\boldsymbol{\Phi}(\boldsymbol{X})\times(\boldsymbol{\Phi}(\boldsymbol{X})^{\mathrm{T}}\sum_p\boldsymbol{\Phi}(\boldsymbol{X})+\sigma_n^2\boldsymbol{I})^{-1}\boldsymbol{\Phi}(\boldsymbol{X})^{\mathrm{T}}\sum_p\boldsymbol{\Phi}(\boldsymbol{x}^*)$。

2.2 核函数方法

假设有

$$k(\boldsymbol{x},\boldsymbol{x}')=\boldsymbol{\Phi}(\boldsymbol{x})^{\mathrm{T}}\sum_p\boldsymbol{\Phi}(\boldsymbol{x}')\tag{1}$$

式中,$\sum_p$ 为 N 阶实对称半正定矩阵。由实对称矩阵性质可知,存在正交矩阵 $\boldsymbol{A}$ 使得 $\sum_p=\boldsymbol{A}\boldsymbol{\Lambda}\boldsymbol{A}^{\mathrm{T}}$。$\boldsymbol{A}$ 是以 $\sum_p$ 的单位特征向量为列构成的正交矩阵,对角矩阵 $\boldsymbol{\Lambda}$ 的元素为 $\sum_p$ 的特征值。进一步地,若令 $\boldsymbol{\Phi}_1(\boldsymbol{x})=\sum_p^{1/2}\boldsymbol{\Phi}(\boldsymbol{x})$,则式(1)可改写为

$$k(\boldsymbol{x},\boldsymbol{x}')=\boldsymbol{\Phi}_1(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{\Phi}_1(\boldsymbol{x}')$$

上式表明,$k(\boldsymbol{x},\boldsymbol{x}')$ 是两个向量的内积,且 $k(\boldsymbol{x},\boldsymbol{x}')$ 满足 Mercer 条件,即 $k(\boldsymbol{x},\boldsymbol{x}')$ 实质为 Mercer 核函数,也称协方差函数。以下为常见的核函数。

线性核函数:

$$k(\boldsymbol{x}_p,\boldsymbol{x}_q)=\boldsymbol{x}_p\cdot\boldsymbol{x}_q$$

P 阶多项式核函数:

$$k(\boldsymbol{x}_p,\boldsymbol{x}_q)=(\boldsymbol{x}_p\cdot\boldsymbol{x}_q)^p$$

径向基核函数:

$$k(\boldsymbol{x}_p,\boldsymbol{x}_q)=\exp\left(-\frac{\|\boldsymbol{x}_p-\boldsymbol{x}_q\|^2}{2\sigma^2}\right)$$

2.3　参数优化

高斯过程回归采用最大边缘似然的策略来获取最优超参数。

步骤 1　建立训练样本集的边缘似然函数:

$$p(\boldsymbol{\omega}\mid\boldsymbol{X},\boldsymbol{y})\sim N\left(\frac{1}{\sigma_n^2}\left(\frac{1}{\sigma_n^2}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}+\sum\nolimits_p^{-1}\right)^{-1}\right)\boldsymbol{X}\boldsymbol{y},\left(\frac{1}{\sigma_n^2}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}+\sum\nolimits_p^{-1}\right)^{-1}\right)$$

式中,$\boldsymbol{\omega}$ 表示超参数。

步骤 2　建立对数边缘似然函数:

$$\log p(\boldsymbol{\omega}\mid\boldsymbol{X},\boldsymbol{y})=-\frac{1}{2}\left(\frac{1}{\sigma_n^2}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}+\sum_p^{-1}\right)\left(\boldsymbol{\omega}-\frac{1}{\sigma_n^2}\left(\frac{1}{\sigma_n^2}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}+\sum\nolimits_p^{-1}\right)^{-1}\boldsymbol{X}\boldsymbol{y}\right)^2+\frac{1}{2}\log\left(\frac{1}{\sigma_n^2}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}+\sum_p^{-1}\right)-\frac{1}{2}\log 2\pi$$

步骤 3　根据无约束最优化原理,以对数边缘似然函数为极大化目标函数,以超参数 $\boldsymbol{\omega}$ 为优化变量,用对数边缘似然函数对超参数求偏导数。

$$\frac{\partial\log p(\boldsymbol{\omega}\mid\boldsymbol{X},\boldsymbol{y})}{\partial\boldsymbol{\omega}}=-\left(\frac{1}{\sigma_n^2}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}+\sum_p^{-1}\right)\left(\boldsymbol{\omega}-\frac{1}{\sigma_n^2}\left(\frac{1}{\sigma_n^2}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}+\sum\nolimits_p^{-1}\right)^{-1}\boldsymbol{X}\boldsymbol{y}\right)$$

最后,根据训练样本实际条件,获得最优超参数 $\boldsymbol{\omega}_{\mathrm{MAP}}$。

3　指标体系构建与测算

3.1　数据来源

本项研究中,2005—2022 年气象数据取自美国国家海洋和大气管理局(NOAA)下设的国家环境信息中心(NCEI)官方网站。经济高质量发展数据来源于国家统计局官方平台公布的 2005—2022 年《城市统计年鉴》以及重庆市统计局官网发布的 2005—2022 年《统计年鉴》。

3.2　指标体系构建

(1) 气象指标选择。根据气象站点数据的可获取性,主要从年均气温(℃)、相对湿度(空气中水汽含量,%)、降水量(mm)、风速(m/s)、气压(100 pa)以及年日照时数(h)的角度研究重庆市不同年份气象的变化情况。将这些因素罗列,得到重庆市 2005—2020 年的年度气象指标值。气象影响因素指标如表 1 所示。

表 1　气象影响因素指标

Table 1　Meteorological influencing factors

指标名	单位
年均气温	℃
年均相对湿度	%
年均降水量	mm
年均风速	m/s
年日照时数	h
年均气压	100 pa

根据指标数值分析,重庆市年均气温在 18.7 ℃左右,常年降水量 1 000~1 500 mm,日照总时数 1 000 h 左右,气压年均约 980 hPa,风速约年均 1.4 m/s,相对湿度在 77%左右。

(2) 经济高质量发展指标体系构建。从引言部分文献梳理成果出发,考虑到重庆地区实际情况,在严格遵循相关性、代表性和可获得性等原则情形下,本文拟从五大新发展理念的角度构建含有 5 个一级指标、10 个二级指标的高质量发展评价指标体系(表 2)。将经济高质量发展水平的指标划分为创新、协调、绿色、开发、共享 5 个一级指标,再对一级指标进行具体细化,形成二级指标。5 个一级指标下的 10 个二级指标参考李红(2021)、程如轩(2022)等[16-17]依据新发展理念以及卢贵慧(2022)等[18]根据重庆市区域特征构建的衡量经济高质量发展的指标体系综合考虑得出。创新指标参考程如轩等的研究,主要从科学技术支出、教育支出占一般公共预算支出比重这样的投入角度反映创新发展的状况;协调指标参考卢贵慧等的研究,主要从第三产业占 GDP 的比重、城镇化率反映协调发展的状况;绿色指标参考李红等的研究,主要从生活垃圾无害化处理率、污水处理厂集中处理率反映绿色发展的状况;开放指标参考李红等的研究,主要从外商直接投资总额、进出口总额占地区 GDP 比重反映开放发展的状况;共享指标参考李红等的研究,由于部分数据难以获取,故用人均社会消费品零售总额替换人均消费支出,两者都能反映消费水平,主要从每万人医院床位数、人均社会消费品零售总额反映共享发展的状况。二级指标均为正向指标,能较好地衡量经济发展状况。

表 2 高质量发展评价指标体系

Table 2 Evaluation index system for high-quality development

一级指标	变量表示	二级指标	单位	指标属性
创新投入	B_1	科学技术支出占一般公共预算支出比重	%	正
	B_2	教育支出占一般公共预算支出比重	%	正
协调	B_3	第三产业占地方 GDP 比重	%	正
	B_4	城镇化率	%	正
绿色生活	B_5	生活垃圾无害化处理率	%	正
	B_6	污水处理厂集中处理率	%	正
对外交易	B_7	外商直接投资总额占地区 GDP 比重	%	正
	B_8	进出口总额占地方 GDP 比重	%	正
社会共享	B_9	每万人医院床位数	张/万人	正
	B_{10}	人均社会消费品零售总额	元/人	正

3.3 指数测算

根据获取的指标数值,首先对数据进行归一化处理,由于均为正向指标,不需逆向指标正向化处理,然后直接通过组间相加、组内相乘的综合指数法合成指数。从而得到重庆市 2005—2021 年经济高质量发展指数如表 3 所示:除 2009 年比比前两年的发展指数低,其余整体呈逐年升高趋势,而从 2018 年至 2021 年,指数虽保持逐年升高的趋势,但增速却是逐年放缓,2021 年经济高质量发展指数是 2018 年的 1.268 倍。

表 3 经济高质量发展指数

Table 3 High-quality economic development index

年份	综合指数	年份	综合指数
2005	0.015	2014	2.158
2006	0.081	2015	2.198
2007	1.069	2016	2.467
2008	1.151	2017	2.512
2009	1.030	2018	2.942
2010	1.082	2019	3.442
2011	1.352	2020	3.696
2012	1.913	2021	3.731
2013	2.057		

4 结果与分析

为验证算法的有效性和适用性,本文将高斯过程回归模型与 K 近邻回归模型、支持向量回归模型横向对比预测结果的误差。

4.1 GPR 模型建立与核函数选择

现将气象因素中的年均气温设置为自变量 X_1,相对湿度为 X_2,年均降水量为 X_3,年均风速为 X_4,年日照时数为 X_5,年均气压为 X_6,则有气象因素 $\boldsymbol{X}=\{X_1,X_2,X_3,X_4,X_5\}$;将经济高质量发展指数设置为因变量 $\boldsymbol{y}$。设 $\boldsymbol{\Phi}$ 为映射函数,构建如下 GPR 模型:

$$\boldsymbol{y}=f(\boldsymbol{X})+\boldsymbol{\varepsilon}=\boldsymbol{\Phi}(\boldsymbol{X})^{\mathrm{T}}\boldsymbol{\omega}+\boldsymbol{\varepsilon}$$

本文建模过程中,选用了 7 种核函数进行比较。分别为

(1) RBF(径向基函数核):

$$K(\boldsymbol{X}_i,\boldsymbol{X}_j)=\exp\left(-\frac{d(\boldsymbol{X}_i,\boldsymbol{X}_j)^2}{2l^2}\right)$$

(2) DotProduct(点积核):

$$K(\boldsymbol{X}_i,\boldsymbol{X}_j)=\sigma_0^2+\boldsymbol{X}_i\cdot\boldsymbol{X}_j\quad i,j=1,2,\cdots,7$$

(3) Exponentiation(幂指核):

$$K_{\exp}(\boldsymbol{X},\boldsymbol{y})=K(\boldsymbol{X},\boldsymbol{y})^p$$

(4) Exp-Sine-Squared(周期核):

$$K(\boldsymbol{X}_i,\boldsymbol{X}_j)=\exp\left(-\frac{2\sin^2(\pi\mathrm{d}(\boldsymbol{X}_i,\boldsymbol{X}_j))/p}{l^2}\right)$$

(5) Matern(调整径向基函数核):

$$K(\boldsymbol{X}_i,\boldsymbol{X}_j)=\frac{1}{\Gamma(v)\cdot 2^{v-1}}\left(\frac{\sqrt{2v}}{l}\mathrm{d}(\boldsymbol{X}_i,\boldsymbol{X}_j)\right)^v$$

$$\boldsymbol{K}_v\ \frac{\sqrt{2v}}{l}\mathrm{d}(\boldsymbol{X}_i,\boldsymbol{X}_j))$$

(6) Sum:

$$K_{\mathrm{sum}}=(\boldsymbol{X},\boldsymbol{y})=K_1(\boldsymbol{X},\boldsymbol{y})+K_2(\boldsymbol{X},\boldsymbol{y})$$

(7) Constant(常值核):

$$K(\boldsymbol{X}_i,\boldsymbol{X}_j)=\mathrm{constant_value}\ \forall\boldsymbol{X}_i,\boldsymbol{X}_j$$

在高斯过程回归(GPR)分析中,首先对 $\boldsymbol{X}$ 与 $\boldsymbol{y}$ 进行归一化处理;然后分别选用不同的核函数在归一化后的样本数据 $\boldsymbol{X}$_norm,$\boldsymbol{y}$_norm 条件下,采用自适应训练率优化器(minimize)优化参数,多次迭代,择出最优超参数,最后对结果进行可视化。预测结果如图 1 所示,核函数选择如表 4 与图 2—图 8 所示,选择 Constant+RBF 核函数的平均标准值(MSE)最小,为 0.382 3。

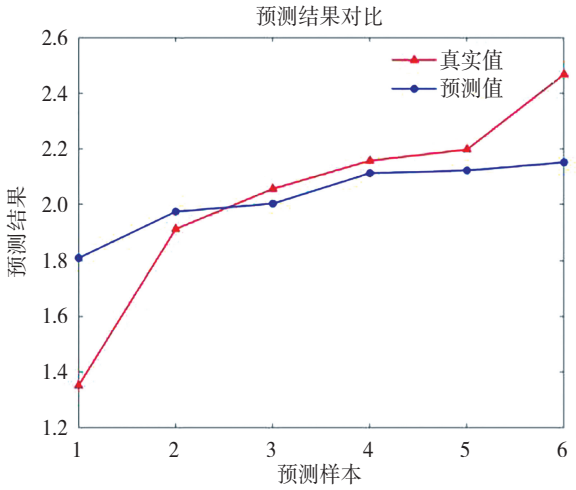

图 1　GPR 的真实值与预测值对比

Fig. 1　Comparison of true and predicted values of GPR

表 4　不同核函数最优参数下的 MSE

Table 4　MSE under the optimal parameters of different kernel functions

核函数	参　数	MSE
RBF	length_scale=9.46	0.462 3
DotProduct	sigma0=3.28	1.625 4
Exponentiation	kernel=RBF(8.52),exponent=2	0.635 8
Exp-Sine-Squared	length_scale=5.19	0.931 9
Matern	length_scale=9.83	0.451 2
Sum	Kernel1=RBF(3.62)+RBF(6.15)	0.433 8
Constant+RBF	constant _ value = 8.091, RBF (9.454 5)	0.382 3

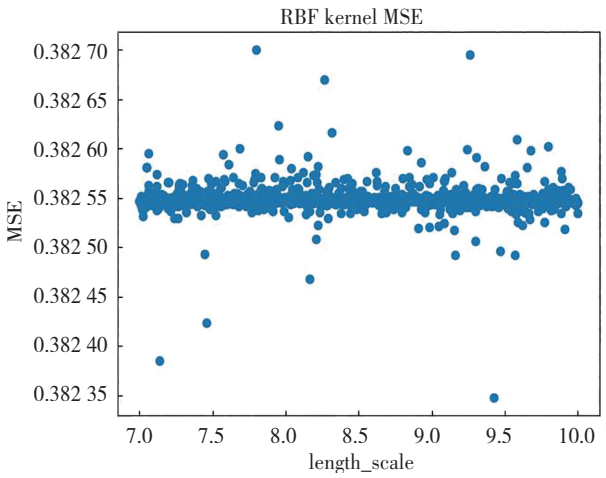

图 2　RBF 核数与 MSE

Fig. 2　RBF kernel parameters and MSE

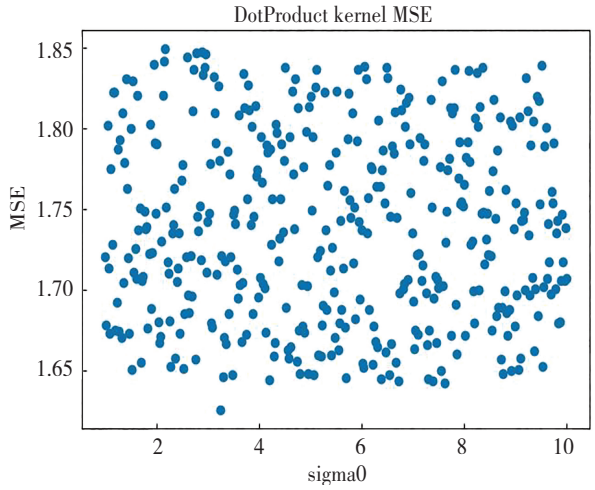

图 3　DotProduct 核函数与 MSE

Fig. 3　DotProduct kernel function and MSE

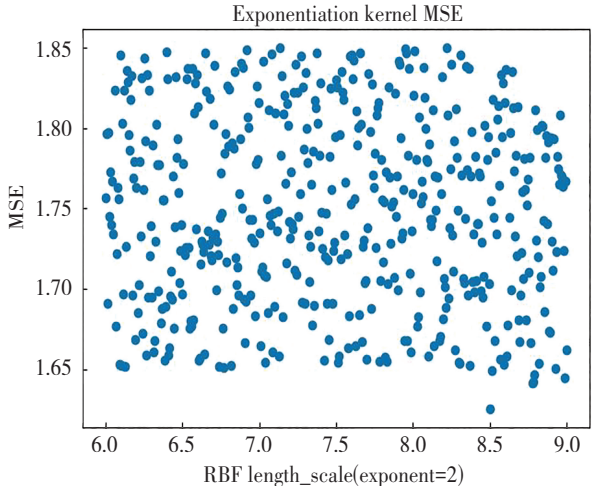

图 4　Exponentiation 核函数与 MSE

Fig. 4　Exponentiation kernel function and MSE

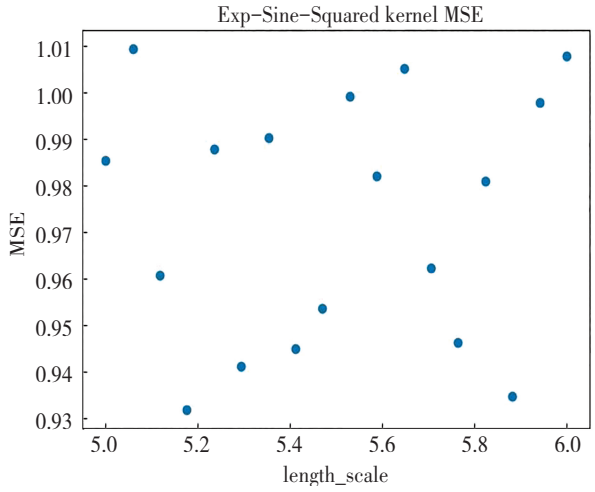

图 5　Exp-Sine-Squared 核函数与 MSE

Fig. 5　Exp-Sine-Squared kernel function and MSE

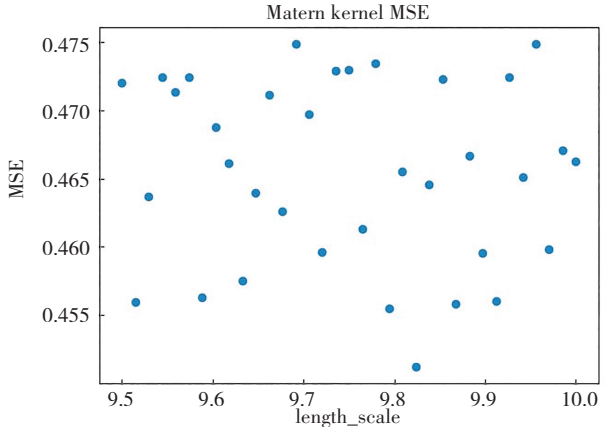

图 6 Matern 核函数与 MSE

Fig. 6 Matern kernel function and MSE

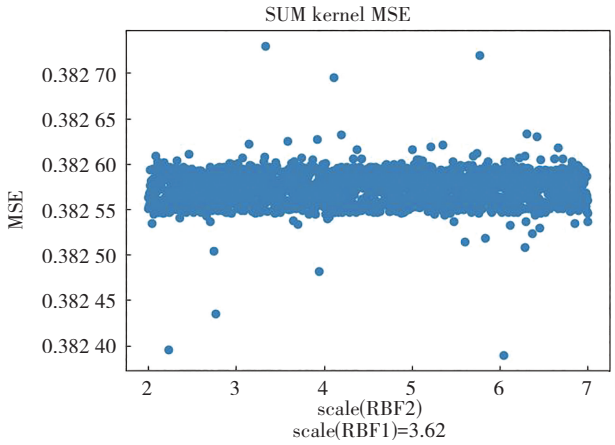

图 7 Sum 核函数与 MSE

Fig. 7 Sum kernel function and MSE

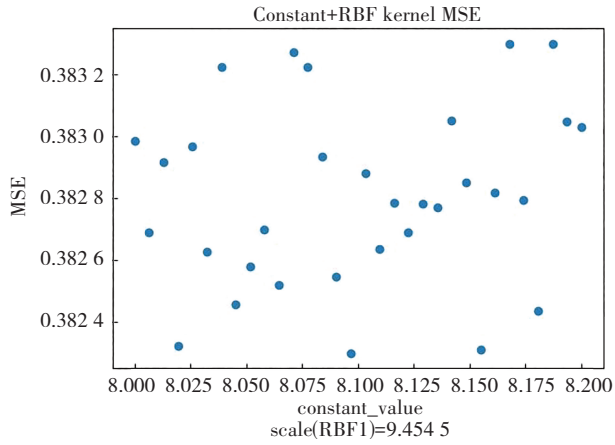

图 8 Constant+RBF 核函数与 MSE

Fig. 8 Constant + RBF kernel function and MSE

4.2 与其他模型的比较

K 近邻回归(KNR)模型通过测量不同特征值之间的距离进行回归,具有精度高、对异常值不敏感、无数据输入假定等优点,根据数据的性质是具有适用性的。KNR 模型在输入待预测样本后,将这个新数据的每个特征与样本集中数据的特征进行比较,然后算法提取样本集中特征最相似数据(最近邻)对应的 $\boldsymbol{y}$。通过试验,当权重参数设置为"距离"会取得相对较好的结果。模型训练后当 $K=4$ 时,效果最好,此时的 MAE 为 0.602,MSE 为 0.669。从图 9 可看出真实值与预测值的对比,拟合效果较图 1 略有不如。

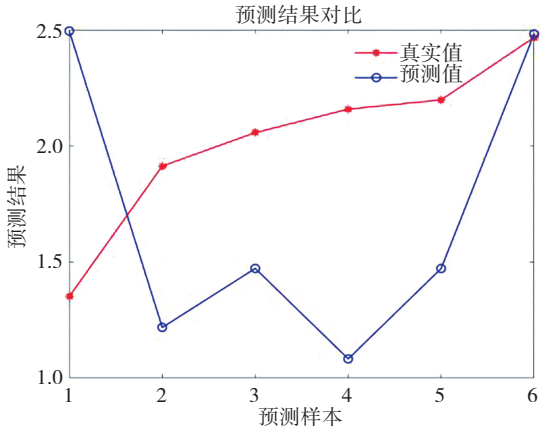

图 9 KNR 的真实值与预测值对比

Fig. 9 Comparison of actual and predicted values for KNR

支持向量回归(SVR)模型的优点在于对异常值具有鲁棒性以及出色的泛化能力等,也较为适用本文数据分析。SVR 模型在考虑模型的复杂性和错误率的情况下,使用对称损失函数进行训练,用 ε 管(ε 表示管子的宽度)的给定余量来接近最佳值。模型训练后准确率为 0.65,此时的 MAE 为 0.647,MSE 为 0.741。从图 10 可看出 SVR 拟合的真实值与预测值的对比,拟合效果相对欠佳。

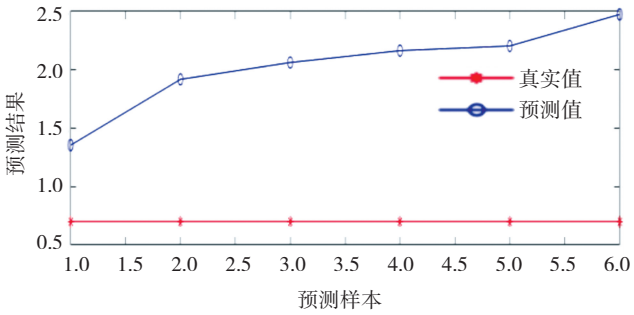

图 10 SVR 的真实值与预测值对比

Fig. 10 Comparison of actual and predicted values for SVR

预测误差比较如表 5 所示。以平均绝对误差、均方误差越小越好为标准。高斯过程回归的 MAE 虽略高于 K 邻近回归,但其 MSE 却明显小于 K 近邻回归。故综合比较,高斯过程回归模型拟合效果优于 K 邻近回归,而 K 邻近回归又优于支持向量回归。

表 5　预测误差比较

Table 5　Comparison of the prediction errors

评价指标	高斯过程回归	K 邻近回归	支持向量回归
MAE(平均绝对误差)	0. 380 2	0. 602	0. 647
MSE(均方误差)	0. 382 3	0. 669	0. 741

4.3　模型结论

选用参数为 8. 091 的常值核与缩放参数为 9. 454 5 的 RBF 核组合而成的混合核作为最佳核函数,将气象各项指标作为自变量,经济高质量发展指数作为因变量建立高斯过程回归预测模型,得到的 RBF 部分为 $K(\boldsymbol{x},\boldsymbol{x}')=\mathrm{e}^{\frac{\|x-x'\|^2}{2\times 2.84^2}}$。

模型在相同尺度范围内的数据下,预测效果比 K 邻近回归、支持向量回归等传统算法效果更好。在气象数据与经济高质量发展数据均为面板数据的情况之下,有着较强的先验信息,且目标的经济高质量发展指数分布未知,非常适合利用高斯过程回归来建模。同时,数据并不存在稀疏问题,且建模选用的气象数据维度不高,可以有效避免高斯过程回归对于稀疏矩阵与高维的不足。

5　结　论

现有文献现研究发现,高斯过程回归的特点,有利于分析气象对经济高质量发展的预测。以重庆市为例的实证分析验证了 GPR 模型相较于 K 邻近回归模型、支持向量回归模型,误差更低,GPR 模型预测点的 $\boldsymbol{y}$ 值绝对误差最大为 0. 548,最小为 0. 094,较为准确。模型真实值与预测值对比显示拟合效果较为良好。故将 GPR 模型运用于气象因素对经济高质量发展的预测分析具有优良性。

同时,通过重庆市气象与经济高质量发展指数分析,认为重庆年均气温、气压适宜,常年降水量和日照都较为充足,这些特征使得重庆农林业、交通、电力、能源等发展较好,进而使得经济高质量发展指数整体呈明显上升态势;但重庆市相对湿度偏高,属于高湿区,易引起物品、设施等霉烂,造成经济损失,夏季易出现极端高温,造成户外工作延期、高耗能等问题,影响经济效益,尤其是 2009 年,年均湿度达到最高,盛夏高温伏旱,加上暴雨等事件影响经济发展,使得当年经济高质量发展指数低于其前两年。

针对此,提出三点建议:一是加强气象预报,充分利用气候条件发展农林业,分时段施工作业,减少耗能;二是提高利用效率,重视设备维护,从而减少经济损耗;三是精准化预测,通过气象辅助预测经济走势,及时调整策略。

本文仍存在着不足,值得进一步探讨。一方面受限于数据的获取,经济高质量发展的指标有待优化;另一方面 GPR 模型的重难点问题以及 GPR 核函数优选问题,这都是需要并且也值得进一步深入探讨的。

参考文献(References):

[1]　李炳昆, 赵超, 李丙辉, 等. 常见气象要素知识简介[J]. 现代农业科技, 2016(2): 253—254.
LI Bing-kun, ZHAO Chao, LI Bing-hui, et al. Introduction of knowledge of common meteorological elements[J]. Modern Agricultural Science and Technology, 2016(2): 253—254.

[2]　刘锴, 周雅慧, 王嵩, 等. 创新驱动下中国区域高质量发展—基于平衡充分发展水平的门槛分析[J]. 技术经济, 2020, 39(12): 1—8.
LIU Kai, ZHOU Ya-hui, WANG Song, et al. Innovation-driven China regional high-quality development: Based on the threshold analysis of balanced and full development level[J]. Technical economy, 2020, 39(12): 1—8.

[3]　鲁邦克, 金大卫. 我国经济发展平衡性充分性的测度与空间动态演进[J]. 统计与决策, 2022, 38(5): 96—100.
LU Bang-ke, JIN Da-wei. The measure and spatial dynamic evolution of the balance and adequacy of China's economic development[J]. Statistics and Decision-making, 2022, 38(5): 96—100.

[4]　史丹, 李鹏. 我国经济高质量发展测度与国际比较[J]. 东南学术, 2019(5): 169—180.
SHI Dan, LI Peng. Comparison of high quality economic development with international[J]. Southeastern Academic, 2019(5): 169—180.

[5]　侯祥鹏. 中国城市群高质量发展测度与比较[J]. 现代经济探讨, 2021(2): 9—18.
HOU Xiang-peng. Measurement and comparison of high-quality development in urban agglomerations in China[J]. Discussion on Modern Economy, 2021(2): 9—18.

[6]　韩永辉, 韦东明. 中国省域高质量发展评价研究[J]. 财贸研究, 2021, 32(1): 26—37.
HAN Yong-hui, WEI Dong-ming. Evaluation of high quality

development in China [J]. Finance and Trade Research, 2021, 32(1): 26—37.

[7] 方晓萍, 廖晓玲, 邓又军. 我国省际高质量发展水平测度[J]. 统计与决策, 2022, 38(5): 111—115.
FANG Xiao-ping, LIAO Xiao-ling, DENG You-jun. High-quality development level measurement in China[J]. Statistics and Decision-making, 2022, 38(5): 111—115.

[8] 王育宝, 陆扬, 王玮华. 经济高质量发展与生态环境保护协调耦合研究新进展[J]. 北京工业大学学报(社会科学版), 2019, 19(5): 84—94.
WANG Yu-bao, LU Yang, WANG Wei-hua. New progress in the coordinated coupling of high-quality economic development and ecological and environmental protection [J]. Journal of Beijing University of Technology (Social Science Edition), 2019, 19(5): 84—94.

[9] 褚钰, 付景保, 陈华君. 区域生态环境与经济耦合高质量发展时空演变分析——以河南省为例[J]. 生态经济, 2022, 38(5): 161—168.

[10] 赵巍, 葛永磊. 金融发展与海洋经济高质量发展的耦合协调分析[J]. 青岛科技大学学报(社会科学版), 2020, 36(4): 1—6.
ZHAO Wei, GE Yong-lei. Coordination analysis of the coupling between financial development and high-quality development of Marine economy [J]. Journal of Qingdao University of Science and Technology (Social Science Edition), 2020, 36(4): 1—6.

[11] 申丹娜, 申丹虹, 李研. 气象科技竞争力及其与经济发展相关性分析[J]. 工程研究-跨学科视野中的工程, 2018, 10(5): 495—502.
SHEN Dan-na, SHEN Dan-hong, LI Yan. Analysis of the competitiveness of meteorological science and Technology and its correlation with economic development [J]. Engineering Research-Engineering in an interdisciplinary sight, 2018, 10(5): 495—502.

[12] 丁锋, 张凯静, 郭文慧. 气象条件对滨海旅游经济产出的弹性影响—以青岛为例[J]. 海洋气象学报, 2021, 41(1): 147—152.
DING Feng, ZHANG Kai-jun, GUO Wen-hui. The elastic influence of meteorological conditions on the economic output of coastal tourism: Take Qingdao as an example[J]. Journal of Marine Meteorology, 2021, 41(1): 147—152.

[13] 邓玲, 李学敏, 谭姗姗, 等. 湖南省暴雨灾害气象防灾减灾服务效益评估研究[J]. 湖北农业科学, 2021, 60(8): 176—179, 189.
DENG Ling, LI Xue-min, TAN Shan-shan, et al. Research on service benefit evaluation of rainstorm disaster prevention and mitigation in Hunan Province [J]. Hubei Agricultural Science, 2021, 60(8): 176—179, 189.

[14] 顾小丽, 钱燕珍, 周伟军, 等. 台风"利奇马"灾害风险及气象服务效益评估[J]. 热带气象学报, 2022, 38(1): 35—42.
GU Xiao-li, QIAN Yan-zhen, ZHOU Wei-jun, et al. Assessment of disaster risk and meteorological service benefit of Typhoon Lekima [J]. Journal of Tropical Meteorology, 2022, 38(1): 35—42.

[15] 田心如, 张雪蓉, 侯俊. 基于改进逆推算法的台风气象服务综合效益评估研究[J]. 气象科学, 2021, 41(6): 835—843.
TIAN Xin-ru, ZHANG Xue-rong, HOU Jun. Study on the comprehensive benefit evaluation of typhoon meteorological service based on the improved inverse calculation method[J]. Meteorological Science, 2021, 41(6): 835—843.

[16] 李红, 曹玲. 长江中游城市群经济高质量发展测度[J]. 统计与决策, 2021, 37(24): 101—105.
LI Hong, CAO Ling. High-quality economic development measure of urban agglomeration in the middle reaches of the Yangtze River[J]. Statistics and Decision-making, 2021, 37(24): 101—105.

[17] 程如轩, 李小胜, 马悦航. 新发展理念下我国经济高质量发展测度[J]. 安庆师范大学学报(社会科学版), 2022, 41(5): 53—63.
CHENG Ru-xuan, LI Xiao-sheng, MA Yue-hang. High-quality development measurement of China' s economy under the new development concept[J]. Journal of Anqing Normal University (Social Science Edition), 2022, 41(5): 53—63.

[18] 卢贵慧, 韦子军. 重庆市经济高质量发展测度评价研究[J]. 现代营销(下旬刊), 2022(6): 98—100.
LU Gui-hui, WEI Zi-jun. Evaluation study on the measurement of high-quality economic development in Chongqing [J]. Marketing Management Review, 2022(6): 98—100.

责任编辑:李翠薇