ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202502_8(2).0034

Forecasting Cross-border E-commerce Sales Based on Gaussian Process Regression

Luotong Tang

Tianjin University of Commerce, Tianjin, China

Abstract

In recent years, cross-border e-commerce is booming. In the cross-border e-commerce model B2C (Business to Customer), the transaction subject is the enterprise and the consumer, due to its characteristics of small batch, multi-batch, facing many customers and order dispersion, the enterprise needs to plan the future experience strategy based on the sales, and the prediction of sales becomes more and more important. In order to meet the enterprise's demand for sales forecasting, this paper establishes a Gaussian process regression model, and selects Squared Exponential Kernel, Rational Quadeatic Kernel and Matern Kernel to construct a combined kernel function for forecasting. Comparing the prediction results with other models, it is found that the Gaussian process regression model using the combined kernel function has better prediction effect than other models, and the prediction results can help cross-border e-commerce enterprises to make decisions.

Keywords

Gaussian process; Time Series; Kernel Function; Forecasting.

1. Introduction

Cross-border e-commerce refers to the transaction subject belonging to different national boundaries or regions, through the Internet platform for transactions and payment and settlement, and through the cross-border logistics to deliver the goods, complete the transaction of a kind of international commercial activities, which is widely recognized by the society as a kind of electronic data exchange and online transactions as the main content of the business model. With the rapid development of economic globalization and e-commerce, cross-border e-commerce is developing rapidly and has become one of the main ways of international trade [1]. The mainstream cross-border e-commerce model of B2C (Business to Customer) is the main body of the transaction between the enterprise and the consumer [2], and its transaction is characterized by small batch, multi-batch, facing many customers and scattered orders. Enterprises need to formulate sales strategies based on the sales volume of goods, such as determining the amount of advertising and the amount of goods to be purchased. Therefore, predicting the sales volume of cross-border e-commerce enterprises can help enterprises formulate correct sales strategies, which can help them develop better.

Sales volume belongs to time series data, and there are many models that can forecast this kind of data, such as the ARMA model, support vector machine, random forest, etc. However, in practice, using these models will encounter some difficulties. However, in practice, the use of these models for forecasting will encounter some difficulties. the ARMA model requires time series data to meet the assumption of smoothness, and the data need to be smoothed [3][4], but the sales volume is affected by a variety of factors, and it is sometimes difficult to be smoothed. And the prediction results of support vector machine [5][6] will change with the change of kernel function. And Random Forest [7] has the problem of overfitting.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202502_8(2).0034

Gaussian Process Regression (GPR) is a Bayesian nonparametric regression method widely used for regression problems in machine learning, especially when the model complexity is high or there are no explicit assumptions about the distribution of the data. Its core idea is to use Gaussian Process (GP) to describe the data generation process in a regression problem, and thus to predict unobserved data [8]. Gaussian Process can effectively model nonlinear relationships in time series data. And Gaussian process regression does not assume a specific model of the data; it flexibly captures the underlying structure of the data through the kernel function. For small amounts of data, Gaussian process can provide more accurate predictions than other methods.

2. Gaussian Process Regression

2.1. Forecast

A Gaussian process is a collection of any finite number of random variables all having a joint Gaussian distribution, the properties of which are completely determined by the mean and covariance functions, i.e:

$$\begin{cases} m(x) = E[f(x)], \\ k(x,x') = E\left[\left(f(x) - m(x)\right)\left(f(x') - m(x')\right)\right], \end{cases} \tag{1}$$

where, $x, x' \in R^d$ are arbitrary random variables, so the Gaussian process can be defined as:

$$f(x) \sim GP\left(m(x), k(x,x')\right). \tag{2}$$

In order to keep the notation concise, the mean function is usually made equal to 0. The Gaussian process regression model is built by taking the noise content into account in the observation y, i.e.

$$y = f(x) + \varepsilon. \tag{3}$$

where, x is the input vector and ε is an independent Gaussian white noise, denoted $\varepsilon \sim N(0, \sigma_n^2)$. Since both ε and $f(x)$ follow a Gaussian distribution, the prior distribution of the observations is:

$$y \sim N(0, K(x,x') + \sigma_n^2 I_n), \tag{4}$$

and the joint prior distribution of observations y and predicted values f_* is

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(X,X) + \sigma_n^2 I_n & K(X,x_*) \\ K(x_*,X) & k(x_*,x_*) \end{bmatrix}\right). \tag{5}$$

where, $K(X,X) = K_n = k_{ij}$ is a symmetric positive definite covariance matrix of order $n \times n$; The matrix element $k_{ij} = k\left(x_i, x_j\right)$ represents the correlation between x_i and x_j; $K(X,x_*) = K(x_*,X)^T$ is the covariance matrix between the test point x_* and the input X of the training set; $k(x_*,x_*)$ is the covariance of the test point x_* itself; I_n is an n-dimensional unit matrix. This gives the posterior distribution of the predicted value f_* as

$$f_* | X, y, x_* \sim N\left(\overline{f}_*, cov(f_*)\right). \tag{6}$$

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202502_8(2).0034

Where

$$\overline{f_*} = K(x_*, X)[K(X,X) + \sigma_n^2 I_n]^{-1} y, \tag{7}$$

$$cov(f_*) = k(x_*, x_*) - K(x_*, X)[K(X,X) + \sigma_n^2 I_n]^{-1} K(X, x_*). \tag{8}$$

Generally, $\overline{f_*}$ is taken as the predicted value of f_* and $cov(f_*)$ is the variance of the predicted value.

2.2. Training

Gaussian process regression models have many kernel functions to choose from, and in order to accomplish the task of prediction, it is often necessary to estimate the hyperparameters of the kernel function using the observations. Let the set of parameters θ be hyperparameters, then the negative log–likelihood function and its partial derivatives with respect to the hyperparameters θ take the following form

$$L(\theta) = \frac{1}{2} y^T C^{-1} y + \frac{1}{2} \log|C| + \frac{n}{2} \log 2\pi, \tag{9}$$

$$\frac{\partial L(\theta)}{\partial \theta_i} = \frac{1}{2} tr\left((\alpha\alpha^T - C^{-1}) \frac{\partial C}{\partial \theta_i} \right), \tag{10}$$

where $C = K_n + \sigma_n^2 I_n$,$\alpha = C^{-1} y$.

The partial derivatives are then minimized using optimization methods such as co-choked gradient method and Newton's method to obtain the optimal solution for the hyperparameters. After obtaining the optimal solution, the predicted values and their variances can be obtained using equations (7) and (8).

3. Covariance Functions

3.1. Covariance function selection

In Gaussian process regression, the kernel function is a key component that determines the relationship between data points. Choosing the appropriate kernel function is crucial in Gaussian process regression, and different kernel functions can lead to very different model fits to the data. With a reasonable kernel function design, Gaussian process regression can not only obtain an accurate prediction of the data, but also provide an estimate of the prediction uncertainty. In this paper, the following three single covariance functions are used [9].

(1). Squared Exponential Kernel(SE)

$$K_{SE}(x_i, x_j) = \sigma_f^2 exp\left[-\frac{1}{2} (x_i - x_j)^T M (x_i - x_j) \right] + \sigma_n^2 \delta^{ij}. \tag{11}$$

(2). Rational Quadeatic Kernel(RQ)

$$K_{RQ}(x_i, x_j) = \sigma_f^2 \left[1 + \frac{(x_i - x_j)^T M (x_i - x_j)}{2\alpha} \right]^{-\alpha}. \tag{12}$$

(3). Matern Kernel

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202502_8(2).0034

$$K_M(x_i, x_j) = \sigma_f^2 \left[1 + \sqrt{3M}(x_i - x_j)exp\left(\sqrt{3M}(x_i - x_j)\right)\right]. \tag{13}$$

where $\theta = \{M, \sigma_f^2, \sigma_n^2\}$ is the parameter vector containing all hyperparameters; $M = diag(l^2)$ is the symmetric matrix of the hyperparameters;l is the correlation measure hyperparameter; σ_f^2 is the signal variance of the kernel function; δ^{ij} is Kronecker function; σ_n is the variance of the noise; α is the shape parameter of the kernel function.

3.2. Covariance function Combinations

Kernel function combination refers to obtaining a new composite kernel function by weighted combination of multiple kernel functions, adding them together, or by some kind of operation. The main purpose of kernel function combination is to utilize the characteristics of different kernel functions to enhance the expressive ability of the model. Therefore, in this paper, we will use Squared Exponential Kernel, Rational Quadeatic Kernel and Matern Kernel to construct the combined kernel function, i.e.

$$K_C(x, x') = \sum_{i}^{n} K_i(x, x'), \tag{14}$$

Consider the four cases of adding two kernel functions and adding three kernel functions respectively, so as to obtain a composite kernel function containing multiple characteristics and improve the prediction accuracy.

4. Application

The data used in this paper is the daily sales of a cross-border e-commerce company from March 1, 2024 to June 28, 2024, with a data volume of 120, taking the first 80 days of data as the training set and the last 40 days of data as the test set.

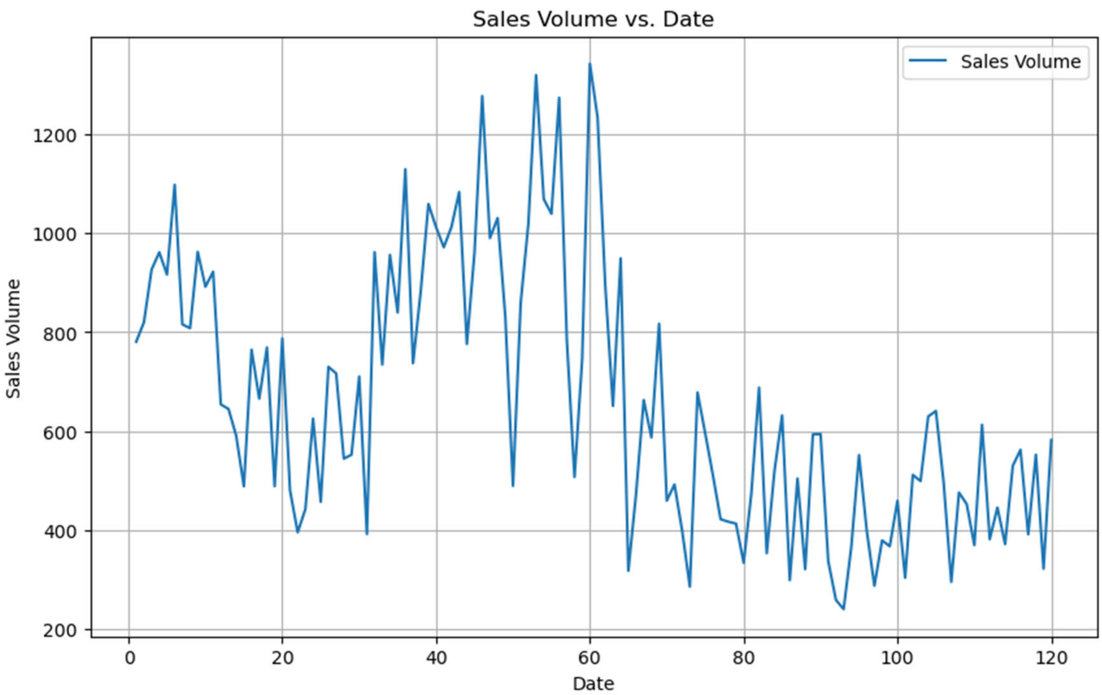

Figure 1. Cross border e-commerce sales time series.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202502_8(2).0034

Choosing the appropriate evaluation metrics can ensure that we comprehensively and accurately assess the performance of the model, in this paper, we choose Root Mean Square Error (RMSE) to test the prediction results. RMSE is an important metric to assess the accuracy of regression models, which is measured by calculating the square root of the prediction error of the model. RMSE provides a simple and effective way to measure and compare the performance of different models, which is calculated as follows:

$$\sigma_{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\bar{y}(i) - y(i)\right)^2}, \tag{15}$$

where N is the number of samples in the test set, $\bar{y}(i)$ is the predicted value, and $y(i)$ is the observed value.

Table 1. Forecasting error analysis.

	GPR						
	SE+RQ+ Matern	SE+RQ	SE+ Matern	RQ+Matern	SVR	DTR	RFR
σ_{RMSE}	125.11	142.44	329.10	144.95	313.36	167.58	147.61

The RMSEs of the different models are given in Table I. It can be seen that for Gaussian process regression, the RMSE of the kernel combination obtained by adding the three kernel functions Squared Exponential Kernel, Rational Quadeatic Kernel and Matern Kernel is the smallest at 125.11. This indicates that its prediction effect is the best. The prediction effect of $K_{\text{SE+RQ+ Matern}}$ is shown in Fig. 2, which shows that the predicted value and the observed value can match well. Therefore the predicted value can be used as a reference for cross-border e-commerce enterprises to make decisions.

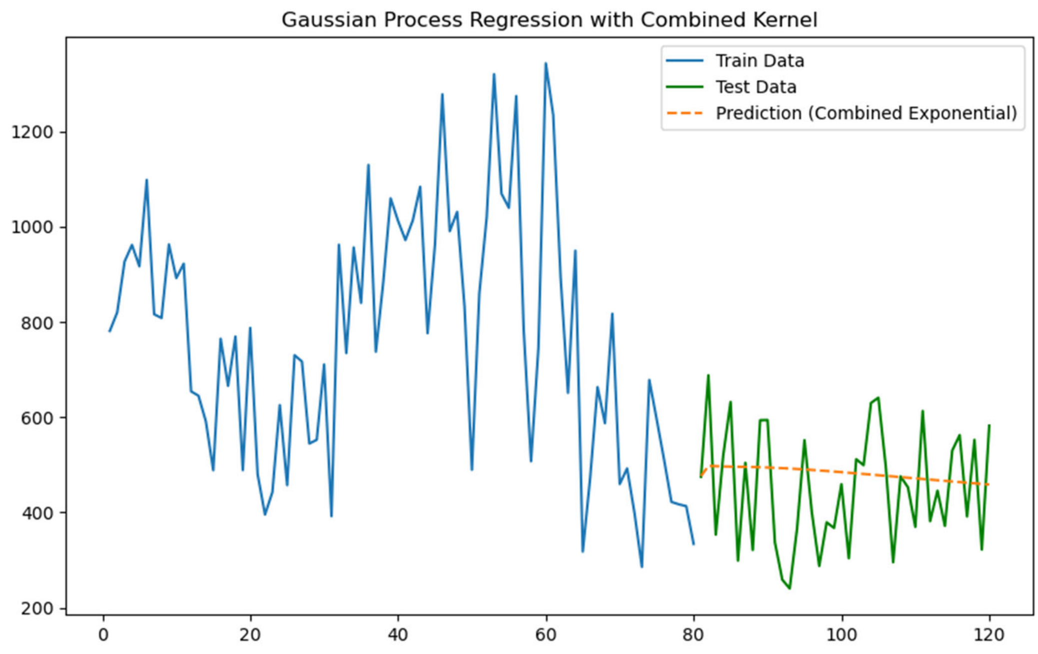

Figure 2. Prediction effect of SE+RQ+ Matern.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202502_8(2).0034

Meanwhile, this paper compares the prediction results of Gaussian process regression model with Support Vector Machine Regression (SVR), Decision Tree Regression (DTR) and Random Forest Regression (RFR). The results show that when the Gaussian process regression model employs a combination of two kernel functions, there is no significant difference in the prediction results with the other models. However, when the Gaussian process uses a combination of three kernel functions, the prediction results are significantly better than the other models.

5. Conclusions

The Gaussian process regression model has excellent performance in the prediction of cross-border e-commerce enterprise sales, which can meet the needs of enterprises as a reference for decision-making. It can provide a positive impact on the development of enterprises. The prediction performance of Gaussian Process Regression based on combined kernel function is stronger than Support Vector Machine Regression, Decision Tree Regression and Random Forest Regression.

The Gaussian process regression model can provide not only the prediction value but also the variance of the prediction value, which can help in the practical application of cross-border e-commerce. In the next step of our research we will explore the form of kernel function combinations of Gaussian process regression models and the effect of different combinations on the prediction performance.

References

[1] E Li-Bin,Huang Yong-Sheng. New way of international trade:the latest research on cross-border e-commerce[J]. Journal of Northeast University of Finance and Economics,2014,(02):22-31.

[2] Zhang Xiaheng. Cross-border e-commerce types and operation mode[J]. China Circulation Economy,2017,31(01): 76-83.DOI:10.14089/j.cnki.cn11-3664/f.2017.01.010.

[3] Zou Baxian, Liu Qiang. Network traffic prediction based on ARMA model[J]. Computer Research and Development, 2002, (12):1645-1652.

[4] ZHANG Mei-Ying, HE Jie. A review of research on time series forecasting models[J]. Practice and Understanding of Mathematics,2011,41(18):189-195.

[5] DING Shifei, QI Bingjuan, TAN Hongyan. A review of support vector machine theory and algorithm research[J]. Journal of University of Electronic Science and Technology,2011,40(01):2-10.

[6] Qi Hengnian. A review of support vector machines and their applications[J]. Computer Engineering, 2004, (10):6-9.

[7] FANG Kuangnan, WU Mibin, ZHU Jianping, et al. A review of random forest methods[J]. Statistics and Information Forum,2011,26(03):32-38.

[8] HE Zhikun, LIU Guangbin, ZHAO Xijing, et al. A review of Gaussian process regression methods[J]. Control and Decision Making,2013,28(08):1121-1129+1137.DOI:10.13195/j.kzyjc.2013.08.018.

[9] Seeger M .Gaussian processes for machine learning[J]. International Journal of Neural System, 2004, 14(2):69-106.