Eng Int Syst (2023) 3: 239–250

Engineering
Intelligent Systems

Application Research and Case Analysis of Elastic-Net Method in Generalized Linear Model

Dengxiang Huang* **and Chunting Lu**

Department of Public Basic Education, Guangxi Financial Vocational College, Nanning 535000, China

Traditional methods apply the generalized linear model in risk assessment, data estimation and other fields, but these approaches have problems such as high time cost and poor accuracy. Therefore, this paper incorporates the Elastic-Net method into the generalized linear model to improve the accuracy of the generalized linear model in terms of calculation and classification. Firstly, the general linear model and the generalized linear model are described, and the Elastic-Net method and its related properties are examined. Then, the Elastic-Net method is incorporated into the generalized linear model, and the classification effect of the model is analyzed by means of an example. Experimental results show that when the Elastic-Net method is introduced into generalized linear models, this can effectively improve the accuracy and efficiency of model classification and the performance of the model.

Keywords: Elastic-Net method; Generalized Linear Model; Classification Results; Data Evaluation.

1. INTRODUCTION

A linear model is an important branch of mathematical statistics with early development, abundant theory and strong application [1]. Over the past few decades, the linear model has not only been very active in theoretical research, but also has been widely used in industry and agriculture, meteorology and geology, economic management, medicine and health, education and psychology. With the continuous progress of modern science and technology, our data collection technology has also been greatly developed [2,3]. Therefore, given the large amount of data being generated, how to extract useful information has become the focus of our attention. Statistical modeling can undoubtedly deal with this problem well. At the beginning of the model, we usually add as many independent variables as possible to the model to reduce the deviation of the model, but given the model can be interpreted, the ease of data collection and calculation cost and other reasons, we need to find the response variables in the modeling process of the most influential independent variable subset, so as to improve the interpretability of the model and the prediction precision [4,5]. Therefore, the selection of variables is a very important and problematic issue in statistical modeling. Relevant scholars have made some progress in this regard.

Teng and Ma [6] proposed a rice-planting risk assessment method based on a generalized linear model. The generalized linear model was used to predict the amount of loss that would be sustained by a rice crop after a natural disaster. Under the assumption that the fluctuations in yield follow the Boolean distribution rule, the linear relationship between the disaster-causing factors and the yield fluctuations is established to assess the vulnerability of the disaster-bearing entity. The weather generator model is used to simulate the occurrence of natural disaster factors in each region to determine the probability of a natural disaster occurring (risk assessment). Finally, the risk assessment result is expressed quantitatively. This method can accurately determine the rice-planting risk. In this instance, the generalized linear model is very effective, although the cost is high. He et al. [7] systematically

*Email of corresponding author: huangdx2022@163.com

introduced the traditional Bayesian small area estimation method, and introduced the generalized linear model into the hierarchical Bayesian method to solve the discrete variable estimation reasoning problem. Then the basic theoretical mechanism and estimation model of classified data are constructed. The results show that the hierarchical Bayesian generalized linear model can accurately estimate the total parameters of the target domain when the sample size is large enough. The new estimation model not only makes full use of prior information and auxiliary information, but also can be used to estimate complex data. Zhu [8] proposed a hybrid attribute data clustering method based on a generalized linear model, and constructed a low-order multivariate generalized linear model to deal with the problem of clustering massive amounts of data. Taking into account the temporal characteristics of data attributes, the attribute time series matrix was obtained. When processing mixed-attribute data based on optimized K-prototypes clustering, the time series matrix of attributes was considered. By considering the distance between the sample and the cluster center, and taking into account the known content of the sample, the optimization method was adopted to calculate the degree of data dissimilarity and the distance between the sample and the cluster set. When the cluster result achieves stability, the operation is terminated and the cluster result is output. This method can achieve accurate clustering of mixed-attribute data, but the clustering process is complicated. Yuan et al. [9] proposed a statistical inference method involving a mixed generalized linear model, based on heterogeneous overall first moment and second moment of the existing conditions. Using the generalized linear mixed models for the overall mean value model, the researchers developed the tectonic extension and the pseudo likelihood function. Then, used the EM algorithm to average and estimate its parameters, divergence and mixture ratio, and through the Monte Carlo simulation to verify the effectiveness of the proposed model parameter estimation method. This method has a good data fitting effect but is costly in terms of time.

To solve the above problems, this paper applies the Elastic-Net method to the generalized linear model, and verifies the application of Elastic-Net method in the generalized linear model by analyzing the effect of data classification and selection by means of examples.

2. METHOD

2.1 Generalized Linear Model

A generalized linear model is a regression model that has greater applicability and practicability based on the generalization of the above model assumptions.

(1) The distribution of response variables can be generalized to an exponential dispersion family, such as normal distribution, Poisson distribution, binomial distribution, negative binomial distribution, gamma distribution, inverse Gaussian distribution [10].

(2) Research object: The main research object of the generalized linear model is still the mean value $R(Y)$ of the response variable.

(3) Connection mode: The connection function used in the generalized linear model can theoretically be arbitrary, not limited to $f(x) = x$. Of course, the connection function that is selected must be adapted to specific research purposes [11]. At the same time, there are standard connection functions corresponding to the distribution mentioned in hypothesis (1), such as normal distribution corresponding to the identity, Poisson distribution corresponding to the natural logarithm function, etc.

It can be seen that the generalized linear model mainly extends the ordinary linear model in two respects: on the one hand, the expectation of the response variable is connected with the linear combination of independent variables by setting a connection function. On the other hand, the distribution of Y is no longer limited to the normal distribution, but is extended to the exponential distribution family [12–15]. These generalizations allow us to study more general problems.

Specifically, the generalized linear model has the following three assumptions:

(1) Random component: that is, the response variable Y obeys the exponential distribution family and the density function is:

$$f(y, \theta, \varphi) = \exp\left\{\frac{y\theta - b(\varphi)}{a(\varphi)} + c(y, \varphi)\right\} \tag{1}$$

where θ is the natural parameter and φ is the graduated parameter [16]. Under certain canonical conditions, it can be proved that: θ passes $b(\cdot)$ associated with $R(Y|X)$, $\mu = R(Y|X) = b'(\theta)$; Given X, the variance of the response variable Y is a function of the mean and φ, i.e.:

$$Var(Y|X) = a(\varphi)b''(\theta) \tag{2}$$

(2) System components: namely, linear combinations of independent variables:

$$\eta = \beta_1 X_1 + \ldots + \beta_p X_p \tag{3}$$

(3) Connection function: The connection function h is a monotone differentiable function, which relates η to $R(Y|X)$:

$$h(\mu) = \eta = \beta_1 X_1 + \ldots + \beta_p X_p \tag{4}$$

The monotonicity of h guarantees that this is a one-to-one mapping, so we can express $R(Y|X)$ in terms of the inverse of the joining function:

$$R(Y|X) = h^{-1}(\beta_1 X_1 + \beta_p X_p) \tag{5}$$

Let's call it $F(X\beta) = h^{-1}(X\beta)$, $V(X\beta) = \sqrt{a(\varphi)b''(\theta)}$, and then we can write the generalized linear model as follows

$$Y = F(X\beta) + V(X\beta)\zeta \tag{6}$$

where $F'(\cdot)$'is bounded, $F(\cdot)$. The second derivative of is the δ-order Hölder continuous function, $V(\cdot)$. Is a non-negative continuous bounded function, ζ obeys an exponential distribution with mean value of 0 and variance of 1 [17].

This paper discusses the generalized linear model as shown in Equation (6).

2.2 Elastic-Net Method

The Elastic-Net method is a variable selection method used to solve strongly-correlated variables based on the Lasso method. It not only inherits several excellent properties of the Lasso method; it also deals effectively with strongly-correlated variables. Therefore, before introducing the Elastic-Net method, Lasso method is explained.

The Lasso method is a variable selection method proposed in 1996, which not only can select the variables, but also obtains parameter estimates.

Consider the following general linear model:

$$Y = X^T \beta + \zeta \tag{7}$$

Among them, $Y = (y_1, y_2, \ldots, y_3)^T$ is the response variable, n is the sample size, $X = (X_1, X_2, \ldots, X_n)$ is the p-dimensional prediction variable, $X_i = (X_{i1}, X_{i2}, \ldots, X_{in})^T$, $i = 1, 2, \ldots, n$, $\beta = (\beta_1, \beta_2, \ldots, \beta_p)^T$ sparsity, that is, many coefficients in $\beta_1, \beta_2, \ldots, \beta_p$ are zero. $\zeta = (\zeta_1, \zeta_2, \ldots, \zeta_n)^T$ is the normal distribution random error [18], that is, $\zeta \sim N(0, \sigma^2 I_n)$. It is assumed that the observed data (y_i, x_{ij}), $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, p$ have been processed by central standardization, that is:

$$\frac{1}{n}\sum_i y_i = 0,\ \frac{1}{n}\sum_i x_{ij} = 0,\ \frac{1}{n}\sum_i x_{ij}^2 = 1 \tag{8}$$

The Lasso method not only inherits the advantages of traditional methods, but also has an effective algorithm, namely the least-angle regression algorithm, which makes it more widely applied and studied in the statistics domain.

2.3 Elastic-Net Method and Related Properties

For ordinary linear models (9), the Elastic-Net method is defined as follows:

$$\beta'(Enet) = \arg\min_{\beta}\{\|Y - X\beta\|^2 + \omega_2\|\beta\|^2 + \omega_1\|\beta\|_1\} \tag{9}$$

Among, $\|\beta\|^2 = \sum_{j=1}^{p}\beta_j^2,\ \|\beta\|_1 = \sum_{j=1}^{p}|\beta_j|$.

For the algorithm of Elastic-Net method, as long as the solution table of Elastic-Net method is transformed to a solution similar to Lasso method, the solution of the Elastic-Net method can be obtained by Lars algorithm, which is expressed as follows by lemma:

Lemma defines a new data set (X^*, Y^*) for given data (X, Y) and fixed parameters (ω_1, ω_2):

$$X^*_{(n+p)\times p} = (1+\omega_2)\begin{pmatrix} X \\ \sqrt{\omega_2} I \end{pmatrix},\ Y^*_{(n+p)} = \begin{pmatrix} Y \\ 0 \end{pmatrix} \tag{10}$$

Let $\gamma = \omega_1/\sqrt{1+\omega_2}$ and $\beta^* = \sqrt{1+\omega_2\beta}$, then the solution of the method is equivalent to the form of the solution of the following method:

$$\beta^{*\prime} = \arg\min_{\beta^*}\left\{\|Y^* - X^*\beta^*\|^2 + \gamma\sum_{j=1|\beta_j^*|}^{p}\right\} \tag{11}$$

Therefore,

$$\beta'(Enet) = \frac{1}{\sqrt{1+\omega_2}}\beta^{*\prime} \tag{12}$$

According to lemma, the algorithm problem of the method is effectively solved, which lays a foundation for further study of its properties and popularization and application. For data with strongly correlated variable groups, a good variable selection method should be able to select all strongly-correlated variable groups inside or outside the model, which is the group effect property of the method [19].

Lemma given data (X, Y), and parameters (ω_1, ω_2), (X, Y) have been processed by central standardization, so that $\beta'(\omega_1, \omega_2)$ represents estimation, assuming $\beta'_{\mathrm{k}}(\omega_1, \omega_2)\beta'_1(\omega_1, \omega_2) > 0$, definition:

$$D_{\omega_1\omega_2}(k, l) = \frac{1}{\|Y\|_1}|\beta'_{\mathrm{k}}(\omega_1, \omega_2) - \beta'_1(\omega_1, \omega_2)| \tag{13}$$

Therefore,

$$D_{\omega_1\omega_2}(k, l) \le \frac{1}{\omega_2}\sqrt{2(1-\rho)} \tag{14}$$

where $\rho = x_k' x_l$ is the sample correlation coefficient. The lemma gives the group effect property of Elastic-Net method. $D_{\omega_1\omega_2}(k, l)$ represents the difference between the coefficient paths of two variables k and l. The lemma shows that if x_k and x_l are strongly correlated, that is, if $\rho = 1$ (or when $\rho = -1$, $-x_l$ is considered), the difference between the coefficient paths of variables k and l is almost zero [20].

The upper bound of inequality is a quantitative description of the group effect of the Elastic-Net method. It can be seen that the Elastic-Net method can effectively deal with the group of strongly-correlated variables and selects all the groups of strongly-correlated variables necessary for the model, while the Lasso method does not have the property of group effect, and therefore cannot deal effectively with data that have strongly-correlated variables.

3. APPLICATION OF GENERALIZED LINEAR MODEL BASED ON ELASTIC-NET METHOD

3.1 Set-up of Simulated Data

The studied region of space is the side length for the $m-1$ unit of distance square, observation point location in the $m \times m$ point on a grid, the horizontal and vertical distance between each point is a unit of length, with u, v, being the observation point of abscissa and ordinate respectively, under which the order of the observation point according

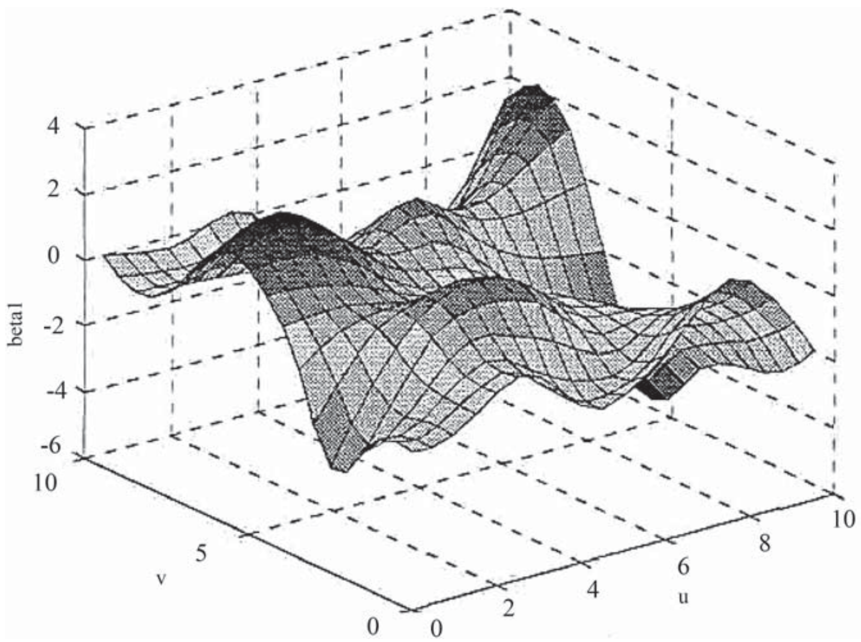

Figure 1 Broadband = 1, Gaussian weight function.

to the left to right, bottom-up order, i.e. for each $u = 0, 1, \ldots, m-1$. Make v take $0, 1, \ldots, m-1$, the coordinate (u_i, v_i) of the position of the ith observation point is $u_i = 0.5 \bmod (i-1, m)$, $v_i = 0.5\left[\frac{i-1}{m}\right]$, $i = 1, 2, \ldots, m^2$, where mod $(\cdot)$ is the mod function and $\cdot$ is the integral function. The generalized geographic weighted regression model is a large class of models, of which the most commonly used are logical regression model and Poisson regression model. In order to facilitate analysis and data simulation, it is advisable to set the geographic weighted regression model for data simulation here as the logical geographic weighted regression model, the expression of which is:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \frac{1}{6}(u_i+v_i) + \ln\left(\frac{1+u_i+v_i}{5}\right)x_i + \zeta_i \quad (15)$$

In the model, the value of independent variable x is a random number generated independently and obeying the uniform distribution on the interval (0,1), and ζ is a random number obeying the standard normal distribution $N(0, 1)$, take $m = 20$. When p_i is greater than or equal to 0.5, y_i is 1, otherwise y_i is 0. After one simulation, 400 sample points (x_i, y_i) can be obtained, and the obtained data can be used to estimate the geographically-weighted logistic regression model.

3.2 Weight Function Selection

To estimate the geographical weighted regression model, we must first determine the type of weight function in order to produce the weight $W_{ij}(u_i, v_i)$ required for parameter estimation. The small space weight function includes two categories; one is the mechanism of fixed weight with constant bandwidth, and the other is the adaptive weight mechanism with variable bandwidth. Three bandwidth invariant weight functions are provided in the spgwr package: the Gaussian weight function, bisquare kernel weight function and cubic kernel weight function. A weight function with variable bandwidth is provided, which is the k nearest neighbor method. Figures 1–6 show the spatial surface diagram of the estimated value of model parameter β_1 obtained by using simulated data samples according to different bandwidth and weight functions.

Figure 1 and Figure 2 both depict Gaussian weight functions, although the bandwidth is different. Therefore, it can be seen that the smaller the bandwidth is, the faster the weight decays with the increase of distance, and the sample points involved in the estimation of each regression point are relatively small, which makes the spatial heterogeneity of the parameter estimation value larger and the spatial surface map uneven. The larger is the bandwidth, the slower the weight will decay with the increase in distance, and the sample points involved in the estimation of each regression point are relatively large, which makes the spatial heterogeneity of parameter estimation less and produces the effect of spatial smoothing, making its spatial surface map smoother. The other four figures show that the double square kernel weight function and the k-nearest neighbor method with variable machine weight mechanism have the same characteristics. Comparing the graphs of different spatial weight functions, when the bandwidth is large or small, the spatial surfaces of their parameter estimates are similar, either smooth or concave convex. From this, we can conclude that the estimation of a geographically-weighted generalized linear model is not very sensitive to the choice of weight function, but it is very sensitive to the broadband of a specific weight function.

3.3 Model Parameter Estimation

The optimal bandwidth of each different weight function is obtained. Next, the weight can be calculated according to the optimal bandwidth to estimate the parameters of the geographically-weighted logistic regression model. Accord-

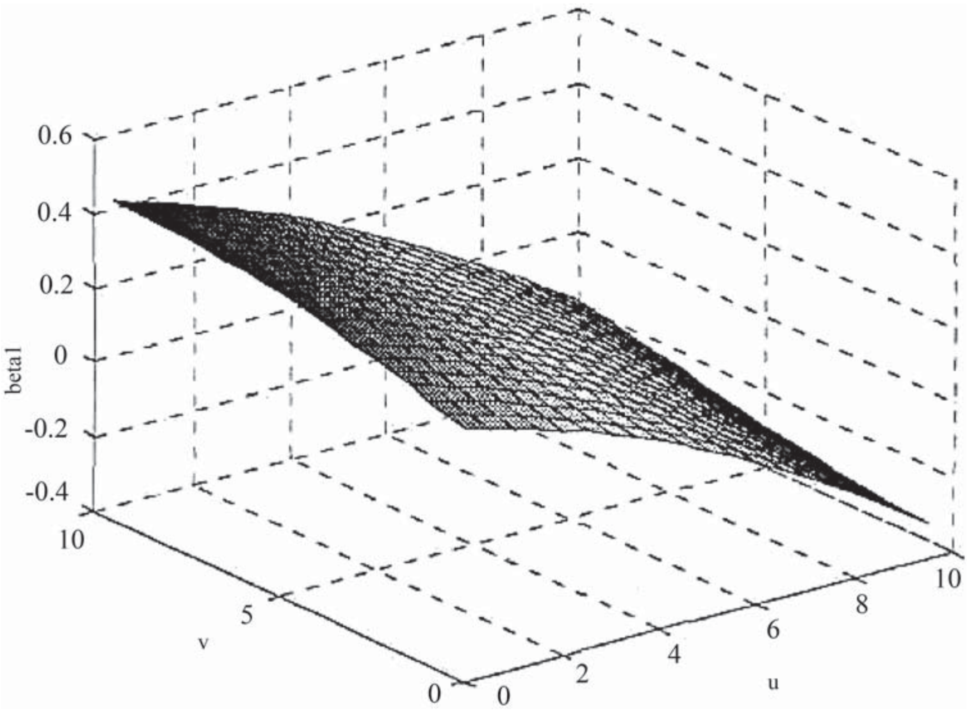

Figure 2 Broadband = 5, Gaussian weight function.

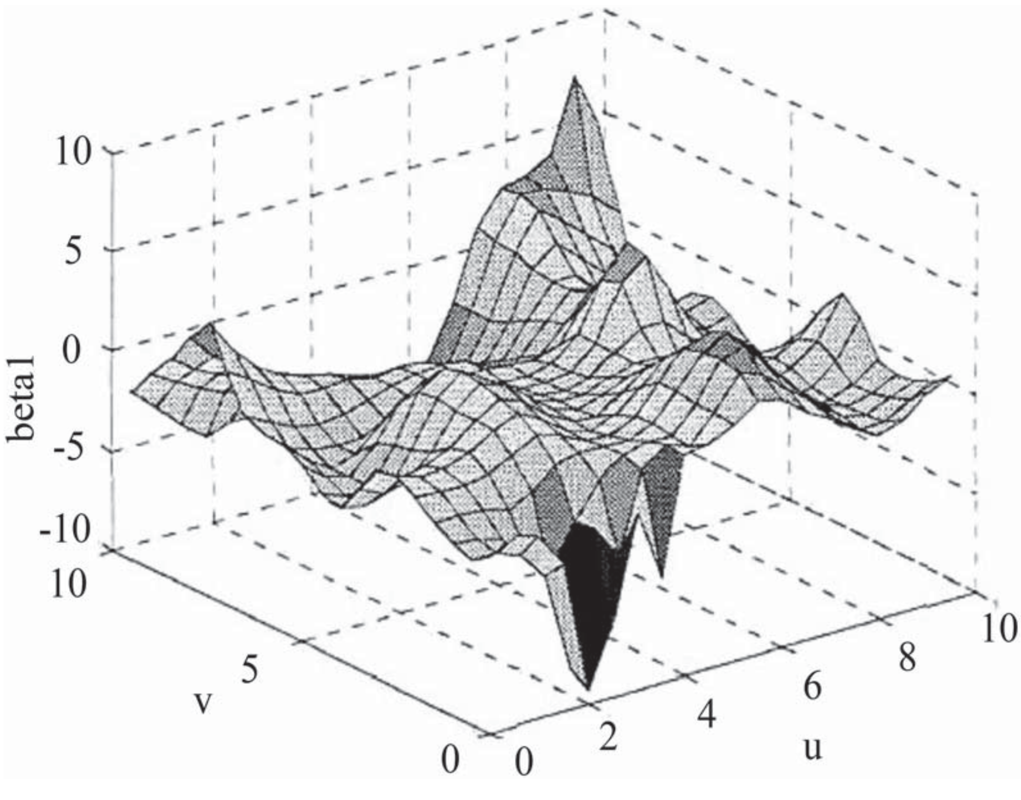

Figure 3 Broadband = 2, double square weight function.

ing to the results of bandwidth optimization, we use Gaussian weight, double square kernel weight and k-nearest neighbor method to estimate the parameters.

We estimate the bandwidth of the model and calculate the optimal value of the software parameters according to the above language weight function. Figures 7 to 11 are the spatial surface diagrams of the estimated value of β_1 obtained by different estimation methods, where Figure 7 is the coefficient 1 of the variable x of the equation to obtain the simulation data, which is taken as $\beta_1 = \ln\left(\frac{1+u_i+v_i}{5}\right)$. Figure 8 is the fitting result of geographic weighted logistic regression model with Gaussian weight function, Figure 9 is the fitting result of bandwidth regression model obtained by k-nearest neighbor method, Figure 10 is the fitting result of double square kernel

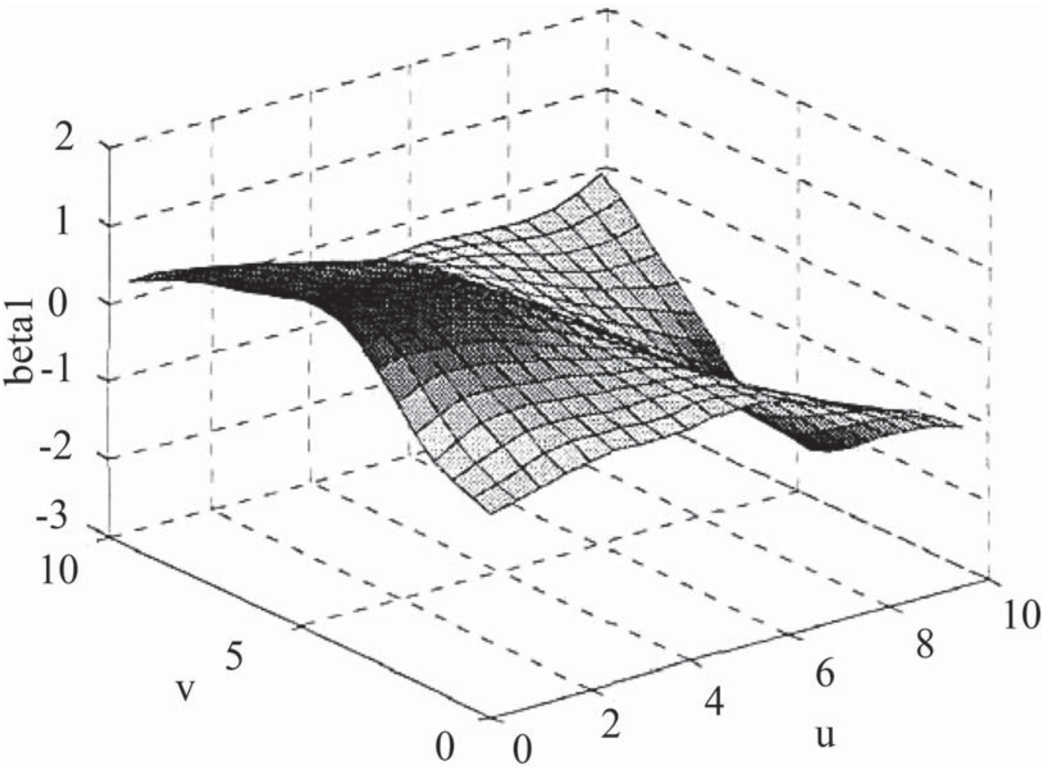

Figure 4 Broadband = 5, double square kernel weight function.

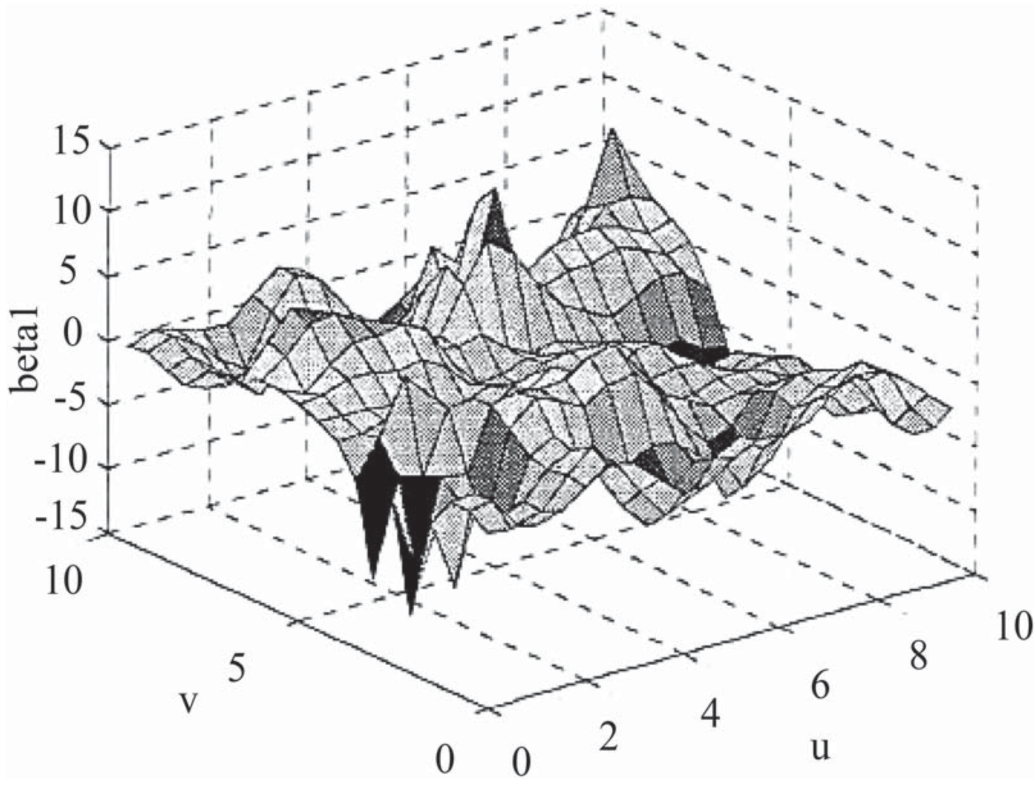

Figure 5 Scale = 0.01, k nearest neighbor method.

regression model, and Figure 11 is the fitting result of the general logistic regression model.

It can be seen from Figure xxx that the general generalized linear model cannot capture the spatial heterogeneity of spatial objects. The parameter estimation obtained by it is the average value of all spatial object parameters and cannot explain the spatial nonstationary risk of spatial objects. However, in Figures 8–10, we cannot determine with the naked eye the model that best fits the original data. Therefore, we need to test the significance of the regression parameters of the model

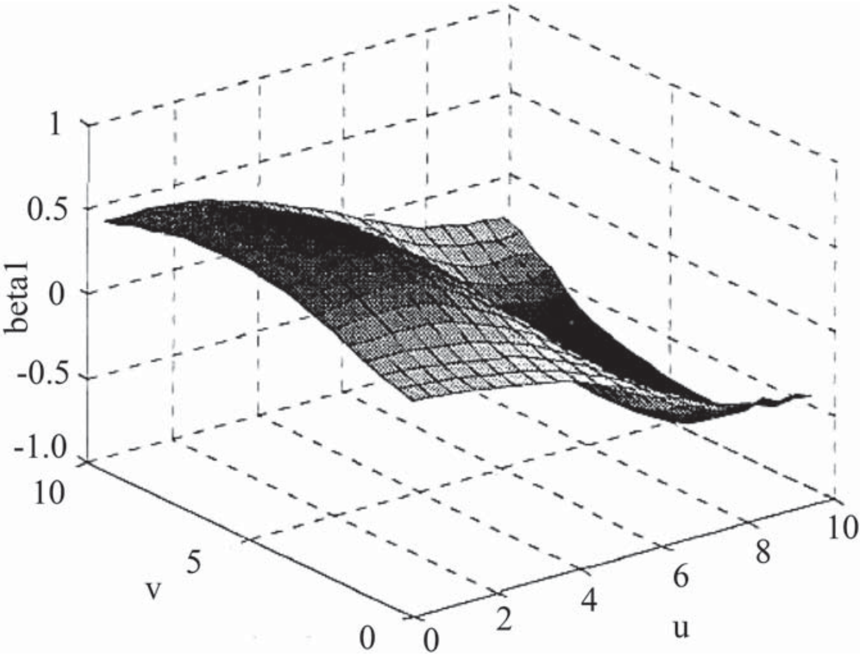

Figure 6 Scale = 0.2, k nearest neighbor method.

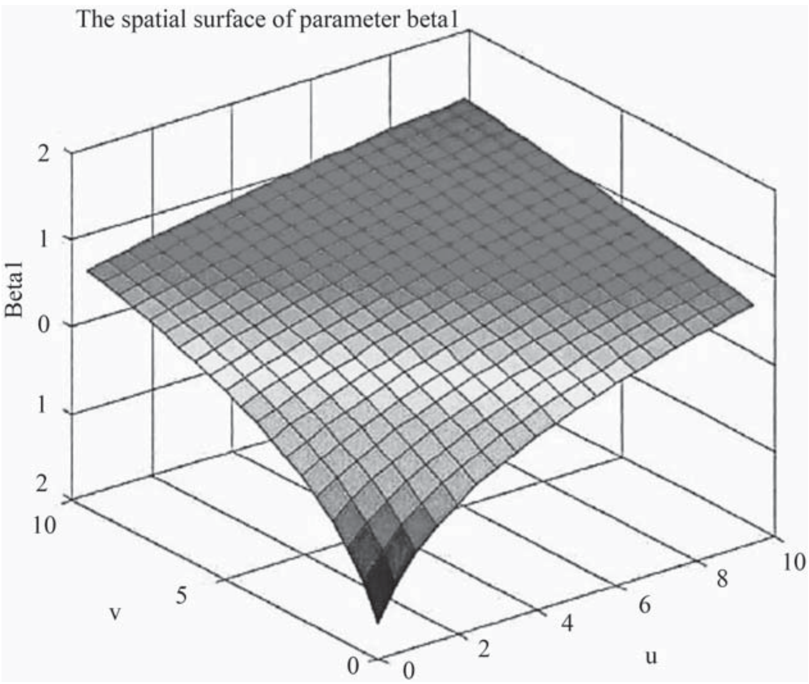

Figure 7 Parameter surface of the original simulation equation.

and compare the goodness of fit of the regression model to select the most appropriate model.

4. EXPERIMENT

The purpose of this section is to investigate whether the better parameter estimators obtained by the Elastic-Net method can be used when there is multicollinearity in the generalized linear model. This section contains a brief description of the way to generate data, and a discussion of results.

4.1 Experimental Design

An important factor of this paper is the correlation between independent variables, and the superiority of ridge parameters is compared by correlation. In order to change the correlation coefficient conveniently, the following equation is used for data generation:

$$x_{ij} = (1 - \rho^2)^{\frac{1}{2}} z_{ij} + \rho z_{ij} \tag{16}$$

where z_{ij} is generated from the standard normal distribution. ρ we take the corresponding four values: 0.75, 0.85, 0.95 and

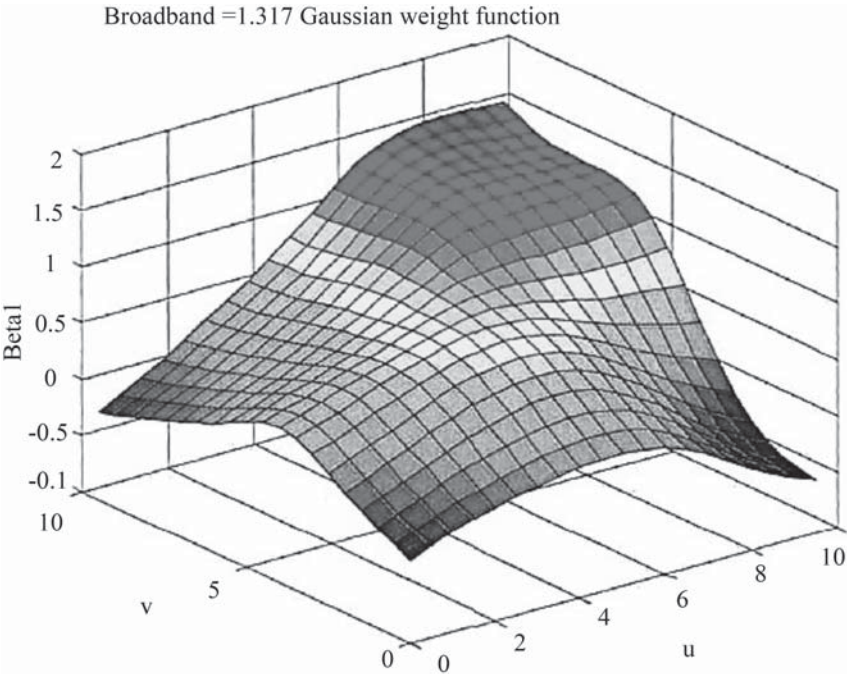

Figure 8 Geographic weighted estimation of Gaussian weight function.

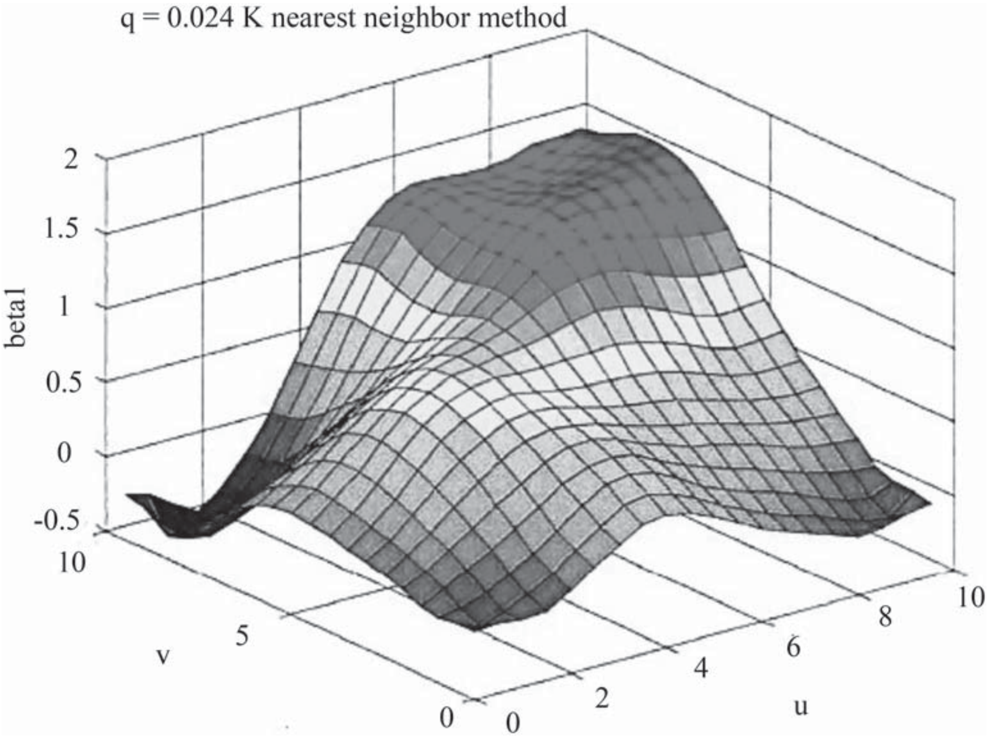

Figure 9 Geographic weighted estimation of k nearest neighbor method.

0.99. The n observations of the dependent variable are derived from the $Be(P_i)$ distribution. Among them:

$$p_i = \frac{\exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_l x_{ip})}{1 + \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_l x_{ip})} \tag{17}$$

We select parameter 1, and the value of $\beta_1, \beta_2, \ldots \beta_p$ satisfies the common constraint $\sum_{i=1}^{p} \beta_i^2 = 1$ in simulation studies and $\beta_1 = \ldots = \beta_p$.

In addition, the other factor we chose was the sample size n, because increasing the sample size can reduce the variance of the parameter estimates. In this simulation study, the sample size n is 50, 100, 150, 200, 250, 300, 350, 400, 450 and 500, respectively.

Finally, we need to consider the number of independent variables. Because this factor determines the number of independent variables that work best with that parameter k, in this simulation study, the number of independent variables is 2, 4, and 8.

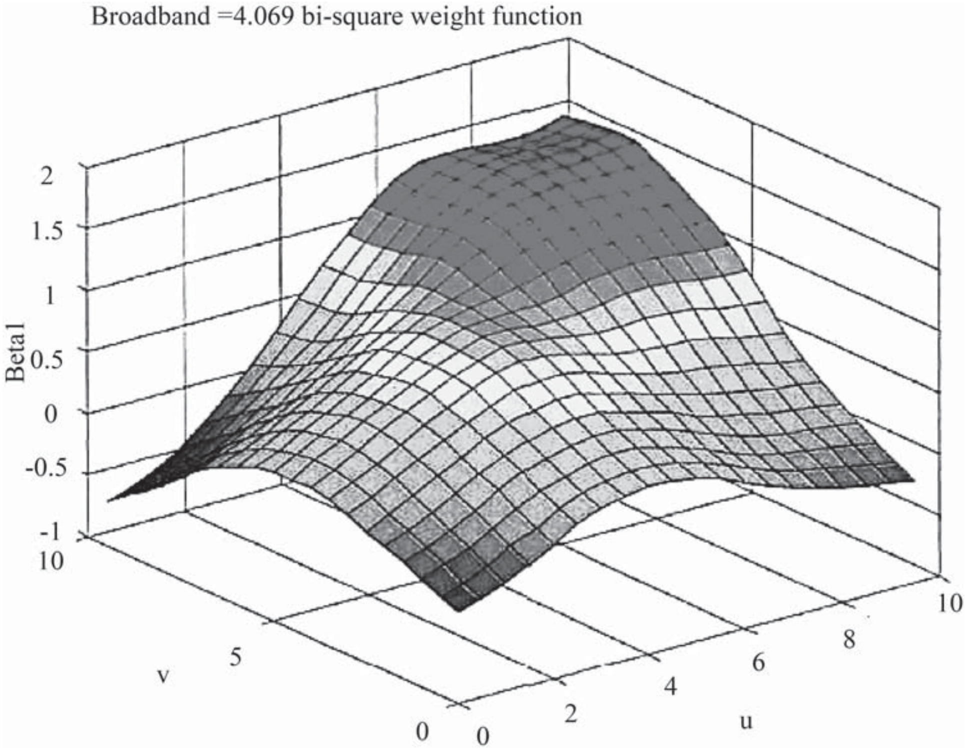

Figure 10 Geographic weighted estimation of bisquare kernel weight function.

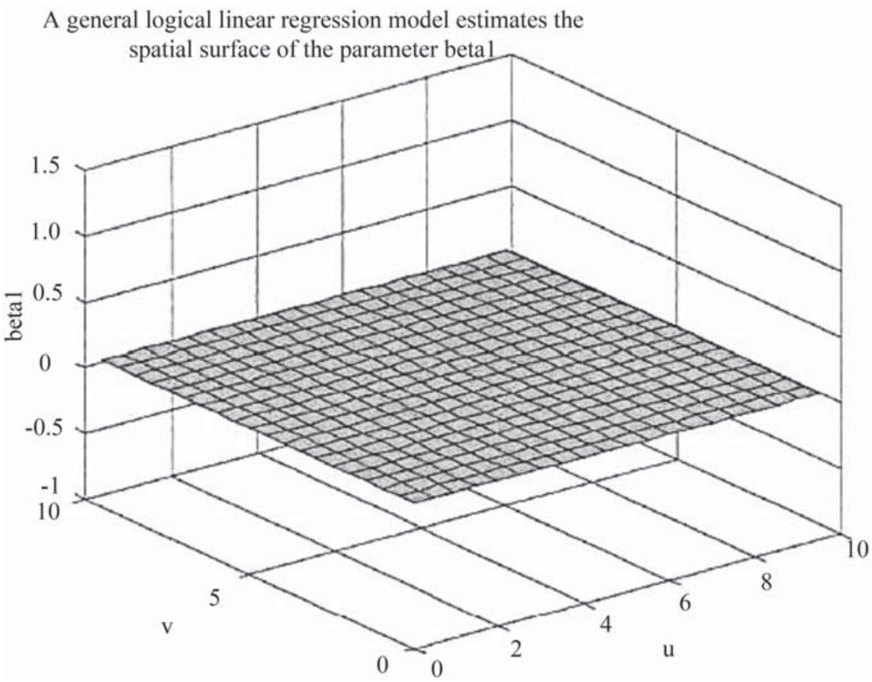

Figure 11 Global generalized model estimation.

4.2 Experimental Results

4.2.1 Accuracy of Model Data Classification

In order to verify the effectiveness of the proposed model for model data classification, Hierarchical Bayesian generalized linear model (Reference [7] method), optimized K-prototypes clustering generalized linear model (Reference [8] method), mixed generalized linear model for statistical inference (Reference [9] method) and mixed generalized linear model for Elastic-Net (this method), were used to detect the classification accuracy of model data. The result for model accuracy are shown in Figure 12.

Figure 12 shows that there are significant differences in the classification accuracy of model data obtained by different methods. When the sample size is 50GB, the classification accuracy of the model data obtained by the method proposed in this paper is 95%, 76% of the model data in Reference [7],

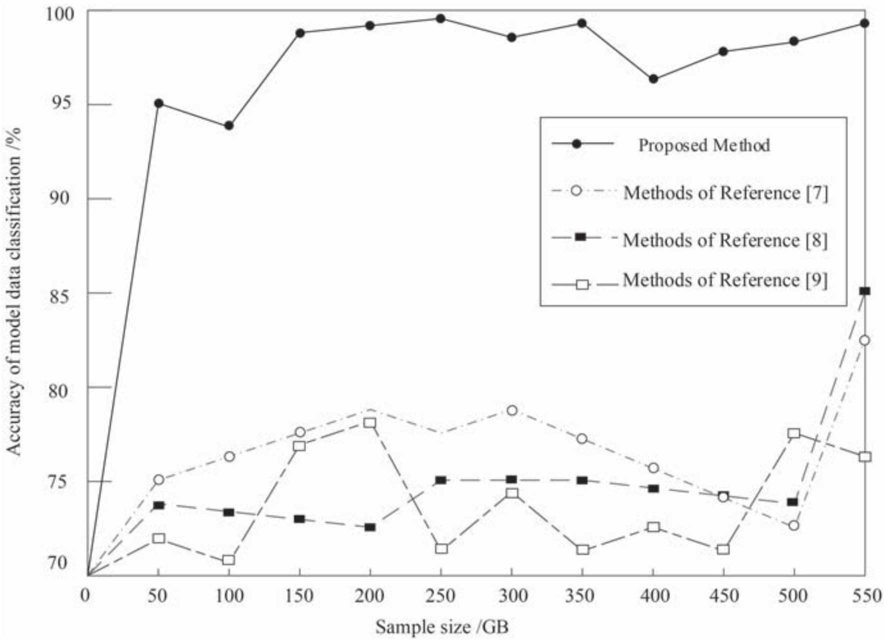

Figure 12 Classification accuracy of model data.

73.5% of the model data in Reference [8], and 72.1% of the model data in Reference [9]. When the sample size is 300GB, the classification accuracy of model data proposed in this paper is 98%, 78% of the model data proposed in Reference [7], 74.8% of the model data proposed in Reference [8] and 74% of the model data proposed in Reference [9]. When the sample size reaches 550GB, the classification accuracy of the model data in this paper is 98.3%, 82.5% of the model data in Reference [7], 85% of the model data in Reference [8] and 76% of the model data in Reference [9]. The proposed method always has a high classification accuracy of model data under the above conditions, indicating that the proposed method has a good classification accuracy of model data.

4.2.2 Recall Rate of Model Data Classification

To verify the effectiveness of our proposed model in terms of model classification, A hierarchical Bayesian generalized linear model (Reference [7]), a generalized linear model for optimized K-prototypes clustering (Reference [8]), a mixed generalized linear model for statistical inference (Reference [9]), and a mixed generalized linear model for Elastic-Net (this method) were used to detect the recall rate of type DATA classification. The recall rate results are shown in Figure 13.

The analysis of Figure 5 shows that there are great differences in the recall rates of model data classification under different methods. When the sample size is 100GB, the classification recall rate of the model data obtained by the proposed method is 97%, 77.6% of the model data in Reference [7], 78% of the model data in Reference [8] and 82.5% of the model data in Reference [9]. When the sample size is 300GB, the classification recall rate of model data in this method is 98%, 77.5% of model data in Reference [7], 78.2% of model data in Reference [8] and 79.8% of model data in Reference [9]. When the sample size is 550GB, the classification recall rate of the model data in this paper is 98.2%, 78% of the model data in Reference [7], 83% of the model data in Reference [8] and 82% of the model data in Reference [9]. The proposed method always has a high recall rate of model data classification, which effectively verifies the effectiveness of the proposed method in generalized linear models.

4.2.3 Time Taken to Classify Model Data

In order to verify the efficiency of the model in the process of model classification, The hierarchical Bayesian generalized linear model (Reference [7]), optimized K-prototypes clustering generalized linear model (Reference [8]), mixed generalized linear model for statistical inference (Reference [9]), and mixed generalized linear model for Elastic-Net (this method) were used to select time series for model data classification. The time result is shown in Figure 14.

Figure 14 shows that the classification time of model data varies according to the different methods applied. When the sample size is 50GB, the model data classification time of the method proposed in this paper is 0.8s, 5.2s for the method in Reference [7], 9s for the method in Reference [8] and 11s for the method in Reference [9]. When the sample size is 200GB, the model data classification time of this method is 1.2s, that of the method in Reference [7] is 9s, that of the method in Reference [8] is 13.2s, and that of the method in Reference [9] is 18s. When the sample size is 550GB, the model data classification time of the method in this paper is 3.2s, the model data classification time of the method in Reference [7] is 30s, the model data classification time of the method in Reference [8] is 21.6s, and the model data classification time of the method in Reference [9] is 28s. The model data classification time of the proposed method is much lower than that of other methods, indicating that the model data classification effect of the proposed method is better.

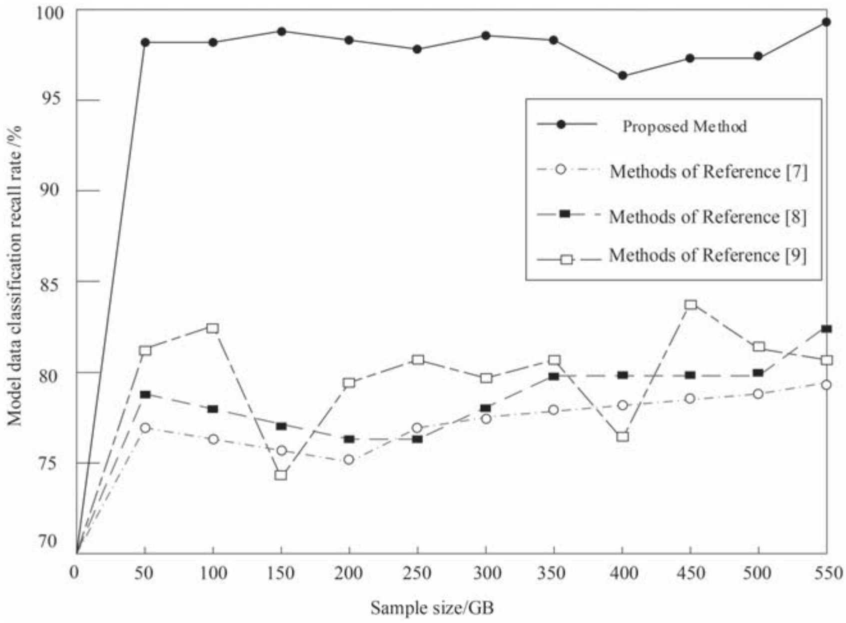

Figure 13 Recall rate of model data classification.

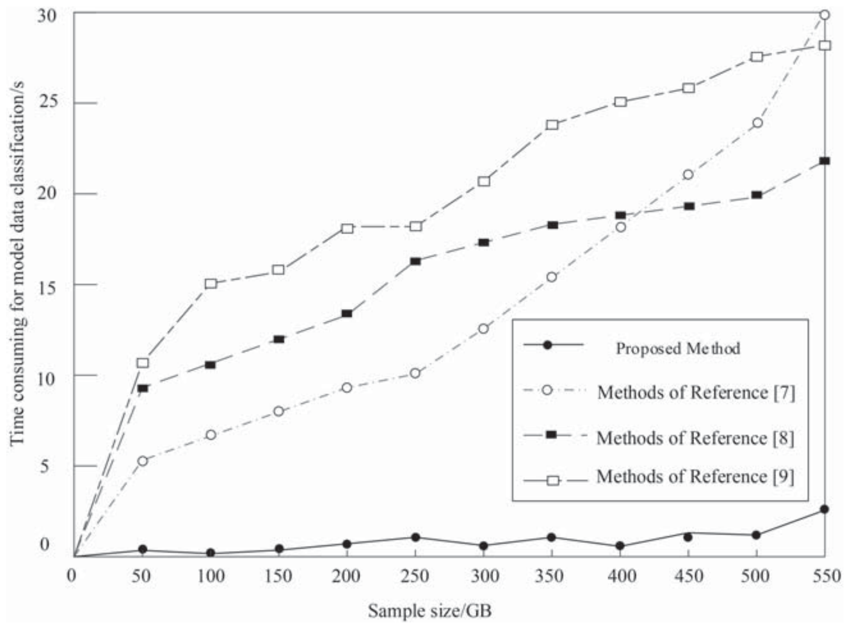

Figure 14 Model data classification time.

5. CONCLUSION

This paper proposes a generalized linear model optimization method based on the Elastic-Net method. It presents the structure of the general linear model and the generalized linear model, analyzes Elastic-Net method and its related properties, incorporates Elastic-Net method into the generalized linear model, and analyzes the classification effect of the model by means of examples. The experimental results show that:

(1) When the sample size reaches 550GB, the model data classification accuracy of the proposed method is 98.3%. Under this condition, the proposed method always has a high classification accuracy of model data, indicating that the model data classification accuracy of the proposed method is good.

(2) When the sample size is 550GB, the model data classification recall rate of the proposed method is 98.2%. The proposed method always has a high model data classification recall rate, which confirms the effectiveness of the proposed method in the generalized linear model.

(3) When the sample size is 550GB, the model data classification time of the proposed method is 3.2s, indicating that the model data classification effect of the proposed method is good.

ACKNOWLEDGEMENT

The research is supported by: The Application and example analysis of Elastic-Net Procedure for generalized linear model (No.2021KY1513); The Application and example analysis of Elastic-Net Procedure for generalized linear model(No.GXJZ202010).

REFERENCES

1. Huang XC. Error probability analysis of web page information search based on generalized linear model. Journal of Heze University. 2019; 41(2):14–20.
2. Sulekan A, Suhaila J, Wahid NAA. Assessing the effect of climate factors on dengue incidence via a generalized linear model. Open Journal of Applied Sciences. 2021; 11:549–563.
3. Amorim DJ, Santos ARP, Piedade GN, Faria RQ, Silva EAA, Sartori MMP. The use of the generalized linear model to assess the speed and uniformity of germination of corn and soybean seeds. Agronomy. 2021; 11(3): 588.
4. Xing YQ, Wu LC, Nie XF. Statistical diagnosis of dual generalized linear model based on Pena distance. Applied mathematics. 2019; 32(4):739–746.
5. Mahmood T, Balakrishnan N, Xie M. The generalized linear model-based exponentially weighted moving average and cumulative sum charts for the monitoring of high-quality processes. Applied Stochastic Models in Business and Industry. 2021; 37(4): 703–724.
6. Teng YQ, Ma WJ. Risk assessment of rice planting based on generalized linear model. Mathematics in Practice and Cognition. 2019; 49(2):1–17.
7. He JF, Fu YC, Xiong J. Research on small field estimation method based on hierarchical Bayesian generalized linear model. Mathematical Statistics and Management. 2019; 38(2):247–260.
8. Zhu YJ. Clustering method of mixed attribute data based on generalized linear model. Science Technology and Engineering. 2021; 21(4):1448–1453.
9. Yuan QL, Wu LC, Dai L. Statistical inference of mixed generalized linear models. Journal of Engineering Mathematics. 2019; 36(5):525–534.
10. Gunnarsdottir KM, Gamaldo C, Salas RM, Ewen JB, Allen RP, Hu K, et al. A novel sleep stage scoring system: Combining expert-based features with the generalized linear model. Journal of Sleep Research. 2020; 18(25):126–132.
11. Zhang H, Zou Q, Yang H. Estimation for high-dimensional multi-layer generalized linear model – Part II: The ML-GAMP Estimator. Information Theory; 2020. Available from: https://doi.org/10.48550/arXiv.2007.09827.
12. He JF, Fu YC, Xiong J. The study of small area estimation method based on hierarchical Bayesian generalized linear model. Journal of Applied Statistics and Management. 2019; 122(8):216–225.
13. Dahl NJ, Ammar AM, Knott A, Andersen MAE. An improved linear model for high-frequency class-de resonant converter using the generalized averaging modeling technique. IEEE Journal of Emerging and Selected Topics in Power Electronics. 2020; 8(3):2156–2166.
14. Hamdanah FH, Fitrianah D. Analisis Performansi Algoritma Linear Regression dengan Generalized Linear Model untuk Prediksi Penjualan pada Usaha Mikra, Kecil, dan Menengah. Jurnal Nasional Pendidikan Teknik Informatika. 2021; 10(1): 23–32.
15. Pare D, Kassienou L, Longin S, Pare Y. Solving generalized linear model of black-scholes with classical finite volume method. International Journal of Numerical Methods and Applications. 2021; 20(1):17–40.
16. Yang XW, Zhao KB, Liu XG, Wang DY. Double logarithm law and strong consistency of weighted generalized linear model selection in independent and dependent cases. Journal of Yichun University. 2020; 42(3):40–49.
17. Ghosal R, Ghosh SK. Bayesian inference for generalized linear model with linear inequality constraints. Computational Statistics & Data Analysis. 2021; 166:107335.
18. Dwinata A, Kurnia A, Sadik K. A Bayesian approach for generalized linear model using non-local prior (Case Study: Poverty Status in East Java). Journal of Physics: Conference Series. 2021; 1863(1): 012026.
19. Guo WJ, Jin H. Design and simulation of financial data evaluation algorithm based on support vector machine. Electronic Design Engineering. 2021; 29(18):17–20+25.
20. Qing MJ, Luo ZN. Nonlinear distortion optimization design of power amplifier in F-OFDM system. Computer Simulation. 2021; 38(8):299–304.