

## Research Article

# Application of a New Loss Function-Based Support Vector Machine Algorithm in Quality Control of Measurement Observation Data

Youping Wu  <sup>1</sup> and Guoqiang Tao  <sup>2</sup>

<sup>1</sup>Department of Applied Engineering, Gandong College, Fuzhou 344000, China

<sup>2</sup>School of Surveying and Mapping Engineering, East China University of Technology, NanChang 330013, China

Correspondence should be addressed to Youping Wu; 199460011@ecut.edu.cn

Received 21 April 2022; Revised 16 May 2022; Accepted 20 May 2022; Published 8 June 2022

Academic Editor: Zaoli Yang

Copyright © 2022 Youping Wu and Guoqiang Tao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The loss function of the traditional support vector machine (SVM) method consists of hinge function and regularization, which is difficult to achieve the quality control of observation data. It requires a new loss function to measure the quality of the observed data. At this stage, researchers will use data cleaning or data preprocessing to process observational data. The data preprocessing method will normalize the data features, which will make the data process the same interval and a distribution. However, this method is also difficult to guarantee the accuracy of the data. This study uses the SVM method to study the integrity, unity, and accuracy of the observation data. This study also proposes a new loss function, which can obtain the uncertainty distribution of the observed data. The loss function of traditional SVM methods is that the uncertainty distribution of the data cannot be obtained. The results show that the SVM algorithm with the new loss function has better accuracy in processing the observed data. For the maximum prediction error of the observed data, the error with the new loss function is only 2.58%, which is reduced by 0.21% compared to the SVM method with the old loss function. However, the prediction error of the SVM algorithm with the old loss function is 2.78%, and the SVM algorithm with the new loss function has obvious prediction accuracy.

## 1. Introduction

With the continuous improvement of computer performance and hardware equipment, machine learning algorithms have been developed rapidly [1, 2]. It can realize automated workflow, and it can well assist people in life and production. Machine learning algorithms mainly include supervised learning, unsupervised learning, and reinforcement learning [3, 4]. Unsupervised learning is a kind of learning algorithm without labeled data, which is mainly used in super resolution and other fields. Reinforcement learning is an algorithm that has a strong correlation with the environment [5]. The SVM algorithm is one of the machine learning methods, and it is a supervised learning algorithm that contains the mapping relationship between input and output. Supervised learning algorithms can map the relationship between input data and output data well,

and they can well reflect the relationship between unknown data. Supervised learning is one of the most widely used algorithms in life compared to unsupervised learning and reinforcement learning. Supervised learning is an algorithm that requires input data and label data [6, 7]. It can establish a nonlinear relationship between input data and output data, which is embodied in the form of weights and biases. A supervised learning algorithm needs to learn complex relationships between large datasets, and then it can predict unknown data based on these weights and biases. This requires the accuracy, integrity, and uniformity of observational data [8, 9]. If the accuracy of the observed data is relatively poor, it is easy to cause poor learning effect and generalization ability. If the integrity of the observed data is relatively poor, this will make it difficult for the learning algorithm to complete the matrix operation [10]. Uniformity is also important for observational data. If the uniformity of

observational data is poor, it will be difficult for the learning algorithm to reach the convergence level. The integrity, uniformity, and accuracy of observational data will affect the quality of observational data. The quality of the observed data also affects the accuracy of the algorithm [11]. The improvement of the quality of the observation data will improve the accuracy and convergence speed of the algorithm. In this study, the relevant data of supervised learning are selected as the research object, because it needs to provide input data and label data. Observational data are the dataset that needs to be collected.

Whether it is a regression task or a classification task, it can be achieved by the support vector machine (SVM) method. This study intends to use the SVM algorithm to evaluate the quality of the observation data [12, 13]. The SVM algorithm has obvious advantages in solving small sample data and high-dimensional data, and it can map the relationship between input and output data well. However, the goal of this study is to look for completeness, accuracy, and uniformity of observational data. The traditional SVM method can only map the relationship between the data, but it is difficult to find the defects of the observation data [14]. This is because the loss function of the SVM method itself has certain defects. The SVM loss function is composed of a hinge function and a regularization term [15]. However, the loss function of the hinge function has relatively high data requirements. The hinge function requires that the data must be correctly classified or regressed and that the loss value is correct only if the data have a high enough confidence. The data source of this study is observational data, which is missing or incomplete [16]. This will cause the hinge function to fail to perform regression predictions correctly, and this incomplete dataset also has high uncertainty. The loss function of the traditional SVM needs to be modified into a new loss function, which can obtain the uncertainty of the data [17, 18]. Once the uncertainty of the data is captured, the researcher can ensure the correct utilization of the hinge function by adjusting the data for the location of the larger uncertainty [19]. Therefore, the loss function of traditional SVM needs to be embedded in the distribution of data uncertainty, which can ensure the utilization of observation data quality monitoring [20, 21].

The SVM algorithm with the new loss function will display the uncertainty of the observation data [22]. It can understand the quality of observation data according to the uncertainty distribution of observation data, which can achieve the goal of quality control of observation data. For the quality control research of observational data, this research is mainly aimed at the integrity, uniformity, and accuracy of observational data. If the SVM algorithm with the new loss function can show the uncertainty of the data, the area of large uncertainty means that the observation data is incomplete or has low accuracy. It can be adjusted by adjusting the composition of the observed data or by adjusting the hyperparameters of the SVM algorithm. However, it only adjusts the hyperparameters of the SVM algorithm, which has a relatively small adjustment range for the quality control of the observed data. Researchers can

adjust the value of the observation data according to the distribution of uncertainty, which can also achieve the purpose of controlling the quality of the observation data. This method can adjust the quality of the observational data to a greater extent. With the SVM method with a new loss function, once it has determined the uncertainty distribution area of the data, it can adjust the data in the areas with more uncertainty. Areas with more uncertainty represent lower quality data.

This study designs a new loss function for the SVM algorithm based on the current quality control of observational data, which can display the uncertainty distribution of observational data. This research is mainly divided into five parts for related introduction: the first part mainly introduces the importance of quality control of observation data and the defects of SVM algorithm. The loss function of the SVM algorithm and related research on the quality of the observed data are studied in Section 2. Section 3 explains the principle of the SVM algorithm and the calculation method of the SVM algorithm with the new loss function. The application accuracy of SVM algorithm with the new loss function and traditional SVM algorithm in observation data quality control is investigated in Section 4. It mainly uses average error, error scatter plot, and error hot spot distribution plot to analyze the accuracy of SVM with the new loss function in predicting the quality of observed data. Section 5 is the summary part of the study.

## 2. Related Work

Machine learning algorithms have been applied in many fields, and the quality of observation data determines the accuracy and generalization ability of the algorithm. Many researchers have conducted research on the quality of observational data. Many researchers have also studied the loss function of SVM for different application objects, and they have also achieved relatively good success. Mahmood et al. [23] proposed a solution for the inaccuracy of data collected by health monitoring equipment. The data collected by this solution need to be able to determine the accuracy of future data quality. It investigates the handling of missing values for observational data according to the chart method of Hotelling T-square control. The research results show that this method can accurately detect the problem of missing data, and it can also verify the existence of errors in the data. This method can solve the problems of missing values and low accuracy of health monitoring equipment. Yang et al. [24] used data on sea surface temperature as their research object. He has analyzed the data for a quality control scheme utilizing validated extreme values and time adjacent-based temporal signature data, and a method that can detect outliers in the collected sea surface temperature. The correlation between these anomalies and the actual data is also relatively poor, which is also the solution for the quality detection of sea surface temperature data. The research results show that the data quality detection task of sea surface problem is an important process for meteorology. Castelao [25] has come to believe that sensor error is an unavoidable error for ocean

measurements. He also believed that oceanographic measurements require a data quality control program to check the integrity and accuracy of the data. Aiming at the low efficiency of manual data quality control procedures, he proposed a quality detection method for marine scientific data based on machine learning methods. This approach can reflect inaccuracies in marine data as outliers. The research results show that this machine learning method can reduce the error rate by 50% in the data quality detection task. Liu et al. [26] believed that the uncertainty of the refractive index of the radio data would lead to inaccuracy in the detection of the refractive index distribution. He uses data quality control methods to eliminate radio-detected low-quality data problems, which are developed based on data assimilation methods. This approach can also quantify deviations and correlations between data. The NCAR data were used for validation by him, which showed that the local spectral width (LSW) data quality control procedure can reduce the presence of low-quality data. The loss function of the SVM algorithm is mainly composed of a hinge function and a regularization term, and it also has certain defects. Many researchers have done a lot of research on the loss function of the SVM algorithm. Panup et al. [27] use the SVM algorithm to solve data classification problems. He combined the stochastic gradient descent method and the traditional SVM algorithm to design a generalized pinball support vector machine GPSVM algorithm. He also uses the kernel method to evaluate the performance of the GPSVM algorithm, which is mainly aimed at nonlinear data classification problems. It proposed a new loss function with good adaptability and weak noise sensitivity for this new SVM algorithm. This novel SVM function and new loss function also have a good performance on large datasets. Zheng et al. [28] already believe that the twin support vector machine TSVM has attracted the attention of a large number of researchers, and it also follows the basic computational principles of SVM. The big difference between TSVM and SVM is that it is more sensitive to outliers. He introduced the related concept of entropy induction into the TSVM algorithm. The research results show that the novel TSVM algorithm has better robustness and accuracy in handling nonlinear data classification tasks. Tanveer et al. [29] also found that the TSVM algorithm also has unstable data and strong noise sensitivity, which will cause the inaccuracy of the algorithm. He proposed a novel Pinball Twins Support Vector Machine (PTSVM) algorithm to solve the classification problem of nonlinear data. The corresponding loss function also changes, which results in a new loss function. The results show that the PTSVM algorithm has better stability in nonlinear data classification tasks, and it is less sensitive to noise. From the above literature review, it can be seen that the quality of observation data and variants of SVM algorithms have always been hot topics of research, both of which will directly lead to inaccuracy and strong sensitivity of prediction or classification tasks. This study proposes an SVM algorithm with a new loss function for quality control of observational data. For the study of observed data quality, this study takes uncertainty into

account in the loss function of SVM. The SVM algorithm with the new loss function will show the uncertainty of the data.

### 3. The Introduction of SVM Algorithm with New Loss Function

*3.1. The Importance of New Loss Function to the Quality of the Observed Data.* The loss function of the traditional SVM algorithm is composed of a hinge function and a regular term, and it is difficult to deal with the quality control task of the observation data [30, 31]. This is because the observed data are incomplete and inaccurate. The hinge function has relatively high requirements on the certainty of the data. Therefore, a new type of loss function is necessary to control the quality of observed data. This study intends to change the form of the loss function and introduce the concept of uncertainty into the loss function of SVM [32]. In this study, the uncertainty distribution of the loss function response is used to determine the distribution area of the incompleteness and inaccuracy of the observed data. Once the incompleteness of observational data and the distribution area of uncertainty or nonuniformity are found, the purpose of quality control of observational data can be achieved. Therefore, the new loss function plays an important role in the quality control of the observed data.

*3.2. The Introduction of the Principle of SVM Algorithm.* SVM is a binary classification algorithm, which is mainly applied to linear data and high-dimensional data. SVM needs to find a hyperplane to cut the samples accordingly. The principle of segmentation is to maximize the hyperplane. The SVM method finally needs to be transformed into a convex quadratic programming problem to solve the maximization problem. Figure 1 shows the schematic diagram of SVM algorithm principle. It can be seen that the sample points are efficiently divided by a hyperplane. Figure 1 shows a technique for dividing data by hyperplanes. For the linear division of the SVM method, there are mainly three ways to divide the data of the samples. If the data have good linear separability, SVM can implement a linear support vector machine by a hard margin maximization method. If the data have approximately linearly separable properties, SVM can achieve linear division of the data by maximizing a soft margin. If the nonlinearity between the data is strong, it is difficult to achieve linear division through the above two maximization methods, and it needs a kernel method to maximize the data.

However, in practical engineering applications or research, the SVM method often faces some nonlinear data, and it is difficult to achieve the current data division only by relying on linear data processing methods. A hyperplane containing a straight line cannot effectively divide nonlinear data. It requires a hyperplane in the form of a circle or an ellipse to divide nonlinear data. Figure 2 shows the flow of nonlinear data partitioning using ellipses. For the hyperplane of Figure 1, it is suitable for processing some linearly distributed data. However, the elliptical hyperplane of

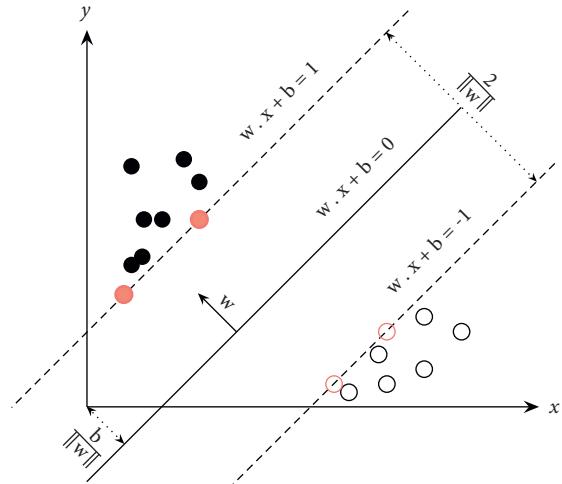


FIGURE 1: The schematic diagram of SVM algorithm principle.

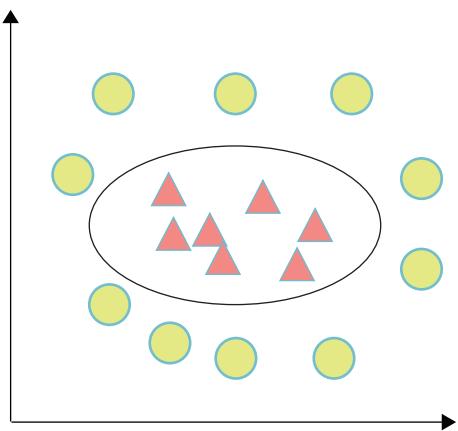


FIGURE 2: The schematic diagram of hyperplane solution for nonlinear data.

Figure 2 can handle some nonlinear data. The solution for nonlinear problems is more difficult. Moreover, the SVM method includes integral and derivation operations, which brings great difficulties to the nonlinear solution of SVM method. In general, nonlinear data need to be mapped from the original space to a high-dimensional space, which can convert the nonlinear distribution of samples into a linear distribution. In this way, the nonlinear problem can be transformed into a linear problem and solved accordingly.

Data for nonlinear data tend to be transformed into calculations for linear data. Therefore, the calculation method of this study is presented according to the way in Figure 1. The first step of the SVM method needs to calculate the interval between two heterogeneous support vectors, the interval is the projected distance on the hyperplane, and the calculation method is shown in equation (1). Equations (2) and (3) show the relationship satisfied by the two positive and negative vectors in the SVM method, where  $\vec{x}_+$  is the positive support vector,  $\vec{x}_-$  is the negative support vector,  $\gamma$  is the vector interval,  $w$  represents the weight of the support vector, and  $b$  represents the bias of the support vector:

$$\gamma = \frac{(\vec{x}_+ - \vec{x}_-) \cdot \vec{w}^T}{\|w\|}, \quad (1)$$

$$1 * (w^T \vec{x}_+ + b) = 1, \quad (2)$$

$$-1 * (w^T \vec{x}_- + b) = 1. \quad (3)$$

The goal of the SVM method is to maximize the margin, and equation (4) shows how the maximized margin is calculated. For the computational source of SVM, it is often treated as a convex quadratic programming problem, which is for ease of solution. This study uses the Lagrange multiplier method to obtain the loss function of the SVM method, and the calculation equation is shown in equation (5), where  $L(w, b, \alpha)$  stands for Lagrangian function and  $\alpha$  is the correlation coefficient:

$$\max_{w,b} \frac{2}{\|w\|}, \quad (4)$$

$$\text{s.t. } y_i(w^T \vec{x}_i + b) \geq 1,$$

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 + \sum_{i=1}^m \alpha_i (1 - y_i(w^T \vec{x}_i + b)). \quad (5)$$

**3.3. The Principle of the New Loss Function.** The main goal of this study is to design a new loss function for the SVM method, which takes into account the uncertainty of the data distribution in this study. The loss function of the traditional SVM method only contains the hinge function and the regularization term, and it does not contain the uncertainty distribution of the data. The traditional SVM method is not conducive to showing the quality of observational data. Generally speaking, the amount of observation data is relatively large. It can identify poor-quality data through the distribution of uncertainty. If there is an area of high uncertainty distribution, then this is an area of poor observational data quality. Figure 3 shows the workflow of the SVM method with the new loss function. The old loss function contained in the SVM method only calculates the corresponding error between the predicted value and the label value, which in turn adjusts the operation direction of the SVM method. The SVM method with the new loss function will implement multiple sampling processing of multiple data, which will show the uncertainty distribution of the observed data. The new loss function with uncertainty distribution will improve the quality of observational data, which can show the inaccuracy and incompleteness of observational data. Multiple sampling is the key to uncertainty research. Multiple sampling can determine the difference between the average value and the peak value. It can determine the area of greater uncertainty through multiple sampling.

The uncertainty distribution will change the distribution method of the weight of the SVM algorithm, and the weight of the traditional SVM will be distributed in the form of

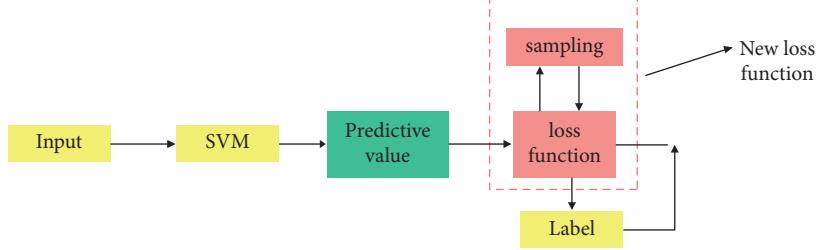


FIGURE 3: The SVM workflow with a new loss function.

points. The SVM with the new loss function needs to show the uncertainty distribution of the observed data. Therefore, the weights in the SVM with the new loss function are distributed probabilistically, and it needs to take into account the prior knowledge of the observed data. Equation (6) shows how the posterior distribution is calculated. The integral operation involved in equation (6) is relatively difficult and equation (7) is a variational approximation operation to equation (6):

$$P(y^*|x^*, X, Y) = \int p(y^*|f^*)p(f^*|x^*, X, Y)df^*, \quad (6)$$

$$P(y^*|x^*, X, Y) = \int p(y^*|f^*)p(f^*|x^*, \omega)p(\omega|X, Y)df^*d\omega. \quad (7)$$

In order to further simplify the integral operation involved in equation (7), the calculation method of Kullback–Leibler (KL) divergence is introduced here. KL divergence is also an approximation of the integral operation. Equation (8) shows how the KL divergence is calculated:

$$\text{KL}(q(\omega)p(\omega)) = \frac{1}{2D} \sum_{i=1}^D \left( p_d \|M_d\|_2^2 + \|b\|_2^2 \right). \quad (8)$$

After a series of variation and approximation operations, the calculation method of the uncertainty of the observed data is shown in equation (9). It is mainly composed of the error term, the KL divergence term, and the log calculation term. This variational and approximate approach can quantify the uncertainty of observed data. This is beneficial to the control of observation data quality:

$$L \approx \frac{1}{2\sigma(f)^2} \|y' - \hat{y}'\|^2 + \frac{1}{2} \log \sigma(f)^2 + \frac{1}{2D} \sum_{i=1}^D \left( p_d \|M_d\|_2^2 + \|b\|_2^2 \right). \quad (9)$$

Equation (10) shows the new loss function of SVM, which is mainly composed of the old loss function and uncertainty terms. The new loss function of SVM will contain two parts: the first part is the hinge function and regularization term of the old loss function and the second part is the uncertainty term of the data.

$$L_{\text{SVM}} = L + L(w, b, \alpha). \quad (10)$$

Although this study needs to deal with the quality of observational data, it also contains more input data. Data cleaning and data preprocessing processes are also required. Data cleaning needs to complete the lack of input data, and it also needs to deal with the places where the data have large defects. Data cleaning does not require processing the output data, which is also the way to find the quality of the observed data through the SVM method. In this study, a normalized data preprocessing method will be used to process the relevant input data.

#### 4. Result Analysis and Discussion

This study changes the loss function of the SVM method, which will be used as a control task for the quality of the observed data. In this study, a small dataset and a large dataset were selected for relevant validation. It will demonstrate the feasibility of the SVM method with a new loss function for quality control of observational data. The quality of observational data will include completeness, accuracy, and uniformity. Figure 4 shows the accuracy of observed data in terms of prediction error using SVM with a new loss function. Overall, the SVM method with the new loss function can predict the quality of the observed data well. The three important factors of the observed data are accuracy, completeness, and uniformity. The SVM method with the new loss function can predict well, and the maximum prediction error is only 2.56%. This part of the prediction error mainly comes from the quality of the accuracy of the observed data. The accuracy of data is different from completeness and uniformity. The characteristics of data integrity and uniformity are easily learned and predicted by SVM, which leads to smaller prediction errors in terms of the integrity and uniformity of observed data. This part of the error is only 1.93% and 1.45%. In general, the SVM algorithm with the new loss function can better predict and control the quality of the observed data. In general, most researchers take 5% as a reasonable prediction error margin. The prediction errors of this study are all within 2%, which further illustrates the feasibility of incorporating a new loss function.

In order to further verify the feasibility and accuracy of the SVM algorithm with the new loss function in the observation data quality control task, this study also selected a large observation dataset for verification. Figure 5 shows the prediction errors on large datasets using two SVM methods.

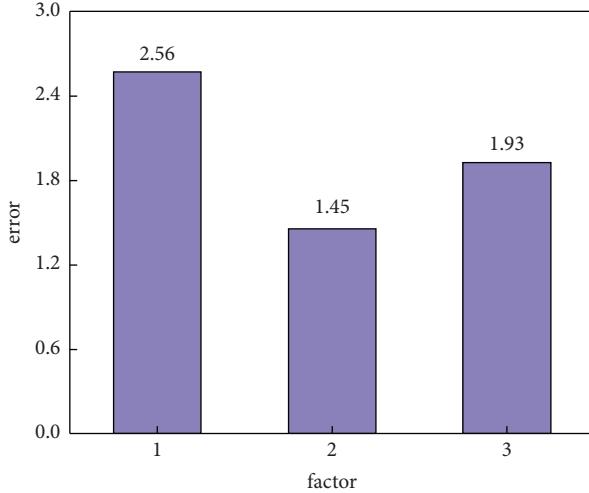


FIGURE 4: The prediction errors on small datasets using the SVM method.

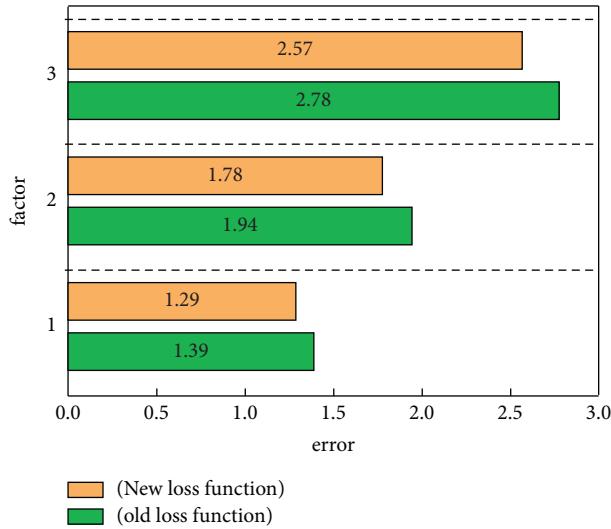


FIGURE 5: The prediction errors on large datasets using two SVM methods.

In general, the prediction error of the SVM algorithm with the new loss function is significantly reduced for the three factors of observed data quality, which shows that the SVM algorithm with the new loss function is more suitable for the prediction task of the factors related to the quality of the observed data. For a large observation dataset, the largest prediction error is only 2.57%. Using the SVM method with a new loss function, this part of the error is also an accuracy factor from the quality of the observation data. The maximum reduction in prediction error of the SVM with the new loss function reaches 0.21% compared to the SVM method with the old loss function. For the other two correlates of observed data quality, prediction errors were within 2%. For the integrity and uniformity of the observed data, the prediction error of the SVM with the new loss function is also reduced by 0.1% and 0.16%, respectively, compared with the SVM method with the old loss function. Compared with

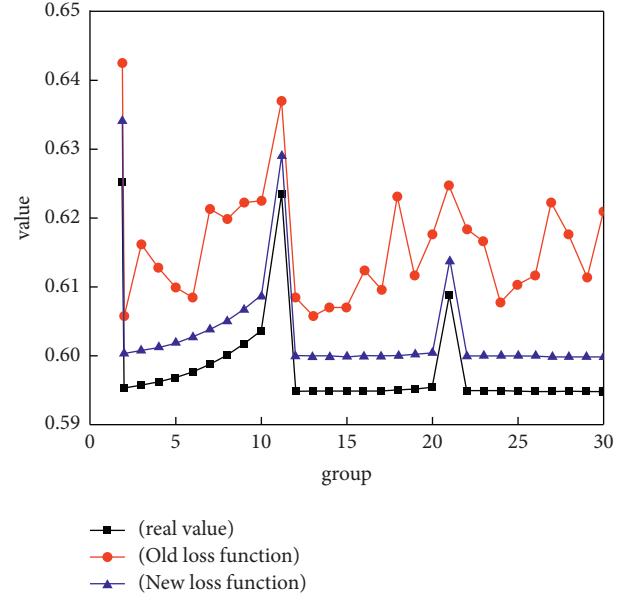


FIGURE 6: The accuracy verification of observational data using two SVM methods.

the SVM method with the old loss function, its prediction error is also reduced to varying degrees.

In order to more intuitively demonstrate the effectiveness of the SVM with the new loss function, this study selects some data from a large dataset to demonstrate the prediction effect of the accuracy factor of the observed data quality. Figure 6 shows the accuracy verification of observational data using two SVM methods. For the change trend of the accuracy factor of the observed data quality, the prediction effect of the SVM method with the new loss function is closer to the actual change trend. Whether it is the peaks or troughs of the data, the SVM method with the new loss function performs better. The change trend of the accuracy predicted by the SVM method with the old loss function is quite different from the actual change trend, and it cannot well reflect the change trend of the accuracy of the observed data. For data values, the data values predicted by the SVM method with the new loss function are in good agreement with the actual data values. However, the data values obtained by the SVM method with the old loss function are quite different from the actual data.

The quality of observational data includes three aspects: accuracy, completeness, and uniformity. Figure 7 shows the integrity error verification of observational data using two SVM methods. Overall, the SVM with the new loss function has a lower error in predicting the completeness of the observed data. Of course, the SVM aspect with the old loss function can also predict the integrity factor of the observed data well, but it only has a larger error than the SVM method with the new loss function. Whether it is the SVM method of the new loss function or the SVM method of the old loss function, the error of the integrity factor of the observed data predicted by it is within 3%. Moreover, for the integrity factor of the observed data, the error between the two is relatively small compared to the accuracy factor of the

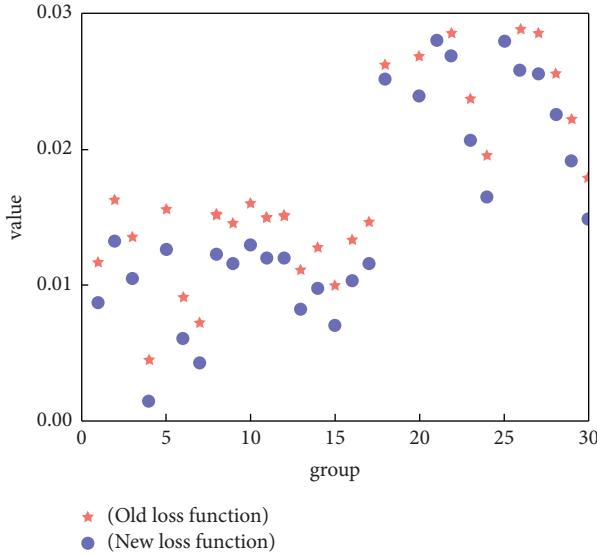


FIGURE 7: The integrity error verification of observational data using two SVM methods.

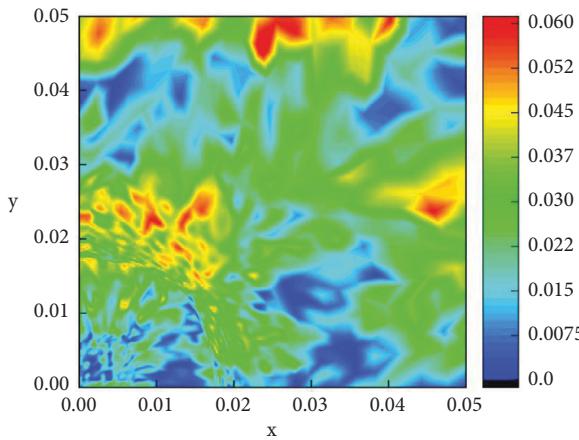


FIGURE 8: The uncertainty error distribution using the SVM method with the old loss function.

observed data. This study embeds the uncertainty of the observational data into a new loss function, which will show the distribution of the quality of the observational data through the distribution of the uncertainty. Figure 8 shows the uncertainty error distribution using the SVM method with the old loss function. The SVM method with the old loss function has a large uncertainty distribution. The regions with greater uncertainty account for a larger proportion of the total observational data. This shows that the SVM method with the old loss function is difficult to complete the task of quality control of observation data.

Figure 9 shows the uncertainty error distribution using the SVM method with the new loss function. The SVM method with the new loss function has a smaller uncertainty error in the prediction of factors related to the quality of the observed data compared to Figure 8. Only a small part of the region has an uncertainty error of more than 5%, and most of the regions have an uncertainty error of less than 2%. For

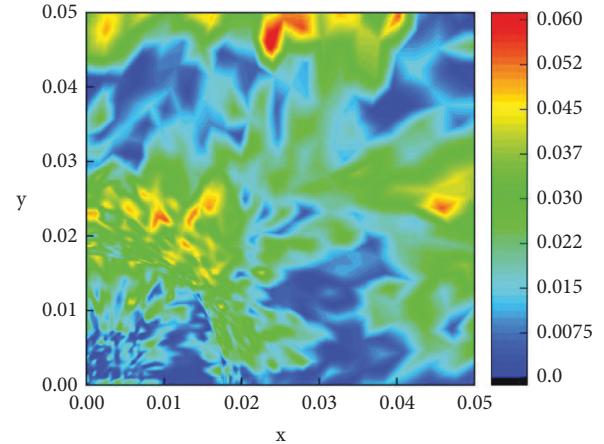


FIGURE 9: The uncertainty error distribution using SVM method with new loss function.

the SVM method with the old loss function, the uncertainty error of most regions is within 4% and more than 5% of the uncertainty error also occupies a certain proportion. For the SVM method with the new loss function, the region where the uncertainty error exceeds 5% occupies only a small proportion. This shows that the SVM method with the new loss function has better performance and accuracy in predicting the quality of the observed data than the SVM method with the old loss function.

## 5. Conclusions

The loss function of the SVM algorithm is mainly composed of a hinge function and a regularization term, and it is difficult to reflect the uncertainty distribution of the data. The traditional SVM loss function is difficult to adapt to the prediction task of observation data quality control. This study proposes an SVM with a new loss function to investigate its feasibility in predicting factors related to the quality of observed data. In this study, large and small datasets were selected to verify the accuracy of the SVM method with the new loss function. For the three factors of observed data quality, the SVM method with the new loss function has better predictive performance. The largest error reduction is 0.21% compared to the SVM method with the old loss function. For the accuracy of the observed data quality, the SVM with the new loss function can not only predict the trend of the accuracy data well, but it also fits the actual data values well. The accuracy data predicted by the SVM method with the old loss function deviate greatly from the actual data trend.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by Science and Technology Research Project of Jiangxi Provincial Department of Education: Application of multiplicative and additive mixed noise model in SAR (or Insar) data processing (no. GJJ218601).

## References

- [1] Y. Qian, "Exploration of machine algorithms based on deep learning model and feature extraction," *Mathematical Biosciences and Engineering*, vol. 18, no. 6, pp. 7602–7618, 2021.
- [2] J.-C. Huang, K.-M. Ko, M.-H. Shu, and B.-M. Hsu, "Application and comparison of several machine learning algorithms and their integration models in regression problems," *Neural Computing & Applications*, vol. 32, no. 10, pp. 5461–5469, 2020.
- [3] R. Li, Q. Zhang, J. Sun, and Y. L. T. Chen, "Smartphone pressure data: quality control and impact on atmospheric analysis," *Atmospheric Measurement Techniques*, vol. 14, no. 2, pp. 785–801, 2021.
- [4] B. Kiran, I. Sobh, V. Talpaert et al., "Deep reinforcement learning for autonomous driving: a survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 25, Article ID 3054625, 2021.
- [5] M. Rothmann and M. Porrmann, "A survey of domain-specific architectures for reinforcement learning," *IEEE Access*, vol. 10, no. 2, Article ID 13767, 2022.
- [6] Y.-F. Li, L.-Z. Guo, and Z.-H. Zhou, "Towards safe weakly supervised learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, p. 1, 2019.
- [7] P. Gkikopoulos, V. Schiavoni, and J. Spillner, "Decentralised data quality control in ground truth production for autonomic decisions," *IEEE Transactions on Parallel and Distributed Systems*, vol. 33, no. 10, pp. 2416–2427, 2022.
- [8] L. Luo, J. Lan, Y. Wang et al., "A novel early warning system (EWS) for water quality, integrating a high-frequency monitoring database with efficient data quality control Technology at a large and deep lake (lake qiandao), China," *Water*, vol. 14, no. 4, p. 602, 2022.
- [9] Y. Min, J. Y. Jeong, I.-K. Min, Y. S. Kim, J. S. Shim, and K. Do, "Enhancement of wave radar observation data quality at the socheongcho ocean research station," *Journal of Coastal Research*, vol. 85, no. 8, pp. 571–575, 2018.
- [10] A. Brotzer, F. Bernauer, K. U. Schreiber, and J. H. Wassermann, "Automated quality assessment of interferometric ring laser data," *Sensors*, vol. 21, no. 10, p. 3425, 2021.
- [11] A. Jurczyk, J. Szturc, I. Otop, and K. P. Ośrodkwa, "Quality-based combination of multi-source precipitation data," *Remote Sensing*, vol. 12, no. 11, p. 1709, 2020.
- [12] S. Chen, J. Cao, and Z. Huang, "Weighted linear loss projection twin support vector machine for pattern classification," *IEEE Access*, vol. 7, no. 6, Article ID 57360, 2019.
- [13] K. Li and Z. Lv, "Smooth twin bounded support vector machine with pinball loss," *Applied Intelligence*, vol. 51, no. 8, pp. 5489–5505, 2021.
- [14] P. Borah and D. Gupta, "Functional iterative approaches for solving support vector classification problems based on generalized Huber loss," *Neural Computing & Applications*, vol. 32, no. 13, pp. 9245–9265, 2020.
- [15] W. Zhu, Y. Song, and Y. Xiao, "A new support vector machine plus with pinball loss," *Journal of Classification*, vol. 35, no. 1, pp. 52–70, 2018.
- [16] W. Huang, H. Shao, J. Zhang, and Y. T. Y. Zhao, "Robust rescaled hinge loss twin support vector machine for imbalanced noisy classification," *IEEE Access*, vol. 7, no. 6, Article ID 65404, 2019.
- [17] J. Wu, L. Shen, and G. G. Zhang, "Efficiency evaluation with data uncertainty," *Annals of Operations Research*, vol. 3, no. 26, Article ID 04636, 2022.
- [18] D. Huang and H. S. Abdel-Khalik, "Nuclear data uncertainty propagation and modeling uncertainty impact evaluation in neutronics core simulation," *Progress in Nuclear Energy*, vol. 128, no. 11, Article ID 103443, 2020.
- [19] M. Tanveer, T. Gupta, and M. Shah, "Pinball loss twin support vector clustering," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 17, no. 2s, pp. 1–23, 2021.
- [20] Y. Ye, J. Gao, Y. Shao, and C. Y. X. Li, "Robust support vector regression with generic quadratic nonconvex  $\epsilon$ -insensitive loss," *Applied Mathematical Modelling*, vol. 82, no. 5, pp. 235–251, 2020.
- [21] H.-J. Xing and M. Ji, "Robust one-class support vector machine with rescaled hinge loss function," *Pattern Recognition*, vol. 84, no. 9, pp. 152–164, 2018.
- [22] W. Panup, W. Ratipapongton, and R. Wangkeeree, "A novel twin support vector machine with generalized pinball loss function for pattern classification," *Symmetry*, vol. 14, no. 2, p. 289, 2022.
- [23] T. Mahmood, P. Wittenberg, I. M. Zwetsloot, and H. K. L. Wang, *International Journal of Medical Informatics*, vol. 126, no. 5, pp. 156–163, 2019.
- [24] H. Yang, Q. Gao, H. Ji, and P. T. He, "Sea surface temperature data from coastal observation stations: quality control and semidiurnal characteristics," *Acta Oceanologica Sinica*, vol. 38, no. 11, pp. 31–39, 2019.
- [25] G. P. Castelão, "A machine learning approach to quality control oceanographic data," *Computers & Geosciences*, vol. 155, no. 8, Article ID 104803, 2021.
- [26] H. Liu, Y.-H. Kuo, S. Sokolovskiy et al., "A quality control procedure based on bending angle measurement uncertainty for radio occultation data assimilation in the tropical lower troposphere," *Journal of Atmospheric and Oceanic Technology*, vol. 35, no. 10, pp. 2117–2131, 2018.
- [27] W. Panup and R. Wangkeeree, "Stochastic subgradient for large-scale support vector machine using the generalized pinball loss function," *Symmetry*, vol. 13, no. 9, p. 1652, 2021.
- [28] X. Zheng, L. Zhang, and L. Yan, "CTSVM: a robust twin support vector machine with correntropy-induced loss function for binary classification problems," *Information Sciences*, vol. 559, no. 5, pp. 22–45, 2021.
- [29] M. Tanveer, A. Sharma, and P. N. Suganthan, "General twin support vector machine with pinball loss function," *Information Sciences*, vol. 494, no. 6, pp. 311–327, 2019.
- [30] H. Dong, L. Yang, and X. Wang, "Robust semi-supervised support vector machines with Laplace kernel-induced correntropy loss functions," *Applied Intelligence*, vol. 51, no. 2, pp. 819–833, 2021.
- [31] I. Razzak, K. Zafar, M. Imran, and G. Xu, "Randomized nonlinear one-class support vector machines with bounded loss function to detect outliers for large scale IoT data," *Future Generation Computer Systems*, vol. 112, no. 9, pp. 715–723, 2020.
- [32] Q. Ren and L. Yang, "A robust projection twin support vector machine with a generalized correntropy-based loss," *Applied Intelligence*, vol. 52, no. 2, pp. 2154–2170, 2022.