Information Fusion 100 (2023) 101966

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Full length article

TWC-EL: A multivariate prediction model by the fusion of three-way clustering and ensemble learning

Xunjin Wu [a], Jianming Zhan [a,*], Weiping Ding [b]

[a] School of Mathematics and Statistics, Hubei Minzu University, Enshi, Hubei, 445000, China
[b] School of Information Science and Technology, Nantong University, Nangtong 226019, China

ARTICLE INFO

Keywords:
Multivariate prediction
Three-way clustering
Ensemble learning
BPNN model
Elman neural network

ABSTRACT

Multivariate data analysis, as an important research topic in the field of machine learning, focuses on how to utilize the intrinsic connection between feature variables and target variables. However, in the face of complex multivariate prediction environments, existing single prediction models often fail to obtain ideal results. Meanwhile, existing ensemble prediction models are not always adapted to certain complex data. Moreover, the randomness in the clustering process cannot guarantee the clustering accuracy. Therefore, to improve the model's prediction accuracy and ability to adapt to complex data and reduce the impact of randomness on clustering accuracy, this paper designs a multivariate prediction model utilizing three-way clustering (TWC) and ensemble learning, which is named the TWC-EL model. First, the initial division of the sample set is realized by k-means clustering algorithm, and further the sample set is divided again via the k-means clustering algorithm to solve the problem of clustering accuracy. Then, the results of clustering twice are combined according to the difference in the number of intersection points and the distance from the samples to the center point of each cluster, and the core and fringe regions of each cluster in the initial clustering results are obtained, forming a new TWC method. Next, based on the correlation between the regions, the obtained core and fringe regions are classified into low-correlation, medium-correlation and high-correlation regions, and an ensemble prediction model is designed by combining the advantages of the Elman neural network model, the Extreme Learning Machine (ELM) model and the back propagation neural network (BPNN) model. Finally, the experimental analysis results exhibit that the constructed TWC-EL model is efficient and feasible, and points out the excellent performance compared with the existing prediction models. The validity of the TWC method and the ensemble prediction model in the proposed TWC-EL model are verified by experiments, respectively.

1. Introduction

During the progressive development of modern sciences, human beings are facing more and more challenges in real world [1]. As an effective method to explore the future trends of things, prediction has been widely used in different aspects of society. Currently, prediction is divided into two main types of prediction, including univariate prediction based on time series [2] and multivariate prediction [3,4]. Meanwhile, as human society continues to move forward, people are more concerned about certain areas, such as health levels, transportation systems, environmental quality, and so on [5,6]. To study specific problems in these areas, adopting prediction methods is indispensable in order to draw desired conclusions.

Linear prediction models, as a relatively simple class of prediction models, are better for linear problems. Recently, there is a research trend among scholars to apply such predictive models to real-world prediction issues. Chen et al. [7] predicted the cold load of energy supply by combining physics and multiple linear regression models. Huang et al. [8] used linear factors of rivers and roads to predict their landslide susceptibility by virtue of buffer analysis methods. Duan et al. [9] combined Rauch Tung Striebel smoothing with an autoregressive integrated moving average model to predict oil and gas production. Oliveira et al. [10] employed an exponential smoothing (denoted as ES) model together with an autoregressive integrated moving average model to predict world energy demand.

Although the above linear prediction models are highly interpretable, they have poor prediction performance for nonlinear problems. To address this shortcoming, neural network models, which are useful tools for coping with nonlinear prediction problems, have attracted many experts' research and achieved fruitful results. Ghose

* Corresponding author.
E-mail addresses: 1397765672@qq.com (X. Wu), zhanjianming@hotmail.com (J. Zhan), dwp9988@163.com (W. Ding).

https://doi.org/10.1016/j.inffus.2023.101966
Received 11 June 2023; Received in revised form 31 July 2023; Accepted 2 August 2023
Available online 6 August 2023

et al. [11] predicted water level fluctuation situation based on the model of BPNN. Li et al. [12] proposed an adaptive runoff prediction method by applying normalized mutual information and kernel principal component analysis to Elman neural networks. Wang et al. [13] analyzed ELM in terms of theory and generalization ability and made various improvements to enhance the prediction performance of ELM. However, in the face of complex big data, a single predictive model cannot make the predicted value and the real value achieve a good fitting effect.

In order to solve this problem and improve the fitting effect and prediction accuracy, the ensemble learning model has received extensive attention from many researchers and some related studies have been done. Chen et al. [14] introduced a new model for the short-term prediction problem of wind speed, which is composed of convolutional neural networks, long and short-term memory networks and an improved BPNN model. Additionally, the model of BPNN was optimized by genetic algorithm (denoted as GA-BPNN) [15,16], the BPNN model is optimized by particle swarm algorithm (denoted as PSO-BPNN) [17,18], the model of BPNN is optimized by artificial bee colony algorithm (denoted as ABC-BPNN) [19,20] and the model of BPNN is optimized by whale algorithm (denoted as WOA-BPNN) [21,22]. Although these ensemble learning models are able to achieve good prediction performance, these models are predicted for all samples, which makes certain samples are not adaptable to the corresponding model, thus affecting the prediction results.

In response to the above deficiencies, a new ensemble prediction model that simultaneously considers the performance of object features and the respective advantages of prediction models is proposed in this study, i.e., utilizing the merits of the Elman neural network model [12], the ELM model [13] and the BPNN model [11,23] to predict samples in the low, medium, and high regions. In view of this, how to effectively accomplish the clustering of samples will be an important task. Currently, clustering techniques are widely studied and applied in various fields. Yang et al. [24] proposed a new hybrid genetic model for clustered ensemble by combining genetic model and HGCEA algorithm. Deng et al. [25] added l_1-norm in order to obtain cleaner data matrices and introduced a graph regularized sparse non-negative matrix trifactorization (GSNMTF) to clustering. In the context of varying rows and columns, Chen et al. [26] designed a fast flexible bipartite graph model for the co-clustering method.

TWD, as a complex problem solving paradigm, was originally designed to give a reasonable semantic interpretation of the three regions of probabilistic rough sets and decision rough sets [27–29]. Since it has good problem handling ability and cognitive advantages, it has aroused the research interests of many scholars. Zhan et al. [30] designed a new TWD model by virtue of outranking relations in the context of multi-attribute decision-making (denoted as MADM). Wang et al. [31] designed a TWD model based on prospect theory to classify the alternatives. While the loss functions of these TWD methods are all subjectively given. Thus, to further compensate for the disadvantages, Jia and Liu [32] objectively derived the loss function based on the differences between the alternatives, and built a novel TWD model to multi-criteria decision-making (MCDM). Evidently, the research on TWD has achieved remarkable achievements.

Inspired by TWD, Yu [33] proposed the concept of TWC based on existing clustering methods and provided a formal representation of it. Notably, this three-way representation is capable of handling both traditional hard and soft clustering tasks. In recent years, TWC has also achieved fruitful research achievements. Ali et al. [34] used blurring and sharpening operations to divide each cluster into three regions. Zhang et al. [35] fused TWD with clustering to establish an adaptive TWC based on feature distribution. Zhang [36] proposed a three-way c-means algorithm based on three-way weights. Due to the stochasticity, the above TWC methods cannot guarantee that a good clustering accuracy has been achieved by performing one clustering. Therefore, in this paper, the method of clustering twice is used to improve the clustering accuracy and fuses the results of the two clusters according to the number of intersections and the principle of shortest distance (the closer an object is to the center of a cluster, the more likely it is to belong to that cluster) to alleviate the error caused by randomness. Meanwhile, TWC has been widely used in overlapping clustering [37], rough sets [38], fuzzy sets [39], mathematical morphology [40], evidence theory [41] and so on.

In summary, considering the weakness of a single neural network model in actual prediction issues and the cognitive complexity, this paper constructs a novel prediction model in light of TWC and ensemble learning by combining the advantages of TWC and ensemble learning.

Based on the preceding analysis, this study presents the following research motivations:

(1) As the prediction environments become more complex, the existing single prediction models often suffer from significant performance limitations [42–45]. For the purpose of effectively enhancing the prediction accuracy, it is of great necessity to consider different prediction models comprehensively.

(2) The selection of clustering numbers and clustering centers, as the core of clustering research, has been much attention and research. The existing clustering numbers and clustering centers are generally chosen randomly or given subjectively [36,46], which makes the classification accuracy of clustering algorithms impossible to be guaranteed. Therefore, exploring a more accurate clustering algorithm will be an issue worth exploring.

(3) Ensemble thinking is an effective strategy to improve prediction accuracy via combining multiple weak predictors to enhance the learning ability of the model [14,47]. It will be also a worthy research problem to adopt an effective ensemble strategy.

In addition, we can point out the following innovations:

(1) We propose a multivariate prediction system by virtue of TWC and ensemble learning to improve the prediction performance. In particular, it enables the prediction system to combine the advantages of different predictors by combining multiple weak predictors.

(2) To improve the clustering accuracy, this paper adopts the k-means clustering algorithm for clustering twice, and gets a new clustering result by the different numbers of clustering intersection points in the clustering twice results. Particularly, this strategy integrates the feature of the data itself and completes the effective mining of data information.

(3) According to the correlation of samples within different regions, a hybrid prediction strategy is constructed. In other words, different prediction strategies are adopted to predict different regions. The prediction approach not only performs well in terms of accuracy but also demonstrates excellent generality.

The rest of the present paper is structured below: In Section 2, TWC, BPNN model, ELman neural network model and ELM model are reviewed. Section 3 outlines a new multivariate prediction model that leverages TWC together with ensemble learning. Subsequently, Sections 4 and 5 validate the viability, effectiveness and generality of the designed TWC-EL method through experimental analysis and evaluation. In the end, Section 6 concludes the paper.

2. Preliminaries

The basics of TWC [33], BPNN models [23], Elman neural network models [12] and ELM models [13] are briefly reviewed in this part.

2.1. TWC

Let the universe $U = \{y_1, y_2, \dots, y_n, \dots, y_N\}$ be a finite set, where $y_n = \{y_n^1, y_n^2, \dots, y_n^m, \dots, y_n^M\}$ is an object that has M attributes, $m \in \{1, 2, \dots, M\}$, $n \in \{1, 2, \dots, N\}$. Suppose that the clustering divides the N objects into K clusters, labeled as $\mathbf{C} = \{C_1, C_2, \dots, C_k, \dots, C_K\}$, here $C_k = \{y_k^1, y_k^2, \dots, y_k^i, \dots, y_k^{|C_k|}\}$, abbreviated as C without ambiguous. Usually, a cluster can be denoted by a single set. To illustrate the

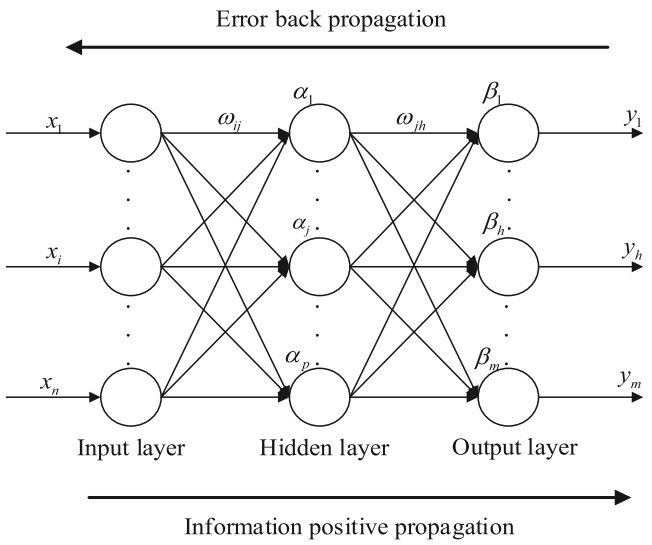


Fig. 1. The topology shape of the model of BPNN.

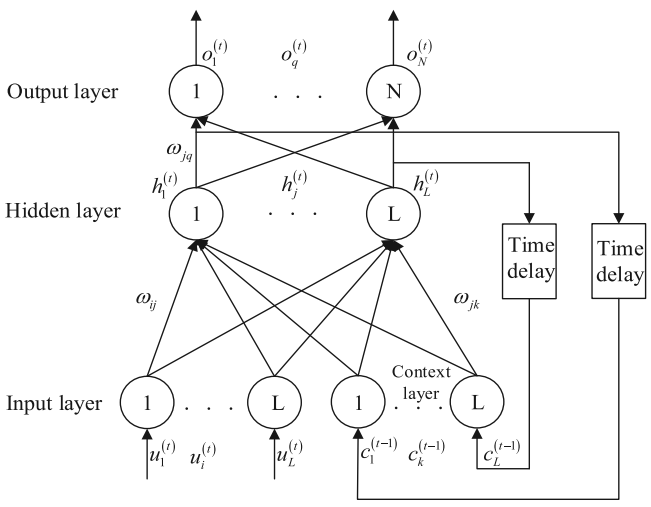


Fig. 2. The topology shape of Elman neural network.

uncertainty of the object in the division process, a cluster is expressed via two sets:

$$C = (Co(C), Fr(C)), \tag{2.1}$$

where $Fr(C) \subseteq U$, $Co(C) \subseteq U$ and $Tr(C) = U - Co(C) - Fr(C)$.

Note that $Co(C)$, $Fr(C)$ and $Tr(C)$ denote the core region, the fringe region, and the trivial region of a cluster respectively. When an object $x \in Co(C)$, then x certainly belongs to the cluster C. When $x \in Fr(C)$, it may belong to the cluster C, and when $x \in Tr(C)$, it certainly does not belong to the cluster C. The subsets possess the subsequent characteristics:

$$\begin{cases} U = Co(C) \cup Fr(C) \cup Tr(C) \\ Co(C) \cap Fr(C) = \varnothing \\ Fr(C) \cap Tr(C) = \varnothing \\ Co(C) \cap Tr(C) = \varnothing \end{cases} \tag{2.2}$$

Notably, each cluster cannot be empty and each object must be divided into at least one cluster. For the cluster **C**, each cluster is represented by the following:

$$\mathbf{C} = \{(Co(C_1), Fr(C_1)), (Co(C_2), Fr(C_2)), \ldots, (Co(C_k), Fr(C_k)), \ldots, (Co(C_K), Fr(C_K))\}. \tag{2.3}$$

2.2. BPNN models

The BPNN model, which is a multiple layer perception network achieved via optimizing back propagation algorithm, exhibits two primary characteristics: backward propagation of errors together with forward propagation of information. Meanwhile, it acts as a multiple layer feed-forward neural network, which involves the correction of weights and thresholds. The specific update process can be referred to [23]. Fig. 1 illustrates the topology shape of BPNN model.

For Fig. 1, the amount of neurons in the input, hidden, and output layer separately are n, p, m, respectively. Simultaneously, in the neural network, $x_i (i = 1, 2, \ldots, n)$ denotes the input while $y_h (h = 1, 2, \ldots, m)$ denotes the actual output; $\alpha_j (j = 1, 2, \ldots, p)$ and $\beta_h (h = 1, 2, \ldots, m)$ are the thresholds; $\omega_{ij} (i = 1, 2, \ldots, n; j = 1, 2, \ldots, p)$ and $\omega_{jh} (j = 1, 2, \ldots, p; h = 1, 2, \ldots, m)$ are the two connection weights.

2.3. Elman models

The Elman neural network model is a feedback model wherein the output of the hidden layer is recurrently associated to the input of the same layer via the context layer, which stores delayed information. The neural network comprises four layers, specifically input layer, hidden layer, context layer together with output layer. The output layer and input layer act as a information delivery and a linear weighting respectively. Meanwhile, a non-linear function is utilized in hidden layer: $f(t) = \frac{1}{1+e^{-t}}$. The context layer unit is utilized to memory the output of the hidden layer unit from the previous moment and feed it back to the input of network, which is considered as a delay operator. For a detailed understanding of the computational and learning process of the Elman neural network, please refer to [12]. Fig. 2 illustrates the topology shape of Elman neural network.

2.4. ELM models

The ELM model is a feed forward neural network known for its fast training and strong generalization. Unlike the BPNN model, the ELM model does not need to adjust the weights and thresholds during training, and only requires setting the amount of nodes in the hidden layer to get the unique optimal solution. The learning process of ELM can be referred to [13]. Moreover, the architecture of ELM has similarities with BPNN.

3. A new prediction model by the fusion of TWC and ensemble learning

Scientific judgments about the future tendency or movement of an object on the basis of available data are the embodiment of forecasting. For the study of prediction methods, many scholars have achieved some stages of results. However, the trustworthiness and precision of forecasted results are not guaranteed. For the purpose of raising the robustness of the prediction model, this article uses an ensemble learning approach for prediction. Meanwhile, considering that objects are similar to each other, we will combine the proposed method with TWC. In particular, Section 3.1 will provide a brief overview of the k-means clustering algorithm [46]. Section 3.2 presents a novel prediction model and algorithm that take into account the characteristics of the correlation between core regions together with the correlation between fringe regions and core regions.

3.1. k-means clustering

Clustering is an approach that groups objects with similar characteristics into a cluster, forming distinct classes. Each class is called a cluster. Moreover, all clusters are disjoint. Among clustering methods,

k-means algorithm has been widely used as a representative method. It obtains the optimal division by minimizing the squared error. Algorithm 1 outlines the specific process, while the following paragraph presents a detailed definition.

Definition 3.1. [47] Given a sample set $D = \{x_1, x_2, \dots, x_n\}$, every sample possesses m attributes, that is, $x_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$ $(i = 1, 2, \dots, n)$. The distance between samples $x_j \in D$ and $x_h \in D$ is defined as follows:

$$d(x_j, x_h) = \sum_{k=1}^{m} (x_{jk} - x_{hk})^2. \tag{3.1}$$

Then, the total squared error of the division result is presented as follows:

$$E = \sum_{i=1}^{K} \sum_{x \in C_i} d(x, \mu_i), \tag{3.2}$$

where $\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$ denotes the center of ith cluster. The target function E is used to measure the similarity among samples belonging to the same cluster.

Remark 3.1. Minimizing E is the main objective of k-means algorithm. Because all possible cluster divisions need to be considered, finding the optimal solution is not an easy task. For obtaining the optimal solution, we use the iterative approach, inspired by the greedy strategy. Hence, the clustering results can be acquired. .

Example 3.1. Given 6 objects and each object has two attributes, namely, $x_1 = (0, 0), x_2 = (2, 1), x_3 = (4, 5), x_4 = (6, 6), x_5 = (7, 8), x_6 = (10, 3)$, let $K = 2$.

(1) Two objects x_1, x_2 are randomly selected from $x_1, x_2, x_3, x_4, x_5, x_6$ as the initial mean vectors, i.e., $P_1^{(0)} = (0, 0), P_2^{(0)} = (2, 1)$.

(2) Compute distance between each sample and initial mean vectors, the initial class: $C_1^{(0)} = \{x_1\}, C_2^{(0)} = \{x_2\}$.

For x_3, $d(x_3, P_1^{(0)}) = 6.40$, $d(x_3, P_2^{(0)}) = 4.47$, then $x_3 \in C_2^{(0)}$.
For x_4, $d(x_4, P_1^{(0)}) = 8.49$, $d(x_4, P_2^{(0)}) = 6.40$, then $x_4 \in C_2^{(0)}$.
For x_5, $d(x_5, P_1^{(0)}) = 10.63$, $d(x_5, P_2^{(0)}) = 8.60$, then $x_5 \in C_2^{(0)}$.
For x_6, $d(x_6, P_1^{(0)}) = 10.44$, $d(x_6, P_2^{(0)}) = 8.25$, then $x_6 \in C_2^{(0)}$.

(3) Obtain the new class $C_1^{(1)} = \{x_1\}, C_2^{(1)} = \{x_2, x_3, x_4, x_5, x_6\}$ and update the mean vector: $P_1^{(1)} = (0, 0), P_2^{(1)} = (5.8, 4.6)$.

(4) Compute distance between each sample and updated mean vectors.

For x_2, $d(x_2, P_1^{(1)}) = 2.24$, $d(x_2, P_2^{(1)}) = 5.23$, then $x_2 \in C_1^{(1)}$.
For x_3, $d(x_3, P_1^{(1)}) = 6.40$, $d(x_3, P_2^{(1)}) = 1.84$, then $x_3 \in C_2^{(1)}$.
For x_4, $d(x_4, P_1^{(1)}) = 8.49$, $d(x_4, P_2^{(1)}) = 1.41$, then $x_4 \in C_2^{(1)}$.
For x_5, $d(x_5, P_1^{(1)}) = 10.63$, $d(x_5, P_2^{(1)}) = 3.60$, then $x_5 \in C_2^{(1)}$.
For x_6, $d(x_6, P_1^{(1)}) = 10.44$, $d(x_6, P_2^{(1)}) = 4.49$, then $x_6 \in C_2^{(1)}$.

(5) Obtain the new class $C_1^{(2)} = \{x_1, x_2\}, C_2^{(2)} = \{x_3, x_4, x_5, x_6\}$ and update the mean vector: $P_1^{(2)} = (1, 0.5), P_2^{(2)} = (6.75, 5.5)$.

(6) Compute distance between each sample and updated mean vectors again.

For x_1, $d(x_1, P_1^{(2)}) = 1.11$, $d(x_1, P_2^{(2)}) = 8.70$, then $x_1 \in C_1^{(2)}$.
For x_2, $d(x_2, P_1^{(2)}) = 1.11$, $d(x_2, P_2^{(2)}) = 6.54$, then $x_2 \in C_1^{(2)}$.
For x_3, $d(x_3, P_1^{(2)}) = 5.41$, $d(x_3, P_2^{(2)}) = 2.80$, then $x_3 \in C_2^{(2)}$.
For x_4, $d(x_4, P_1^{(2)}) = 7.43$, $d(x_4, P_2^{(2)}) = 0.90$, then $x_4 \in C_2^{(2)}$.
For x_5, $d(x_5, P_1^{(2)}) = 9.60$, $d(x_5, P_2^{(2)}) = 2.51$, then $x_5 \in C_2^{(2)}$.
For x_6, $d(x_6, P_1^{(2)}) = 9.34$, $d(x_6, P_2^{(2)}) = 4.10$, then $x_6 \in C_2^{(2)}$.

Therefore, the final clustering outcomes are $C_1^* = \{x_1, x_2\}$ and $C_2^* = \{x_3, x_4, x_5, x_6\}$.

Remark 3.2. The k-means clustering algorithm is a converging process which divides the universe D into k disjoint regions, where each object belongs to only one cluster. Meanwhile, compared with other clustering algorithms, k-means clustering algorithm has the characteristics of low time cost and fast convergence speed. At the same time, k-means clustering algorithm has simple principle and strong interpretability. Therefore, it is conducive to human cognition.

3.2. A novel designed prediction model

Despite the availability of numerous forecasting methods, their limited predictive performance and the growing complication of the data make it difficult to achieve the desired results with a single forecasting method. In this regard, the ensemble learning approach can improve the accuracy of prediction, i.e., combining some weak learners together to make prediction. In addition, taking into account the similarity between objects, clustering methods are applied to classify. Moreover, each cluster of the clustering result is trisected and then predicted according to the degree of correlation between regions. Next, we will propose a new model by means of TWC and ensemble learning. Furthermore, because of the different correlations between the regions, we adopt different prediction models for the objects in the different regions.

Let $S = (D, U)$ be a given predictive knowledge representation system, where $D = \{x_1, x_2, \dots, x_n\}$ indicates the set with n objects together with $U = \{u_1, u_2, \dots, u_m\}$ denotes m features. D is normalized by using the following [0,1] normalization expression:

$$y_{ij} = \frac{x_{ij} - \min_{1 \le h \le n} x_{hj}}{\max_{1 \le h \le n} x_{hj} - \min_{1 \le h \le n} x_{hj}}, \quad j = 1, 2, \dots, m; i = 1, 2, \dots, n. \tag{3.3}$$

Remark 3.3. In light of the large differences in the values of some data or different units, errors will occur in subsequent discussions. To address this issue, the idea of normalization can avoid errors in the discussion process.

The selection of clustering centers is random, which makes the clustering results not precise. Therefore, the aim of another clustering is to acquire the final clustering outcome by combining clustering twice effects, i.e., the normalized D is clustered 2 times through adopting Algorithm 1. This makes the acquired clustering result relatively accurate and more convincing. The result obtained will be written as follows:

$$\mathbf{C_1} = \{C_{11}, C_{12}, \dots, C_{1K}\}, \tag{3.4}$$

$$\mathbf{C_2} = \{C_{21}, C_{22}, \dots, C_{2K}\}. \tag{3.5}$$

According to the clustering results, $C_{1i} (i = 1, 2, \dots, K)$ is intersected with each cluster in $\mathbf{C_2}$, respectively. There are two scenarios that may occur, as follows:

(1) If the number of intersection points is not the same, the cluster with the highest number of intersection points in $C_{1i} (i = 1, 2, \dots, K)$ and $\mathbf{C_2}$ will be merged. Moreover, the intersection points are used as the core region of $C_{1i} (i = 1, 2, \dots, K)$ and the other points are viewed as the fringe region of $C_{1i} (i = 1, 2, \dots, K)$.

(2) If $\mathbf{C_2}$ has two or more clusters with the same maximum number of intersection points with $C_{1i} (i = 1, 2, \dots, K)$, calculating the distance between cluster center of $C_{1i} (i = 1, 2, \dots, K)$ and cluster centers of its corresponding $\mathbf{C_2}$ with the same number of intersection points. Then, taking the cluster nearest to $C_{1i} (i = 1, 2, \dots, K)$ and combine it with $C_{1i} (i = 1, 2, \dots, K)$. Moreover, the intersection points are used as the core region of $C_{1i} (i = 1, 2, \dots, K)$ and the other points are viewed as the fringe region of $C_{1i} (i = 1, 2, \dots, K)$.

Therefore, the new clustering result is noted as follows:

$$\mathbf{C'} = \{(Co(C_1'), Fr(C_1')), (Co(C_2'), Fr(C_2')), \dots, (Co(C_k'), Fr(C_k')), \dots, (Co(C_K'), Fr(C_K'))\}. \tag{3.6}$$

Remark 3.4. Relying on the notion of clustering, under the same number of intersection points, the reason for merging the class with the highest number of intersections or the class with the closest center

Algorithm 1: The k-means clustering algorithm

Input: $D = \{x_1, x_2, \cdots, x_n\}$, K
Output: Cluster result
begin
  K objects are randomly selected from D as the initial mean vectors: $\{\mu_1, \mu_2, \cdots, \mu_K\}$
  repeat
  Let $C_i = \emptyset (1 \le i \le K)$
  for $j = 1, 2, \cdots, n$ **do**
    calculate the distance between sample x_j and initial mean vectors μ_i $(1 \le i \le K)$: $d_{ji} = \|x_j - \mu_i\|_2$;
    determine the labels of the clusters corresponding to the object x_j according to the closest distance principle: $\lambda_j = \arg \min_{i \in \{1,2,\cdots,K\}} d_{ji}$;
    categorize the object x_j into the cluster $C_{\lambda_j} = C_{\lambda_j} \bigcup \{x_j\}$;
  end
  for $i \in \{1, 2, \cdots, K\}$ **do**
    calculate new mean vector: $\mu_i' = \frac{1}{|C_i|} \sum_{x \in C_i} x$;
    if $\mu_i' \neq \mu_i$ **then**
      the recent mean vector μ_i is updated to μ_i'
      else
        the recent mean vector is not updated
      end
    end
  end
  until all recent mean vectors are not updated
  return: Cluster result $\{C_1, C_2, \cdots, C_K\}$
end

distance between the two classes is that the objects in these two classes are more similar than other clusters. Meanwhile, the intersection points are most likely to be the core region of the merged class. Therefore, intersection points are taken as the core region of the class.

In what follows, since the points in the fringe region of every cluster may belong to the core region or fringe region of other clusters, an evaluation function, i.e., Eq. (3.1), is defined to divide the points in the fringe region, namely $\{Fr(C_1'), Fr(C_2'), \ldots, Fr(C_k'), \ldots, Fr(C_K')\}$.

Before dividing, the centroid of $\{Co(C_1'), Co(C_2'), \ldots, Co(C_k'), \ldots, Co(C_K')\}$ needs to be calculated. First, calculating the sum of the distances between an object in each core region and the other objects in cluster. Then, it can be inferred that the object with the smallest sum of distances serves as the centroid of the core region, since it is the most probable center of the corresponding core region. Let the length of $Co(C_j')(j = 1, 2, \ldots, K)$ be b. Moreover, the objects in $Co(C_j')$ are denoted as $\{x_1, x_2, \ldots, x_b\}$ based on an order relation in $Co(C_j')$. The formula is given as follows:

$$x_i = \min_{x_i \in Co(C_j')} \sum_{h=1}^{b} \sqrt{\sum_{k=1}^{m} (x_{ik} - x_{hk})^2}, \quad i \in \{1, 2, \ldots, n\}. \tag{3.7}$$

Depending on whether the fringe region of each cluster have intersections, the following two cases arise:

Case 1: If there is an intersection point in the fringe domain of any two clusters, then the intersection point is classified as fringe region of its corresponding cluster.

Case 2: In Case 1, the distance between the points outside the intersection in fringe region of each cluster and cluster centers of clusters can be calculated. It is worth noting that fringe regions of each cluster intersect with core regions of other clusters. Therefore, it is necessary to eliminate these intersection points in these fringe regions to ensure that the core region of each cluster has no intersection points with the fringe region. Comparing the distance between point and cluster center of corresponding cluster and cluster center of other cluster, one of two situations will happen:

(1) The distance in the corresponding cluster is minimum. Then, this point is classified as fringe region of cluster;

(2) The distance is minimum in one of the other clusters. Then, computing the maximum radius of cluster, namely, the farthest distance from a point in core region of cluster to center of cluster is calculated. The corresponding equation is listed as follows:

$$\delta_j = \max_{x_q \in Co(C_j')} \sqrt{\sum_{k=1}^{m} (x_{qk} - x_{ik})^2}, \quad j = 1, 2, \ldots, K; q \in \{1, 2, \ldots, n\}, \tag{3.8}$$

where x_i is the cluster center of C_j'.

When this distance is less than or equal to the maximum radius of the corresponding cluster, the point is classified as core region of the corresponding cluster, otherwise it is classified as the fringe region of the corresponding cluster. The final result is expressed as follows:

$$\mathbf{Co(C)} = \{Co(C_1), Co(C_2), \ldots, Co(C_K)\}, \tag{3.9}$$

$$\mathbf{Fr(C)} = \{Fr(C_1), Fr(C_2), \ldots, Fr(C_K)\}. \tag{3.10}$$

Further, **Co(C)** will be classified. During the process, objects in the core region are treated as random variables, which possible values are the corresponding eigenvalues. In the context of measuring the correlation between two stochastic variables, the improved covariance expression is used.

$$C(x_r, x_q) = |\sum_{h=1}^{m} \sum_{k=1}^{m} (\|x_{rk}\| - \|\overline{x_r}\|)(\|x_{qh}\| - \|\overline{x_q}\|)|, \quad r, q \in \{1, 2, \ldots, n\}, \tag{3.11}$$

where $\forall x_r \in Co(C_i), \forall x_q \in Co(C_j)$, $i, j \in \{1, 2, \ldots, K\}, i \neq j$. Moreover, $\overline{x_i}$ represents the average value of the object x_i characteristic values, $\| * \|$ are denoted as the norm length of $*$.

Definition 3.2. Let the length of $Co(C_i)$ and $Co(C_j)$ be a and c, respectively. Moreover, the objects in $Co(C_i)$ and $Co(C_j)$ are denoted as $\{x_1, x_2, \ldots, x_a\}$ and $\{x_1, x_2, \ldots, x_c\}$ based on an order in $Co(C_j)$ and $Co(C_j)$, respectively. The overall correlation between $Co(C_i)$ and $Co(C_j)(j \neq i)$ is introduced below:

$$C(Co(C_i), Co(C_j)) = \sum_{r=1}^{a} \sum_{q=1}^{c} C(x_r, x_q), \quad i, j \in \{1, 2, \ldots, K\}, i \neq j. \tag{3.12}$$

Based on Eq. (3.12), we merge the two largest and two smallest clusters of $C(Co(C_i), Co(C_j))$. If the two clusters of the maximum merge and the minimum merge have the same part, the maximum merge is taken. Meanwhile, the two clusters with the largest combination are removed, and the two clusters with the smallest overall correlation among the remaining clusters are merged. Then, we repeat constantly until the merge is complete. Notably, the obtained result is denoted as $NCo(1)$ and $NCo(2)$.

Remark 3.5. Generally speaking, the correlation between fringe regions is low and the correlation between core regions is high. However, in the core regions, the classification error due to the randomness in the clustering process makes the correlation between some core regions are not high, so it is necessary to carry out secondary division of the core region to alleviate the classification error. Moreover, the secondary division can be executed to divide all the objects in the core region into two regions with different characteristics, so that different prediction models can be adopted to predict for different characteristics to improve the prediction accuracy.

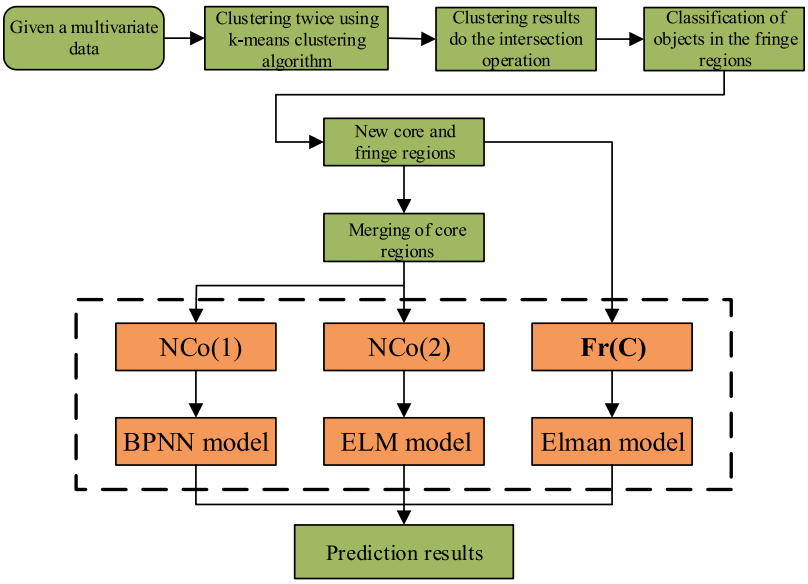


Fig. 3. The flow of TWC-EL model.

Finally, according to the degree of correlations of $NCo(1)$, $NCo(2)$ and $\mathbf{Fr}(\mathbf{C})$, the BPNN model, the ELM model together with the Elman neural network model are utilized for predicting, respectively.

The detailed process of the TWC-EL model is presented in a flow chart (Fig. 3) as well as Algorithm 2.

Remark 3.6. The proposed prediction model in this ensemble framework of this paper simultaneously considers the characteristics of the objects and the respective advantages of the prediction model and the applicable data characteristics, so that the regions with different characteristics and the prediction model achieve the highest degree of matching, which is conducive to the improvement of the prediction accuracy and prediction performance. In addition, the impact of classification errors in the clustering process on the prediction performance can also be reduced.

Remark 3.7. Let $|D| = n$ and $|U| = m$. The time complexity of every step is calculated: Step 1 normalizes data, leading to a computational complexity of $O(mn)$. Step 2 uses k-means clustering algorithm to clustering, leading to a computational complexity of $O(mk)$. Step 3 uses intersection operations based on clustering results, leading to a computational complexity of $O(k^2)$. Step 4 classifies objects in the fringe region, leading to a computational complexity of $O(k^2)$. Step 5 merges the core regions and computational complexity, were it is $O(k^2 - 1)$. Finally, we use BPNN, ELM, Elman neural network to predict. Let the input layer, hidden layer together with output layer nodes in the BPNN model, the ELM model, the Elman neural network model be n, p, and m, respectively, and context layer nodes in the Elman neural network model be L. Then the time complexity is $O(n(np+pm)) = O(np)$, $O(n(np + pm)) = O(np)$, and $O(n(np + Lp + pm)) = O(np)$, respectively. Therefore, this time complexity of the TWC-EL model can be computed as $O(mn)$.

Remark 3.8. For the objects in the low, medium, and high regions, 30% of the objects in each region are randomly taken as the test set. Then, three prediction models are used to predict the test set in the three regions. Next, the prediction results of the three prediction models are put together according to the prediction order as the final prediction result, which ensures the reasonableness and credibility of the prediction results.

4. Data set analysis and evaluations

This section uses QASR aquatic toxicity data[1] to illustrate the prediction capability of the presented model. Particularly, the advantages of the presented TWC-EL model are summarized.

4.1. Data set descriptions and evaluations

With the development of modern sciences, scientists explore more and more profoundly the substances that are closely related to human beings, for example, studies on biological toxicity, housing prices, safety of drinking water, etc. Among them, the study of biological toxicity is a popular topic of exploration which may endanger human life. Therefore, by adopting QASR aquatic toxicity data set, the proposed TWC-EL model is evaluated. This data set consists of 8 conditional attributes and 1 decision attribute, where 8 conditional attributes denote TPSA(Tot) (Molecular properties), SAacc (Molecular properties), H-050 (Atom-centered fragments), MLOGP (Molecular properties), RDCHI (Connectivity indices), GATS1p (2D autocorrelations), nN (Constitutional indices) and C-040 (Atom-centered fragments), respectively. The decision attribute is quantitative response, LC50. In addition, to illustrate the practicability and usefulness of the suggested TWC-EL model, 70% of data set is randomly selected as the training set and the rest as the test set. Notably, some evaluation indicators are used to assess the prediction results, including the mean square error (MSE), the root mean square error (RMSE), the sum of squared errors (SSE) together with the mean absolute error (MAE), respectively [47].

4.2. Comparative analysis

To validate the predictive power of the TWC-EL model, eight known prediction models will be compared with it. In these models, the model of ELM [13], the model of Elman neural network [12] together with the model of BPNN [23] are three classical prediction models. the model of ABC-BPNN [19], the model of GA-BPNN [15], the model of PSO-BPNN [17], the model of WOA-BPNN [21] together with ECP-TWD model [47] are the more advanced ensemble prediction models. Meanwhile, evaluation metrics MSE, RMSE, SSE and MAE are used to evaluate the capability of the TWC-EL model.

[1] https://archive.ics.uci.edu/ml/datasets

Algorithm 2: Algorithm of the TWC-EL method

Input: $S = (D, U)$, K.
Output: The TWC-EL method.

begin
  for $i = 1, 2, \cdots, n$ **do**
    normalization x_i by Eq. (3.3).
  end
  compute clustering results by Algorithm 1.
  for $i = 1, 2, \cdots, K$ **do**
    for $j = 1, 2, \cdots, K$ **do**
      judgment the number of intersection points between C_{2j} and C_{1i}.
    end
    if *the number of intersection points is not the same* **then**
      take the cluster with the most intersection points and merge it with C_{1i};
      else
        take the cluster center of C_{1i} and merge it with the closest cluster center of C_{2j} that has the same number of intersection points
      end
    end
  end
  for $i = 1, 2, \cdots, K$ **do**
    calculate the center point of $Co(C_i^{'})$.
  end
  for $i = 1, 2, \cdots, K$ **do**
    for $j = i + 1, i + 2, \cdots, K$ **do**
      find the intersection point of $Fr(C_i^{'})$ with $Fr(C_j^{'})$.
      if *intersection point exists* **then**
        the intersection point belongs to all fringe domains containing the intersection point
      end
    end
  end
  for $i = 1, 2, \cdots, K$ **do**
    eliminate the intersection points of $Fr(C_i^{'})$.
    in $Fr(C_i^{'})$, remove the intersection point of $Fr(C_i^{'})$ and each core region in Eq. (3.6)
  end
  for $i = 1, 2, \cdots, K$ **do**
    according to Eqs. (3.8) and (3.9), calculate the distance from the points other than the intersection in $Fr(C_i^{'})$ to the cluster center of each cluster and the maximum radius of $Co(C_i^{'})$.
  end
  for $i = 1, 2, \cdots, K$ **do**
    for $j = i + 1, i + 2, \cdots, K$ **do**
      compute the overall correlation between $Co(C_i^{'})$ and $Co(C_j^{'})$ by Eq. (3.12).
    end
  end
  merge the two largest and two smallest clusters of $C(Co(C_i), Co(C_j))(i, j \in \{1, 2, \cdots, K\}, i \neq j)$ and eliminate
  until merge completed
  for $NCo(1)$, $NCo(2)$, ***Fr(C)*** **do**
    the the BPNN model is used to predict $NCo(1)$.
    the the ELM model is used to predict $NCo(2)$.
    the the Elman neural network model is used to predict **Fr(C)**.
  end
  return: the prediction outcomes via the TWC-EL model.
end

In general, the values of these four evaluation indicators show an inverse ratio to the prediction accuracy. Relying on QASR aquatic toxicity data set, the prediction outcomes of eight models are displayed in Table 1, from which some conclusions can be drawn:

Table 1
Evaluation metrics for different models based on QASR aquatic toxicity data set.

Models	MSE	SSE($\times 10^3$)	MAE	RMSE
BPNN [23]	2.5116	4.1189	1.1915	1.5848
ELM [13]	2.4590	4.0328	1.1295	1.5681
Elman [12]	1.5928	2.6123	0.9704	1.2621
ABC-BPNN [19]	1.3879	0.7578	0.8847	1.1781
GA-BPNN [15]	1.3339	0.5522	0.8980	1.1549
PSO-BPNN [17]	1.4702	0.6087	0.9260	1.2125
WOA-BPNN [21]	0.9065	0.3753	0.7032	0.9521
ECP-TWD [47]	0.6805	0.7671	0.6040	0.8250
TWC-EL	**0.0707**	**0.1669**	**0.1972**	**0.2461**

Table 2
Comparative analysis of different prediction models.

Model	Classification capability	Prediction performance	Cognitive advantages
BPNN [23]	×	Low	×
ELM [13]	×	Low	×
Elman [12]	×	Low	×
ABC-BPNN [19]	×	High	×
GA-BPNN [15]	×	High	×
PSO-BPNN [17]	×	High	×
WOA-BPNN [21]	×	High	×
ECP-TWD [47]	√	High	√
TWC-EL	√	Higher	√

(1) Clearly, the TWC-EL model outperforms other models. Meanwhile, each indicator value is also the best among all models, where MSE, SSE, MAE and RMSE are 0.0707, 0.1669×10^3, 0.1972, 0.2461, respectively.

(2) The state-of-the-art prediction models have better predictive performance than classical prediction models, which suggests that the combined models can enhance the accuracy of prediction. For instance, MSE, SSE, MAE together with RMSE of the BPNN model are 2.5116, 4.1189×10^3, 1.1915, 1.5848, respectively; MSE, SSE, MAE together with RMSE of the TWC-EL model are 0.0707, 0.1669×10^3, 0.1972, 0.2461, respectively. Then, by comparing the evaluation indicators results of these two models, the TWC-EL model significantly reduces the error. This indicates that the presented model illustrates its feasibility to some extent.

(3) The forecasting accuracy of the TWC-EL model is remarkably higher compared to the state-of-the-art models. For example, four assessment indicator values of ABC-BPNN model are 1.3879, 0.7578×10^3, 0.8847, 1.1781, respectively. It can be found that the values of these evaluation indicator values are greater than the values of each evaluation indicator of the presented model. That is to say, the presented TWC-EL model has better predictive capability. Similarly, Similarly, compared with other state-of-the-art models, the proposed TWC-EL model exhibits the best predictive performance, which fully proves its effectiveness.

To clearly explain the prediction accuracy of all types of models, according to QASR aquatic toxicity data set, the fit of the actual and predicted values of various models will be shown in Figs. 4 and 5. Among the all models, the BPNN model is the worst effect. Furthermore, it is evident that the fitting effect of the TWC-EL model is significantly superior to others, which confirms effectiveness and superiority of the TWC-EL model.

4.3. Discussion

In this section, the advantages of the model in qualitative perspective will be elaborated. Meanwhile, we also present an analytical comparison of the presented model with the subsistent models in Table 2.

Based on Table 2, the advantages of the TWC-EL model can be concluded as follows:

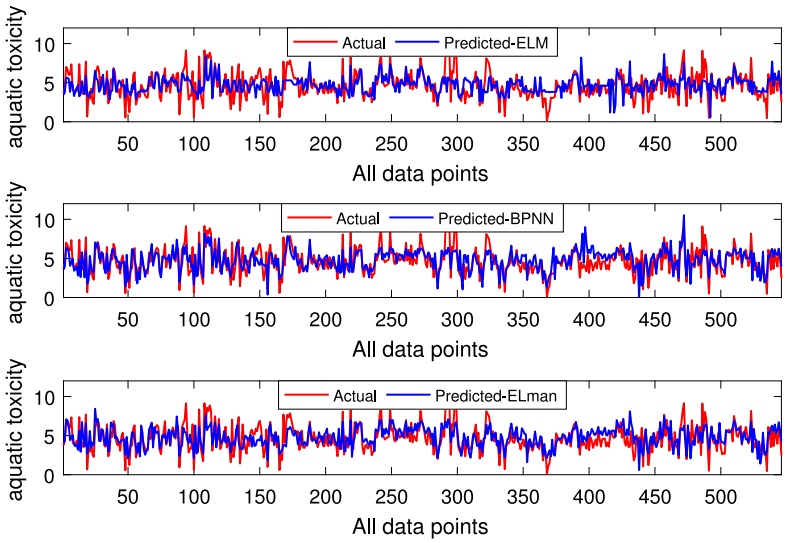


Fig. 4. Prediction results of the classical prediction models.

(a) The BPNN model [23], the ELM model [13], the Elman neural network model [12] together with the TWC-EL model can all handle nonlinear prediction problems. Nevertheless, the prediction performance of the BPNN model, the ELM model together with the Elman neural network model are limited. In this article, for the degree of relevance of objects in the same data set, the proposed TWC-EL model is used to predict objects of the same degree, which effectively improves the prediction performance.

(b) Despite the PSO-BPNN model [17], the WOA-BPNN model [21], the GA-BPNN model [15] together with the ABC-BPNN model [19] combine two weak learners to form new prediction models, the prediction results are still not very accurate. In contrast, the proposed TWC-EL model both enhances prediction accuracy and strengthens cognitive advantages.

(c) Both the ECP-TWD model [47] and the TWC-EL model have classification capability and cognitive merits. However, the ECP-TWD model adopts multiple linear regression model in the prediction process, which is only applicable to linear prediction problems and limited for dealing with complex nonlinear data problems. Therefore, the TWC-EL model solves this problem well and enhances prediction accuracy and capacity.

In summary, the TWC-EL model firstly clusters similar objects together by means of clustering. Then, all objects in the clusters are classified into three categories based on TWD, that is, the core region, fringe region together with trivial region. Next, all objects are classified into three categories according to the degree of correlation between regions and predicted by using different models. Finally, the TWC-EL model offers an innovative way of thinking about the research problem of time series forecasting.

5. Experimental analysis and assessments

This section will verify the superiority of the TWC-EL model through some experimental analyses. In Section 5.1, training period of BPNN model together with the number of clustering will be analyzed. The generalization capability of the constructed model is demonstrated in Section 5.2 through a comparison of the evaluation index values derived from various data sets. Section 5.3 employs 10-fold cross-validation method to identify the optimal model for the QASR aquatic toxicity data set. In Section 5.4, T-test and F-test are done based on seven data sets. In Section 5.5, validity analysis are further done.

Table 3
Experimental results of the training period.

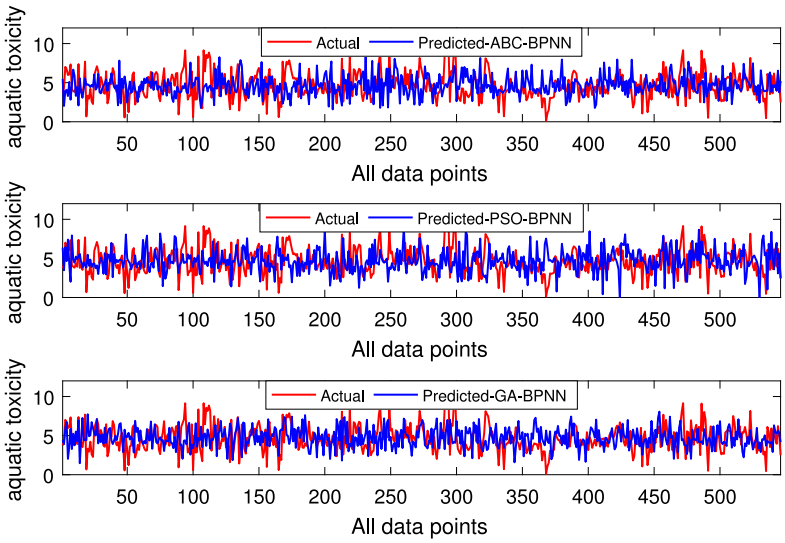
Training period (times)	MSE	SSE($\times 10^3$)	MAE	RMSE
1	0.0483	0.0275	0.1412	0.1801
2	0.0358	0.0150	0.1312	0.1667
3	0.0252	0.0028	0.1039	0.1385
4	0.0192	0.0018	0.0951	0.1232
5	0.0152	0.0013	0.0826	0.1216
6	0.0154	0.0012	0.1211	0.1081
7	**0.0117**	**0.0011**	**0.0581**	**0.0718**
8	0.0229	0.0048	0.0707	0.1003
9	0.0263	0.0171	0.1614	0.1448
10	0.0487	0.0320	0.1635	0.2081
11	0.0627	0.0427	0.1804	0.2437

Table 4
Experimental results of the number of clustering.

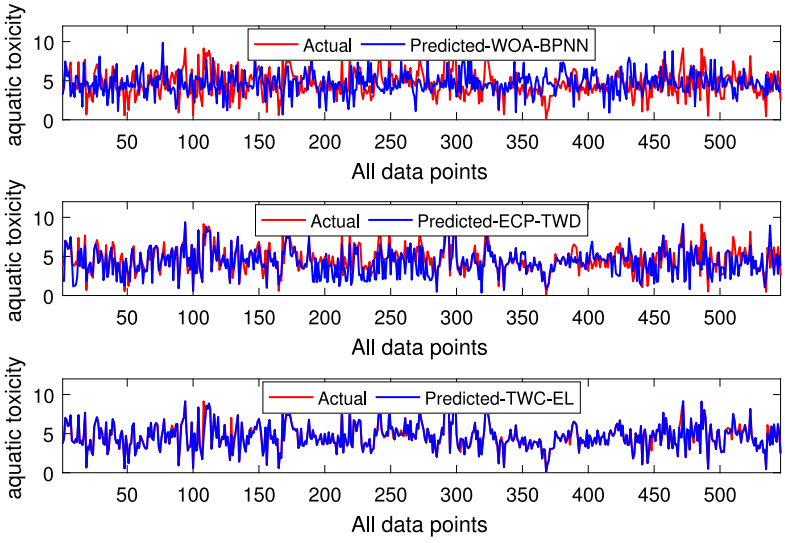
Clustering number	MSE	SSE($\times 10^3$)	MAE	RMSE
6	0.3289	0.0341	0.1872	0.2342
7	0.1933	0.0112	0.1079	0.1398
8	0.1605	0.0044	0.1066	0.1257
9	0.1307	0.0015	0.0932	0.1109
10	**0.1096**	**0.0011**	**0.0782**	**0.1017**
11	0.1396	0.0106	0.0953	0.1244
12	0.1458	0.0142	0.1238	0.1513
13	0.1501	0.0187	0.1262	0.1592
14	0.1754	0.0218	0.1565	0.2022

5.1. Training period, number of clustering and clustering times experiments

When making predictions, the parameters of prediction model will affect prediction accuracy. Meanwhile, in this paper, because clustering is involved, the number of cluster will impact the final prediction outcome. Moreover, clustering times also affects the prediction accuracy. Therefore, the training period of model, the number of clustering and the clustering times are experimented in this section, where take the training period of the BPNN model as an example for the training period experiment. Notably, the Elman neural network model and ELM model do the same experiment. The results in Tables 3 and 4 indicate that the TWC-EL model achieves the best prediction performance at a cluster number of 10 and a training period of 7. Meanwhile, it can be clearly seen that the prediction accuracy is optimal when clustering times is 2 from Table 5.



(a) Prediction results of the ABC-BPNN model, the PSO-BPNN model together with the GA-BPNN model.



(b) Prediction results of the WOA-BPNN model, the ECP-TWD model together with the TWC-EL model.

Fig. 5. Prediction outcomes of the models of state-of-the-art prediction.

Table 5
Experimental results of the clustering times.

Clustering times	MSE	SSE($\times 10^3$)	MAE	RMSE
2	**0.0707**	**0.1669**	**0.1972**	**0.2461**
3	0.1008	0.1739	0.2372	0.3174
4	0.1917	0.2501	0.2740	0.4379
5	0.1994	0.2786	0.3589	0.5468
6	0.2393	0.2885	0.3690	0.7728
7	0.3083	0.3344	0.6033	1.0719
8	0.3419	0.5814	0.9667	1.3932

In addition, during the experiment, the size of MSE is used as a standard to acquire the optimal training period together with number of clustering. Figs. 6 and 7 display the results, with "Best" representing the error associated with the best prediction performance, and "Goal" indicating the optimal training period and number of clusters. As shown in Fig. 6, the TWC-EL model achieves the best prediction performance

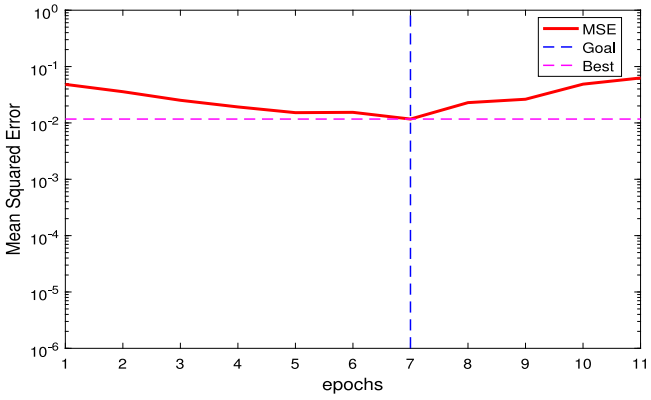


Fig. 6. Experimentation of training period.

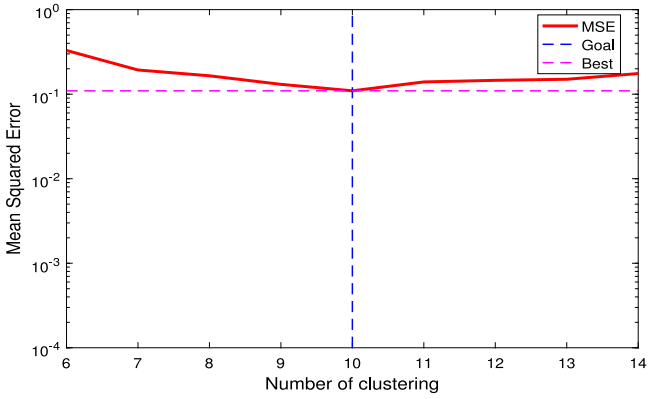


Fig. 7. Experiment on the number of clustering.

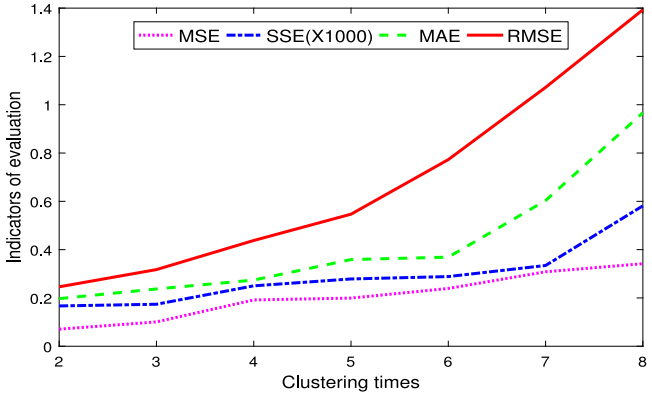


Fig. 8. Experiment on the clustering times.

at a training period of 7 for the BPNN model. Fig. 7 illustrates that the model attains the highest prediction accuracy at 10 clusters. Two figures provide evidence for the superiority of the presented prediction model. Similarly, the same experiments are conducted for other evaluation metrics and the same conclusions are also obtained. Meanwhile, it can be seen that the evaluation metrics are monotonically increasing from Fig. 8, which further indicates that clustering twice leads to the best predictive performance of the proposed model.

5.2. Data set experiments

In this section, the availability and validity of the TWC-EL model by means of different data set experiments are fully validated. In particular,seven data sets are "QSAR aquatic toxicity", "Heart failure clinical records", "Boston house price", "QASR fish toxicity", "Appliances energy prediction", "Air Quality" and "Concrete Slump Test", where "QSAR aquatic toxicity" and "QASR fish toxicity" have the same decision attribute, but they have different conditional attributes. The conditional attributes of "QASR fish toxicity" are MLOGP (molecular properties), CIC0 (information indices), GATS1i (2D autocorrelations), NdssC (atom-type counts), NdsCH (atom-type counts), SM1-Dz(Z) (2D matrix-based descriptors), while the conditional attributes of "QSAR aquatic toxicity" TPSA(Tot) (Molecular properties), SAacc (Molecular properties), H-050 (Atom-centered fragments), MLOGP (Molecular properties), RDCHI (Connectivity indices), GATS1p (2D autocorrelations), nN (Constitutional indices), C-040 (Atom-centered fragments), respectively. Table 6 presents the particular circumstances of these data sets.

Next, the TWC-EL model is experimentally evaluated with seven data sets and compared against five models of state-of-the-art prediction. The results of the four evaluation indicators discussed in Section 4.1 are presented in Table 7. As shown in the table, it is evident that the TWC-EL model outperforms the others significantly. In addition, the reason for selecting these five models for comparison is the superior prediction accuracy of latest models as compared to classical prediction models.

The superiority of the TEC-EL model is further demonstrated by presenting the evaluation metrics of six models in Fig. 9. However, some of the evaluation indicators are not visible due to significant variations among the results. Overall, the presented TWC-EL model shows the best fit for different data sets. This is verified by the smallest error in each case, which fully supports the generality of the TEC-EL model.

5.3. Cross validation examinations

The selection of the most suitable model to adapt to a given data set is typically done by choosing from a range of available methods. In this article, the K-fold cross-validation method is utilized to check the presented TWC-EL model with the following procedure:

Step 1: Assume that n samples are involved in the experiment. We divide these samples into K parts, where each part contains $\frac{n}{K}$ samples. Then, a part is randomly chosen as the testing set and the rest as the training set.

Step 2: The TWC-EL model is trained by training set, and then the goodness of the TWC-EL model is tested via tthe raining set. At the same time, the prediction error is calculated.

Step 3: Cross-validation is performed K times and the average prediction error will be seen as the final prediction error.

Note that the choice of K is the key step in the above steps. In this paper, let $K = 10$. Furthermore, as the proposed model in this article is an ensemble model, the five state-of-the-art prediction models are selected to compare with it. Table 8 gives the average results of the 10-fold cross-validation. As can be observed, the proposed TWC-EL model has a more excellent performance compared to other models. This phenomenon effectively illustrates the superiority and stability of the TWC-EL model.

5.4. Hypothesis testing

Hypothesis testing is an effective statistical inference method, which can determine whether the differences between samples and the whole are caused by sampling errors or by essential differences. In this section, T-test and F-test are employed to check significant differences between true value and predicted value. Based on the seven data sets from Section 5.2, the outcomes of T-test and F-test on these data sets through using the same method are shown in Table 9, where CST, HFCR, BHP, QAT, QFT, AQ, and AEP are the abbreviations of seven data sets, namely, "Concrete Slump Test", "Heart failure clinical records", "Boston house price", "QSAR aquatic toxicity", "QASR fish toxicity", "Air Quality" and "Appliances energy prediction", respectively.

In this article, both the T-test and F-test are conducted with a significance level of 0.05, and the null hypothesis, which assumes that the forecasting value together with the true value follow the same distribution, is evaluated accordingly. If $H = 0$, the null hypothesis is accepted. Conversely, if $H = 1$, the null hypothesis will be rejected. Moreover, the calculated probability, denoted as P, indicates the likelihood of the observed value occurring when the null hypothesis is true. The probability principle suggests that smaller probabilities are more likely to result in the null hypothesis being rejected, while higher probabilities result in the null hypothesis being retained.

Table 9 confirms that all values of H are zero, implying the null hypothesis is correct. Therefore, the forecasting values are rational. Additionally, all P-values for the seven data sets under the T-test range from 0.9787 to 0.9993, and for the seven data sets under the F-test range from 0.9838 to 0.9957. These values are all close to 1, which further supports the accuracy and validity of our forecasting outcomes.

Table 6
The introduction of seven data sets.

Data set	Number of objects	Condition attributes	Decision attribute
Concrete Slump Test	103	9	Flow
Heart failure clinical records	299	12	Death event
Boston house price	506	13	House price
QASR aquatic toxicity	546	8	Quantitative response, LC50
QASR fish toxicity	908	6	Quantitative response, LC50
Air Quality	9357	14	AH
Appliances energy prediction	19734	26	Tdewpoint

Table 7
Evaluation metric values based on different data sets.

Data sets	Models	MSE	SSE($\times 10^3$)	MAE	RMSE
Concrete Slump Test	ABC-BPNN	5.1722	5.3274	4.4842	7.1918
	GA-BPNN	1.0279	1.0587	1.9212	3.2060
	PSO-BPNN	8.8461	9.1115	6.3389	9.4054
	WOA-BPNN	1.5512	0.1598	0.6912	1.2455
	ECP-TWD	1.5154	1.5609	4.1502	1.2310
	TWC-EL	**0.0205**	**0.2279**	**0.1168**	**0.1410**
Heart failure clinical records	ABC-BPNN	0.1673	0.0501	0.3149	0.4090
	GA-BPNN	0.1285	0.0384	0.2723	0.3584
	PSO-BPNN	0.1618	0.0484	0.2695	0.4022
	WOA-BPNN	0.1131	0.0338	0.2426	0.3363
	ECP-TWD	0.2381	0.0712	0.2792	0.4879
	TWC-EL	**0.1027**	**0.0046**	**0.1711**	**0.3205**
Boston house price	ABC-BPNN	1.2361	0.0066	2.6776	3.6002
	GA-BPNN	1.0513	0.0153	2.3154	3.2373
	PSO-BPNN	1.4428	0.0073	2.6619	3.7897
	WOA-BPNN	0.7025	0.0136	1.9887	2.6497
	ECP-TWD	0.3286	0.0166	1.4570	1.9266
	TWC-EL	**0.2067**	**0.0061**	**0.2876**	**0.1961**
QASR aquatic toxicity	ABC-BPNN	1.3879	0.7578	0.8847	1.1781
	GA-BPNN	1.3339	0.5522	0.8980	1.1549
	PSO-BPNN	1.4702	0.6087	0.9260	1.2125
	WOA-BPNN	0.9065	0.3753	0.7032	0.9521
	ECP-TWD	0.6805	0.7671	0.6040	0.8250
	TWC-EL	**0.0707**	**0.1669**	**0.1972**	**0.2461**
QASR fish toxicity	ABC-BPNN	1.2850	1.1668	0.6697	1.1336
	GA-BPNN	0.8673	0.7875	0.6526	0.9313
	PSO-BPNN	1.0063	0.9137	0.6682	1.0031
	WOA-BPNN	0.7454	0.6769	0.6199	0.8634
	ECP-TWD	0.5792	0.5259	0.1533	0.2407
	TWC-EL	**0.0183**	**0.0014**	**0.0958**	**0.1311**
Air Quality	ABC-BPNN	0.0595	0.5566	0.1916	0.2439
	GA-BPNN	0.0154	0.1442	0.0932	0.1242
	PSO-BPNN	0.0335	0.3139	0.1415	0.1832
	WOA-BPNN	0.0041	0.0387	0.0486	0.0643
	ECP-TWD	0.2097	0.1962	0.2718	0.1448
	TWC-EL	**0.0022**	**0.0071**	**0.0284**	**0.0541**
Appliances energy prediction	ABC-BPNN	0.0027	0.0537	0.0403	0.0522
	GA-BPNN	0.0028	0.0559	0.0409	0.0532
	PSO-BPNN	0.0025	0.0491	0.0385	0.0499
	WOA-BPNN	0.0022	0.0427	0.0362	0.0465
	ECP-TWD	1.5154	0.0607	0.0438	0.0555
	TWC-EL	**0.0016**	**0.0204**	**0.0215**	**0.0394**

Table 8
Results of cross-validation examinations according to QSAR aquatic toxicity data set.

Model	MSE	SSE($\times 10^3$)	MAE	RMSE
ABC-BPNN	1.7062	0.9316	0.8902	1.2768
GA-BPNN	1.2260	0.6694	0.8201	1.1042
PSO-BPNN	1.5313	0.8361	0.8567	1.2259
WOA-BPNN	1.0272	0.5608	0.7622	1.0122
ECP-TWD	0.1435	0.7836	0.3197	0.3788
TWC-EL	**0.0658**	**0.4940**	**0.1498**	**0.2021**

5.5. Validity analysis

In this section, to illustrate the effectiveness of the proposed ensemble prediction strategy, we constructed cross-over experiments. In other words, we use the experiments to illustrate the rationality of using the BPNN model to predict high correlation regions, the ELM model to predict medium correlation regions, and the Elman neural network model to predict low correlation regions. The experimental results obtained based on seven real-world datasets are displayed in Table 10. In Table 10, it is concluded that the strategy of using the BPNN to predict high correlation region, the ELM model to predict medium correlation region, and the Elman network model to predict low correlation region is the best under all the indicators in the selected data sets, which fully demonstrates the effectiveness and reasonableness of the strategy of

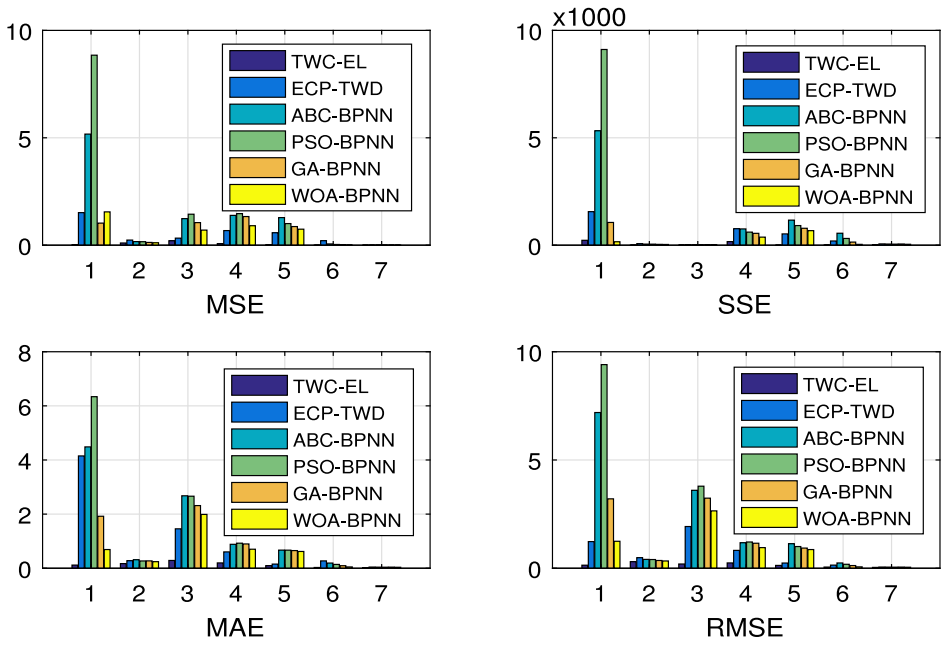


Fig. 9. Comparison of models based on different data sets.

Table 9
Outcomes of hypothesis testing.

Approaches	Metrics	CST	HFCR	BHP	QAT	QFT	AQ	AEP
T-test	H	0	0	0	0	0	0	0
	P	0.9989	0.9942	0.9855	0.9850	0.9787	0.9993	0.9923
F-test	H	0	0	0	0	0	0	0
	P	0.9931	0.9955	0.9838	0.9862	0.9905	0.9957	0.9934

adopting these three prediction models to predict the corresponding region in this paper.

In addition, this section illustrates the validity of the proposed TWC-EL model by comparing three existing TWC methods, including Zhang's model [36], BS3 model [34] and 3WC-D model [35] are utilized to compare with the TWC model of this paper. To illustrate the differences between the different TWC models, the differences are shown in Table 11. Moreover, the comparative experimental results are shown in Tables 12. Obviously, compared with above three existing TWC methods, the TWC method proposed in this article is the most effective in improving the prediction accuracy, which further illustrates the superiority of the proposed TWC method.

6. Conclusions

In the context of multivariate information systems, a new multivariate prediction model has been constructed in light of TWC and ensemble learning. This study has utilized experimental analysis to compare classical and the models of state-of-the-art prediction with the novel TWC-EL model to demonstrate its superiority. The particular contributions of the present article are displayed as follows:

(1) The objects have been categorized into three regions, namely, low, medium and high correlation regions, based on their correlation degrees. This categorization enhances the fit of between actual and predicted values and facilitates cognition.

(2) For these three regions that are divided and completed, the Elman neural network model, ELM model, together with BPNN model have been unified to forecast separately to improve prediction accuracy, instead of adopting single models to predict. Meanwhile, the importance of ensemble learning ideas in considering prediction problems has been demonstrated.

(3) The k-means clustering algorithm has been performed twice to acquire two different clustering results for the object set. These two results have been then combined to generate a new clustering result by comparing the number of intersection points between the clusters in the first result and the clusters in the second result. This step aims at enhancing the classification accuracy rate of clustering.

Future research can focus on the following aspects: (1) More methods from machine learning to prediction problems will be explored [48]; (2) For data with very many attributes [3,49], how to remove the redundant attributes will be a prospective topic; (3) In the case of incomplete data sets [50,51], the prediction problems are worth considering.

CRediT authorship contribution statement

Xunjin Wu: Investigation, Conceptualization, Methodology, Writing – original draft. **Jianming Zhan:** Methodology, Investigation, Writing – original draft. **Weiping Ding:** Methodology, Writing – review & editing.

Declaration of competing interest

All authors declare that there is no conflict of interest regarding the publication of this manuscript.

Data availability

Data will be made available on request.

Acknowledgments

The authors are extremely grateful to the editors and three anonymous referees for their valuable comments which helped us improve the presentation of this article.

The work was partially supported by grants from the NNSFC (12271146; 12161036; 61976120).

Table 10
Prediction results based on different combinations of different data sets.

Data sets	$NCo(1)$, $NCo(2)$, $Fr(C)$	MSE	SSE($\times 10^3$)	MAE	RMSE
CST	Elman, ELM, BPNN	0.0372	0.4726	0.1402	0.1705
	Elman, BPNN, ELM	0.1510	0.8948	0.2809	0.3376
	ELM, BPNN, Elman	0.0270	0.3175	0.1371	0.1639
	ELM, Elman, BPNN	0.0858	0.4306	0.2104	0.2327
	BPNN, Elman, ELM	0.0485	0.4411	0.1713	0.1918
	BPNN, ELM, Elman	**0.0205**	**0.2279**	**0.1168**	**0.1410**
HFCR	Elman, ELM, BPNN	0.2526	0.0834	0.3684	0.4935
	Elman, BPNN, ELM	0.1368	0.0517	0.2701	0.3559
	ELM, BPNN, Elman	0.2235	0.0621	0.3201	0.4706
	ELM, Elman, BPNN	0.1923	0.5521	0.3481	0.4313
	BPNN, Elman, ELM	0.2869	0.0104	0.4048	0.5004
	BPNN, ELM, Elman	**0.1027**	**0.0046**	**0.1711**	**0.3205**
BHP	Elman, ELM, BPNN	0.2148	2.0364	0.9235	0.9515
	Elman, BPNN, ELM	0.2658	3.5894	0.4125	0.4681
	ELM, BPNN, Elman	0.9047	2.2131	0.6390	0.7187
	ELM, Elman, BPNN	1.1334	1.3949	0.3260	0.3089
	BPNN, Elman, ELM	0.9685	2.4167	0.2958	0.2822
	BPNN, ELM, Elman	**0.2067**	**0.0061**	**0.2876**	**0.1961**
QAT	Elman, ELM, BPNN	0.1243	0.3436	0.2953	0.3525
	Elman, BPNN, ELM	0.0806	0.2691	0.2244	0.2838
	ELM, BPNN, Elman	0.1051	0.2296	0.2031	0.3242
	ELM, Elman, BPNN	0.1272	0.2427	0.2482	0.2334
	BPNN, Elman, ELM	0.1205	0.1768	0.2035	0.2571
	BPNN, ELM, Elman	**0.0707**	**0.1669**	**0.1972**	**0.2461**
QFT	Elman, ELM, BPNN	0.0359	0.4012	0.1323	0.1806
	Elman, BPNN, ELM	0.0281	2.7234	0.1325	0.1625
	ELM, BPNN, Elman	0.0239	0.1217	0.1140	0.1418
	ELM, Elman, BPNN	0.1070	0.4748	0.1956	0.2605
	BPNN, Elman, ELM	0.0290	0.3367	0.1158	0.1610
	BPNN, ELM, Elman	**0.0183**	**0.0014**	**0.0958**	**0.1311**
AQ	Elman, ELM, BPNN	0.0031	0.3912	0.0568	0.0875
	Elman, BPNN, ELM	0.0264	0.1460	0.0826	0.1032
	ELM, BPNN, Elman	0.0090	1.0886	0.0350	0.0781
	ELM, Elman, BPNN	0.0030	1.4194	0.0341	0.0889
	BPNN, Elman, ELM	0.0063	0.5069	0.0603	0.0793
	BPNN, ELM, Elman	**0.0022**	**0.0071**	**0.0284**	**0.0541**
AEP	Elman, ELM, BPNN	1.1048	1.9605	0.7022	0.9467
	Elman, BPNN, ELM	0.5349	1.3474	0.4824	1.0342
	ELM, BPNN, Elman	1.5291	0.7424	0.8109	0.9528
	ELM, Elman, BPNN	0.6594	0.9400	1.5159	1.7975
	BPNN, Elman, ELM	0.6474	0.5885	0.5445	0.6615
	BPNN, ELM, Elman	**0.0016**	**0.0204**	**0.0215**	**0.0394**

Table 11
Comparison among the proposed TWC method and other TWC models.

Models	Clustering times	Object feature performance	Clustering process
Zhang's model [36]	once	No	Three-way weight are proposed based on the characteristics of each object and then, the objects are assigned to three regions by combining the three assignments.
BS3 [34]	once	No	Using blurring and sharpening operations to divide each cluster into three regions.
3WC-D [35]	once	No	Based on the interrelationships between the classes, the classes in the initial clustering results are further processed and then the remaining objects are divided.
TWC-EL	twice	Yes	The results of clustering twice are fused and the final clustering result is obtained based on the characteristic properties of the objects.

Table 12
Experimental results on the proposed TWC method compared with other TWC models.

Models	MSE	SSE($\times 10^3$)	MAE	RMSE
Zhang's model [36]	0.1644	0.5129	0.3554	0.4054
BS3 [34]	0.2889	1.4095	1.1686	1.5061
3WC-D [35]	1.2174	0.4017	0.8427	1.1034
TWC-EL	**0.0707**	**0.1669**	**0.1972**	**0.2461**

References

[1] F. Petropoulos, D. Apiletti, V. Assimakopoulos, et al., Forecasting: theory and practice, Int. J. Forecast. 38 (3) (2022) 705–871.
[2] C.Y. Ding, S.L. Sun, J. Zhao, MST-GAT: A multimodal spatial–temporal graph attention network for time series anomaly detection, Inform. Fusion 89 (2023) 527–536.
[3] X.F. Huang, J.M. Zhan, W.P. Ding, W. Pedrycz, Regret theory-based multivariate fusion prediction system and its application to interest rate estimation in multi-scale information systems, Inform. Fusion 99 (2023) 101860.
[4] J. Enes, R. Expósito, J. Fuentes, J. Cacheiro, J. Touriño, A pipeline architecture for feature-based unsupervised clustering using multivariate time series from HPC jobs, Inform. Fusion 93 (2023) 1–20.
[5] Y. Li, K.L. Liu, A. Foley, et al., Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review, Renew. Sust. Energ. Rev. 113 (2019) 109254.
[6] P.H. Li, Z.J. Zhang, R. Grosu, Z.W. Deng, J. Hou, Y.J. Rong, R. Wu, An end-to-end neural network framework for state-of-health estimation and remaining useful life prediction of electric vehicle lithium batteries, Renew. Sust. Energ. Rev. 156 (2022) 111843.
[7] S.H. Chen, X.Q. Zhou, G. Zhou, C.L. Fan, P.X. Ding, Q.L. Chen, An online physical-based multiple linear regression model for building's hourly cooling load prediction, Energy Build. 254 (2022) 111574.
[8] F.M. Huang, L.H. Pan, X.M. Fan, S.H. Jiang, J.S. Huang, C.B. Zhou, The uncertainty of landslide susceptibility prediction modeling: Suitability of linear conditioning factors, Bull. Eng. Geol. Environ. 81 (5) (2022) 182.
[9] Y.G. Duan, H. Wang, M.Q. Wei, L.J. Tan, T. Yue, Application of ARIMA-RTS optimal smoothing algorithm in gas well production prediction, Petroleum 8 (2) (2022) 270–277.
[10] E.M. Oliveira, F.L. Oliveira, Forecasting mid-long term electric energy consumption through bagging ARIMA and exponential smoothing methods, Energy 144 (2018) 776–788.
[11] D. Ghose, S. Panda, P. Swain, Prediction of water table depth in western region, Orissa using BPNN and RBFN neural networks, J. Hydrol. 394 (3–4) (2010) 296–304.
[12] C.M. Li, L. Zhu, Z.Y. He, H.M. Gao, Y. Yang, D. Yao, X.Y. Qu, Runoff prediction method based on adaptive elman neural network, Water 11 (6) (2019) 1113.
[13] J. Wang, S.Y. Lu, S.H. Wang, Y.D. Zhang, A review on extreme learning machine, Multimedia Tools Appl. 81 (29) (2022) 41611–41660.
[14] G.G. Chen, L.J. Li, Z.Z. Zhang, S.Y. Li, Short-term wind speed forecasting with principle-subordinate predictor based on conv-LSTM and improved BPNN, IEEE Access 8 (2020) 67955–67973.
[15] W.D. Jiao, C. Ma, Track prediction algorithm based on GA-BPNN, in: 3rd International Conference on Civil Aviation Safety and Information Technology, (ICCASIT), IEEE, 2021, pp. 211–217.
[16] R.Z. Wang, H.Y. Bi, A predictive model for chinese children with developmental dyslexia-based on a genetic algorithm optimized back-propagation neural network, Expert Syst. Appl. 187 (2022) 115949.
[17] Y.B. Bi, S.L. Wang, C.S. Zhang, H.Y. Cong, B. Qu, J.Z. Li, W. Gao, Safety and reliability analysis of the solid propellant casting molding process based on FFTA and PSO-BPNN, Process Saf. Environ. 164 (2022) 528–538.
[18] L.G. Cui, Y.Q. Tao, J. Deng, X.L. Liu, D.Y. Xu, G.F. Tang, BBO-BPNN and AMPSO-BPNN for multiple-criteria inventory classification, Expert Syst. Appl. 175 (2021) 114842.
[19] L.X. Chen, T.H. Wu, Z.C. Wang, X.L. Lin, Y.X. Cai, A novel hybrid BPNN model based on adaptive evolutionary artificial bee colony algorithm for water quality index prediction, Ecol. Indic. 146 (2023) 109882.
[20] Q. Wang, Z.X. Xu, Z.J. Xu, Y.Y. Shi, H.L. Wu, PM2.5 prediction model based on ABC-BP, in: International Conference on Communications, Information System and Computer Engineering, (CISCE), IEEE, 2019, pp. 140–143.
[21] H.F. Xian, J.X. Che, Unified whale optimization algorithm based multi-kernel SVR ensemble learning for wind speed forecasting, Appl. Soft Comput. 130 (2022) 109690.
[22] Z. Wang, Optimizing BP neural network prediction model based on WOA, Int. Core J. Eng. 7 (9) (2021) 342–348.
[23] W. Sun, C.C. Huang, A carbon price prediction model based on secondary decomposition algorithm and optimized back propagation neural network, J. Clean. Prod. 243 (2020) 118671.
[24] W.L. Yang, Y.H. Zhang, H.J. Wang, P. Deng, T.R. Li, Hybrid genetic model for clustering ensemble, Knowl.-Based Syst. 231 (2021) 107457.
[25] P. Deng, T.R. Li, H.J. Wang, D.X. Wang, S. Horng, R. Liu, Graph regularized sparse non-negative matrix factorization for clustering, IEEE Trans. Comput. Soc. Syst. 10 (3) (2023) 910–921.
[26] W. Chen, H.J. Wang, Z.G. Long, T.R. Li, Fast flexible bipartite graph model for co-clustering, IEEE Trans. Knowl. Data Eng. 35 (7) (2023) 6930–6940.
[27] Y.Y. Yao, Three-way decisions with probabilistic rough sets, Inform. Sci. 180 (3) (2010) 341–353.
[28] Y.Y. Yao, The superiority of three-way decisions in probabilistic rough set models, Inform. Sci. 181 (6) (2011) 1080–1096.
[29] Y.Y. Yao, Three-way decisions and cognitive computing, Cogn. Comput. 8 (4) (2016) 543–554.
[30] J.M. Zhan, H.B. Jiang, Y.Y. Yao, Three-way multiattribute decision-making based on outranking relations, IEEE Trans. Fuzzy Syst. 29 (10) (2021) 2844–2858.
[31] T.X. Wang, H.X. Li, X.Z. Zhou, B. Huang, H.B. Zhu, A prospect theory-based three-way decision model, Knowl.-Based Syst. 203 (2020) 106129.
[32] F. Jia, P.D. Liu, A novel three-way decision model under multiple-criteria environment, Inform. Sci. 471 (2019) 29–51.
[33] H. Yu, Three-way decisions and three-way clustering, in: Rough Sets: International Joint Conference, (IJCRS), Springer, 2018, pp. 13–28.
[34] B. Ali, N. Azam, J.T. Yao, A three-way clustering approach using image enhancement operations, Internat. J. Approx. Reason. 149 (2022) 1–38.
[35] R.T. Zhang, X.L. Ma, J.M. Zhan, Y.Y. Yao, 3WC-D: A feature distribution-based adaptive three-way clustering method, Appl. Intell. 53 (2023) 15561–15579.
[36] K. Zhang, A three-way c-means algorithm, Appl. Soft Comput. 82 (2019) 105536.
[37] M. Afridi, N. Azam, J.T. Yao, Variance based three-way clustering approaches for handling overlapping clustering, Internat. J. Approx. Reason. 118 (2020) 47–63.
[38] X.L. Chu, B.Z. Sun, X. Li, K.Y. Han, J.Q. Wu, Y. Zhang, Q.C. Huang, Neighborhood rough set-based three-way clustering considering attribute correlations: An approach to classification of potential gout groups, Inform. Sci. 535 (2020) 28–41.
[39] X.W. Xin, C.L. Shi, J.B. Sun, Z.A. Xue, J.H. Song, W.M. Peng, A novel attribute reduction method based on intuitionistic fuzzy three-way cognitive clustering, Appl. Intell. 53 (2) (2023) 1744–1758.
[40] P.X. Wang, Y.Y. Yao, CE3: A three-way clustering method based on mathematical morphology, Knowl.-Based Syst. 155 (2018) 54–65.
[41] H. Yu, L.Y. Chen, J.T. Yao, A three-way density peak clustering method based on evidence theory, Knowl.-Based Syst. 211 (2021) 106532.
[42] H. Moayedi, H. Nguyen, L. Foong, Nonlinear evolutionary swarm intelligence of grasshopper optimization algorithm and gray wolf optimization for weight adjustment of neural network, Eng. Comput. 37 (2021) 1265–1275.
[43] A. Kammoun, M. Alouini, On the precise error analysis of support vector machines, IEEE Open J. Signal Process. 2 (2021) 99–118.
[44] G. Dudek, P. Pelka, S. Smyl, A hybrid residual dilated LSTM and exponential smoothing model for midterm electric load forecasting, IEEE Trans. Neural Netw. Learn. 33 (7) (2021) 2879–2891.
[45] Z.Q. Li, G. Tanaka, Multi-reservoir echo state networks with sequence resampling for nonlinear time-series prediction, Neurocomputing 467 (2022) 115–129.
[46] A. Likas, N. Vlassis, J. Verbeek, The global k-means clustering algorithm, Pattern Recognit. 36 (2) (2003) 451–461.
[47] X.F. Huang, J.M. Zhan, W.P. Ding, W. Pedrycz, An error correction prediction model based on three-way decision and ensemble learning, Internat. J. Approx. Reason. 146 (2022) 21–46.
[48] A. Kwekha-Rashid, H. Abduljabbar, B. Alhayani, Coronavirus disease (COVID-19) cases analysis using machine-learning applications, Appl. Nanosci. 13 (3) (2023) 2013–2025.
[49] X.D. Na, M. Han, W.J. Ren, K. Zhong, Modified BBO-based multivariate time-series prediction system with feature subset selection and model parameter optimizationms, IEEE Trans. Cybern. 52 (4) (2022) 2163–2173.
[50] X.F. Huang, J.M. Zhan, Z.S. Xu, H. Fujita, A prospect-regret theory-based three-way decision model with intuitionistic fuzzy numbers under incomplete multi-scale decision information systems, Expert Syst. Appl. 214 (2023) 119144.
[51] J.M. Zhan, J. Ye, W.P. Ding, P.D. Liu, A novel three-way decision model based on utility theory in incomplete fuzzy decision systems, IEEE Trans. Fuzzy Syst. 30 (7) (2021) 2210–2226.