Deepar-Based Ground Subsidence Prediction Method

Tianyu Li[1(✉)], Feng Xiao[1], Jiaying Li[1], and Jiaqing Zhang[2]

[1] School of Computer Science, Guangdong University of Technology, Guangzhou, China
1755681356@qq.com, {2111905198,2111905224}@mail2.gdut.edu.cn
[2] The Third Affiliated Hospital of Southern Medical University, Guangzhou, China
zjq1190@smu.edu.cn

Abstract. Ground subsidence near transmission lines is frequent, and there is a certain safety hazard for the normal operation of transmission lines. The existing deterministic models for ground settlement prediction are complicated to apply and require specified data parameters, which are difficult to obtain; the time series prediction based on single point historical observation data has problems such as lack of data volume. In this paper, we propose a model that uses smoothed formulation of DTW (Soft-DTW) to measure the similarity of ground settlement time series and combines Kmedoids for clustering, and then uses Autoregressive recurrent neural network (DeepAR) to build a prediction model for the clustered data. It achieves a unified prediction model for multiple observation points, and is simple to apply, reducing the requirement for the amount of historical data from a single observation point. The experimental results based on the subsidence data of Qujing Sentinel observation show that the accuracy of its subsidence prediction trend established by DeepAR has been improved more obviously after the classification of Kmedoids clustering method based on Soft-DTW.

Keywords: Ground subsidence · Soft-DTW · Kmedoids · DeepAR

1 Introduction

Ground subsidence is a phenomenon in which the surface soil of the earth's crust settles to varying degrees, resulting in a decrease in the height of the ground in different areas, with the total performance of ground subsidence being longer in duration, slower in development, and larger in regional impact [13]. Ground settlement disasters have had a serious impact on many countries and regions in Asia, Europe, America, Africa, and Oceania in the world's five continents, becoming a global problem [6], and the safety hazards brought about by ground settlement threaten the normal working operation of cities and have caused widespread concern.

Ground subsidence has attracted widespread attention and research because of its frequency and catastrophic nature. Many scholars have carried out a lot of research work in the two fields of ground settlement monitoring and prediction, and have made some progress, which is of guiding significance for ground settlement prevention and control.

Z. Zeng et al. (Eds.): ECIT 2022, AHE 11, pp. 760–771, 2023.
https://doi.org/10.2991/978-94-6463-005-3_77

At present, the ground settlement prediction methods are mainly physical models and time series models based on historical observation data. Physical models are limited by complex physical parameters such as regional geology and hydrology, and are difficult to be applied.

At present, the commonly used time series models are support vector machine, gray model, BP model, etc. and various improved and combined models. Reference improved the SVM model, combined with phase space reconstruction to change time series data and improved gray Wolf algorithm to optimize support vector machine, and established the deformation prediction model of PSR-IGWO-SVM, with a high prediction accuracy [9]. Literature combined the grey theory model and markov chain theory, a new dimension GM(1,1) settlement prediction model based on markov correction was established to solve the shortcomings of prediction of settlement data series with large random fluctuations [11]. It was applied to the prediction of surface settlement of buildings around foundation pit, and verified the rationality of the model. Integrated combinatorial models have been widely studied and applied in several engineering fields such as foundation pits [8], urban ground [3], and rock dams [15]. In recent years, neural networks have been widely used in subsidence prediction due to their ability to fit nonlinear problems and their power. on the basis of combining the advantages of the equal-dimensional neoclinic GM(1,1) model and BP model, an improved GM-BP combined model is established, and the prediction accuracy is higher than that of GM(1,1) and improved GM(1,1) by mining and updating the internal information of the original data series [15].

The advantage of physical model prediction of ground settlement is that it has theoretical basis and physical background, and high credibility. However, its application scenario is more limited and requires simulation prediction based on specific detailed ground parameters and a large amount of reliable measured data, which requires high data requirements. Moreover, there are uncertainties in the factors affecting settlement, and the engineering properties and physical nature of soil are complex, and this type of prediction method is not highly efficient and timely. Time series models, on the other hand, use observed historical data to achieve short-term trend predictions for the future. However, the current time series model predicts settlement only for a single observation point, and cannot predict ground settlement for a large sample with a high density of observation points.

In view of the above reasons, this paper firstly measures the similarity of subsidence time series based on Soft-DTW and conducts clustering with Kmedoids. Then, DeepAR is used to establish a prediction model for the classified data and evaluate the prediction accuracy of the model in the time series with correlation, in order to provide an efficient and convenient new method for the analysis of large sample subsidence data.

2 Methods

2.1 Model Building

The existing physical models are complicated in application and difficult in data collection, and most of the time series prediction models based on historical data only predict a single observation point. In this paper, the clustering method of large sample land subsidence sequence based on Soft-DTW Kmedoids and the subsidence prediction method

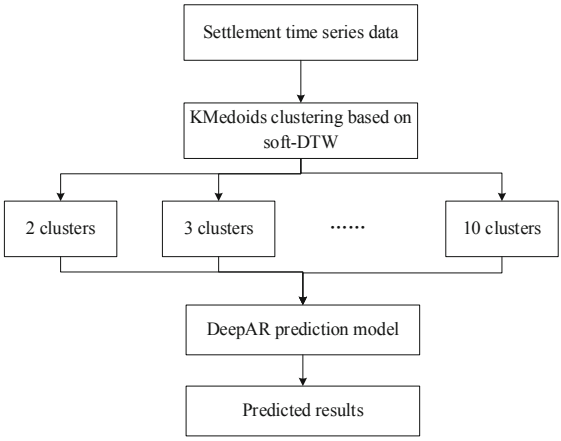

Fig. 1. Settlement prediction process.

based on DeepAR are proposed. A given set of time series is first clustered based on the similarity of the time series, and then a set of prediction models is trained. One prediction model is trained for each time series cluster, and the training set for each prediction model consists of the training parts of the corresponding time series clusters. The prediction model corresponding to each time series is used to obtain its future values, and the process can be repeated several times by changing the number of clusters to predict the subsidence trend. The overall technical idea is shown in Fig. 1.

2.2 Kmedoids Clustering Method Based on Soft-Dtw

The Dynamic Time Warping (DTW) metric was applied, originally used mainly in the field of speech recognition, to time series to solve the similarity of two sequences by dynamic programming [1]. To regularize the DTW by smoothing it, Cuturi and Blondel proposed to use the Soft minimum unification algorithm [2], as shown in Eq. (1). Since *min* is a discrete process, Soft-DTW uses min^{γ} instead. Where $x = (x_1, \ldots, x_n) \in R^{(p*n)}$ and $y = (y_1, \ldots, y_m) \in R^{(p*m)}$ are two sequences that define the cost matrix $\Delta(x,y) := ([\delta(x_i,y_i)])_{ij} \in R^{(n*m)}$, where δ is the differentiable cost function $\delta: R^p \times R^p \to R_+$ (p-dimensional information on x at a moment + p-dimensional information on y at a moment $\to$ a real value). Usually $\delta(\cdot,\cdot)$ can be used as a Euclidean distance. Define the set $A_{(n,m)} \subset \{0,1\}^{(n\times m)}$,where each element A is a matrix identifying the alignment matrix between two time series x and y.

$$\mathrm{dtw}_{\gamma}(\mathrm{x}, \mathrm{y}) = \min^{\gamma}\{\mathrm{A}, \Delta(\mathrm{x}, \mathrm{y}), \mathrm{A} \epsilon \mathrm{A}(\mathrm{n}, \mathrm{m})\} \tag{1}$$

Among them

$$\min^{\gamma}\{\mathrm{a}_1, \ldots \mathrm{a}_\mathrm{n}\} = \begin{cases} \min_{\mathrm{i}\le\mathrm{n}} \mathrm{a}_\mathrm{i}, & \gamma = 0 \\ -\gamma \log \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{e}^{-\frac{\mathrm{a}_\mathrm{i}}{\gamma}}, & \gamma > 0 \end{cases} \tag{2}$$

KMedoids clustering is a clustering method proposed by Kaufman [4, 5]. That uses the data similarity center to represent the cluster center. The KMedoids algorithm is an

improvement on the Kmeans [7] algorithm, with the difference that the initial clustering center of KMedoids is the median point of the current cluster, the sample point with the smallest sum of distances to the sample points in the other clusters. This has two major advantages: 1) it weakens the negative impact of outliers; 2) it extends the application of K-Means to the discrete sample space.

An initial cluster center is randomly selected for each cluster, and the remaining objects are assigned to the nearest specified cluster according to their distance from the center;

The centroids are replaced with non-centroids and their distances to each sample point in the cluster are recalculated, and the cluster centers are replaced if the conditions are met.

Traditional KMedodids use Euclidean distance to calculate the distance between sample points. The Euclidean distance between two n-dimensional vectors $a = (x_{11}, x_{12}, \ldots, x_{1n})$ and $b = (x_{21}, x_{22}, \ldots, x_{2n})$ is:

$$d_{ab}=\sqrt{\sum_{k=1}^{n}(x_{1k}-x_{2k})^2} \tag{3}$$

The quality of the clustering results is evaluated with an objective function that assesses the degree of similarity between the object and its clustering center. The calculation process ends when the change in the clustering centers is less than a certain threshold value. The sum of squares of its error is:

$$J=\sum_{j=1}^{K}\sum_{p\in c_j}\left(\|p-o_i\|^2\right) \tag{4}$$

where: p is a point currently belonging to a cluster of a certain class, and o_i is a representative object of a cluster of a certain class.

The clustering method of KMedoids based on Soft-DTW is used in this paper. The clustering method of KMedoids uses Soft-DTW as a subsidence time series metric and replaces Eq. (3) with Eq. (1) in order to cluster ground subsidence sequences with different subsidence trends.

2.3 Time Series Prediction of Ground Subsidence Based on Deepar

Historical ground settlement observations are characterized by nonlinearity and non-smoothness. Traditional temporal methods such as ARIMA [16], Holt-Winters [12] are usually modeled for one-dimensional time series. DeepAR is a prediction algorithm for unified modeling of a large number of correlated time series with a strong nonlinear fitting capability, which effectively learns from a large number of correlated time series global model and thus forecasting each time series [10]. Therefore, in this paper, the DeepAR algorithm is used to predict the sedimentation time series.

Denote by $c_{(i,t)}$ the value of the ith sequence at time step t, $x_{(i,t)}$ the covariate, and t0 the prediction start moment. DeepAR predicts the probability distribution of $c(i, t)$ based on an autoregressive recurrent neural network, represented by the likelihood function $Q(c_{(i,t)}|\theta_{(i,t)})$. The model is shown in Fig. 2.

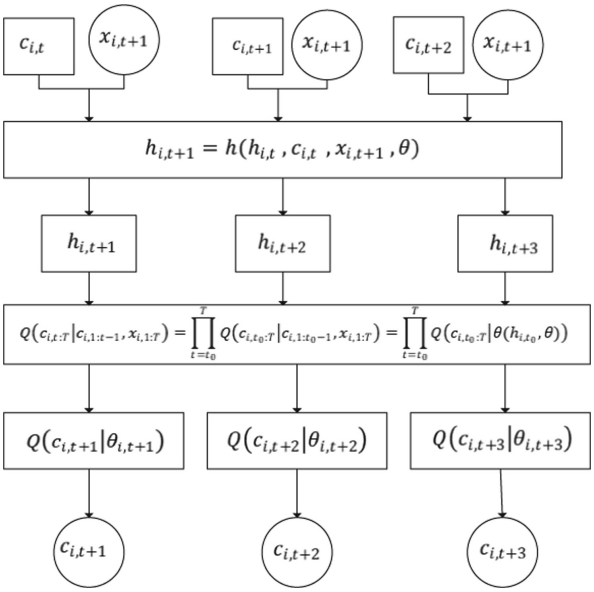

Fig. 2. Schematic diagram of DeepAR model.

where the likelihood factor of the model distribution is defined as shown in Eq. (5) (6).

$$\begin{aligned} Q\left(c_{i,\,t:T}|c_{i,\,1:t-1},\,x_{i,\,1:T}\right) &= \prod_{t\,=\,t_0}^{T} Q\left(c_{i,\,t_0:T}|c_{i,\,1:t_0-1},\,x_{i,\,1:T}\right) \\ &= \prod_{t=t_0}^{T} Q\left(c_{i,t_0:T}|\theta(h_{i,t_0},\,\theta)\right) \end{aligned} \tag{5}$$

$$h_{i,\,t\,+\,1} = h(h_{i,\,t},\,c_{i,\,t}\,,\,x_{i,\,t\,+\,1},\,\theta) \tag{6}$$

$h_{(i,t)}$ is obtained by the neural network, and the parameters in the likelihood function $Q(c_{(i,t\_0:T)}|\theta(h_{(i,t0)},\theta)$ are obtained by the affine projection of the output $h_{(i,t)}$ of the neural network through the function $\theta(h_{(i,t0)},\theta)$. In this paper, we use the Gaussian likelihood function $Q(c|\mu,\sigma)$, as shown in Eq. (7). Where μ is given by the affine transformation function of the network output, and σ is obtained by following the affine transformation of the dsoftplus activation function.

$$Q(c|\mu,\,\sigma) = (2\pi\pi^2)^{-\frac{1}{2}}\,\exp(-(c-\mu)^2/(2\sigma^2)) \tag{7}$$

$$\mu\left(h_{i,\,t}\right) = w_{\mu}^{T} h_{i,\,t} + b_{\mu} \tag{8}$$

$$\sigma\left(h_{i,\,t}\right) = \log(1 + \exp(w_{\sigma}^{T} h_{i,\,t} + b_{\sigma})) \tag{9}$$

where w is the weight matrix, b is the bias matrix, and $\mu(h_{(i,t)})$, $\sigma(h_{(i,t)})$ are the mean and standard deviation of the GaussiCan distribution function.

3 Data

The data used in the experiment of this paper are the subsidence observation data obtained by interferometeric synthetuc aperture radar (InSAR) from January 2, 2018 to May 15, 2020 in Qujing City, Yunnan Province, and the buffer zone of 500 m on both sides of the Yu Luo transmission line is selected as the study area, with a total of 11237 observation points, as shown in Fig. 3. The subsidence data are obtained from Sentine-1A radar images in C-band, and the observation data period is once a month. Among them, the observation data in June and July 2018 and July 2019 are missing, and the average of the observation values before and after the missing values is interpolated, unit is mm. Some examples of the data are shown in Table 1.

3.1 Cluster Analysis

The experiments used the Kmedoids clustering algorithm based on soft-dtw as a distance metric to unsupervisedly cluster 11237 historical time-series observations within

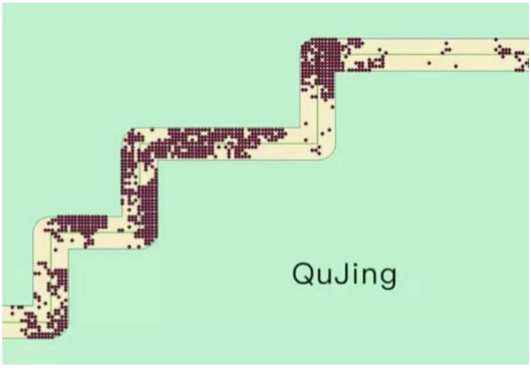

Fig. 3. Distribution of settlement observation points.

Table 1. Example of selected subsidence data in the study area.

2018-01-02	2018-02-07	…	2019-06-26	Missing Value	2019-08-01	…	2020-05-03
0	−10.7626		−10.7626		−95.842		−138.762
0	2.309245		2.309245		−83.4163		−118.475
0	5.584481		5.584481		−80.3967		−114.863
0	−17.4795		−17.4795		−106.649		−125.468
0	9.298269		9.298269		−98.8107		−81.4196
0	9.298269		9.298269		−98.8107		−81.4196
⋮	⋮		⋮		⋮		⋮

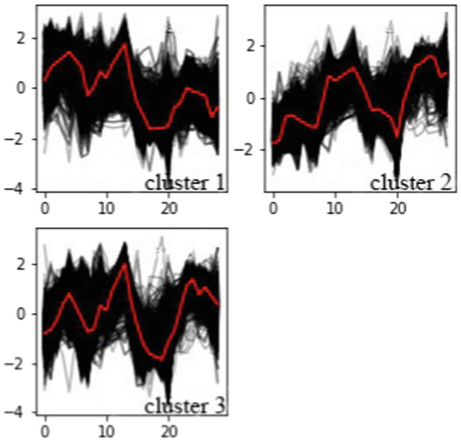

Fig. 4. Number of clusters k = 3.

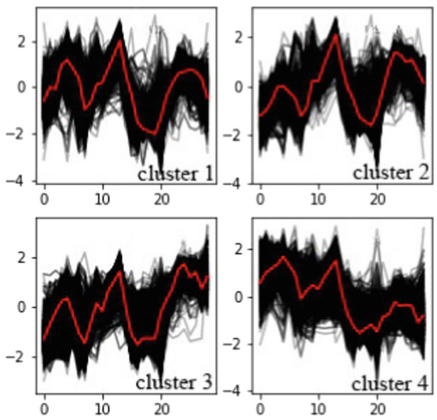

Fig. 5. Number of clusters k = 4.

a 500m buffer around the Yurow transmission line, since there is no direct method to determine the optimal number of clusters based on soft-dtw Kmedoids clustering. In this experiment, multiple classifications were performed on the sedimentation data without dimensionality and the number of clusters ranged from [3, 10]. Figures 4, 5, 6, 7, 8, 9, 10 and 11 show the trend plots after classification with different number of clusters, where the red line is the center of mass cluster.

3.2 Evaluation Metrics

The experiments use symmetric mean absolute percentage error (sMAPE) and mean absolute proportional error (MASE) to measure the performance of the model, which are commonly used error metrics in the field of time series forecasting. Equations (9) and (10) define the sMAPE and MASE error measures, respectively. Where N is the

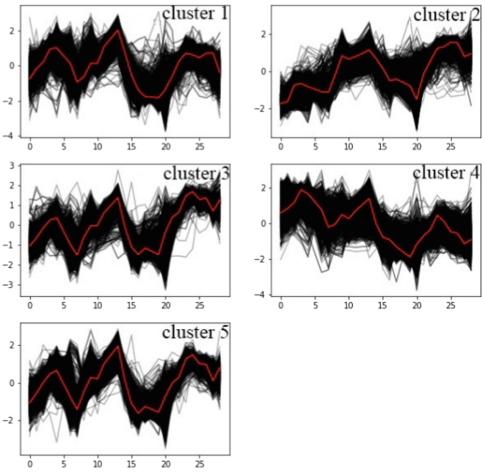

Fig. 6. Number of clusters k = 5.

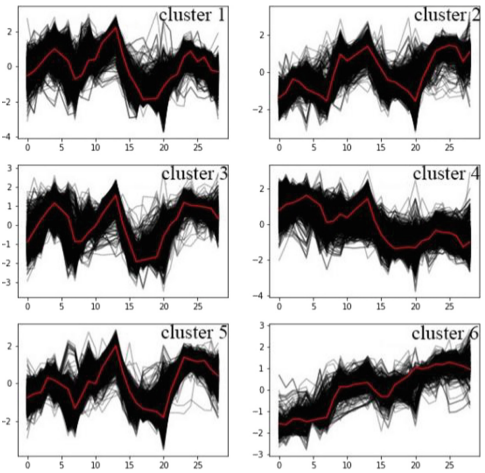

Fig. 7. Number of clusters k = 6.

number of data points in the test set, Fk is the forecast, and Yk is the actual value of the desired forecast period.

$$\text{sMAPE} = \frac{100\%}{N} \sum_{k=1}^{N} \frac{|F_k - Y_k|}{(|Y_k| + |F_k|)/2} \tag{10}$$

$$\text{MASE} = \frac{\sum_{k=1}^{N} |F_k - Y_k|}{\frac{1}{N} \sum_{k=1}^{N} |Y_k - \overline{Y}|} \tag{11}$$

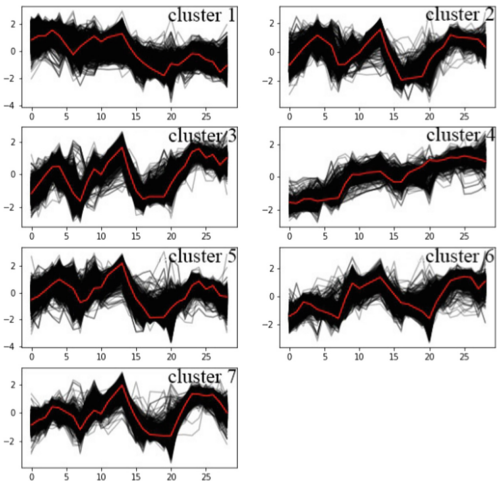

Fig. 8. Number of clusters k = 7.

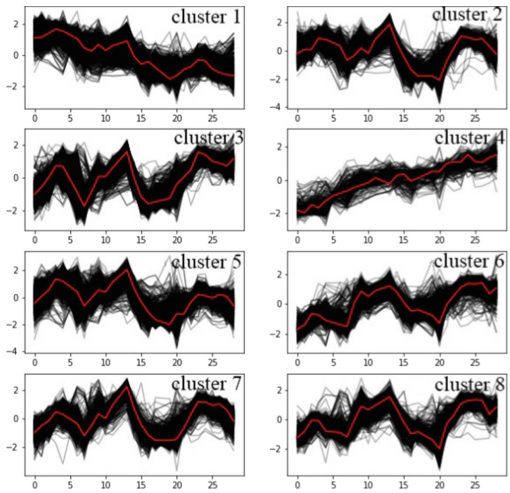

Fig. 9. Number of clusters k = 8.

3.3 Experimental Results

Each sedimentation time series is divided into a training part and a test part, and the last three data of the sedimentation data are taken as the test set, while the rest is the training set. In order to evaluate the performance of the subsidence prediction model proposed in this paper, it was compared with the direct use of DeepAR, and the MQCNN was chosen, and the transformer algorithm was compared on the dataset after using the clustering algorithm, for the number of clusters [3, 10] respectively, and the average of sMAPE and MASE was taken. From Table 2, it can be seen that after performing the clustering analysis, the model is better able to predict the subsidence trend from the time series

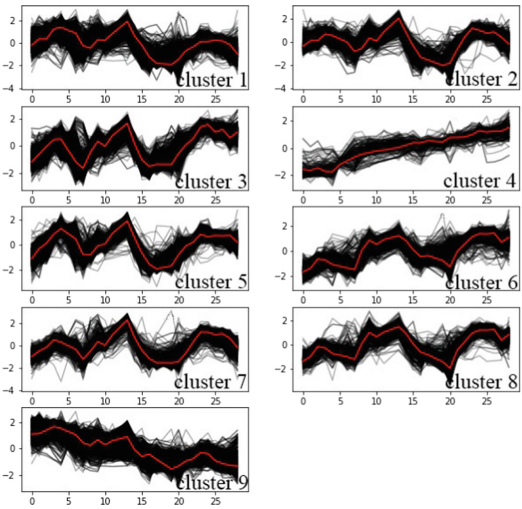

Fig. 10. Number of clusters k = 9.

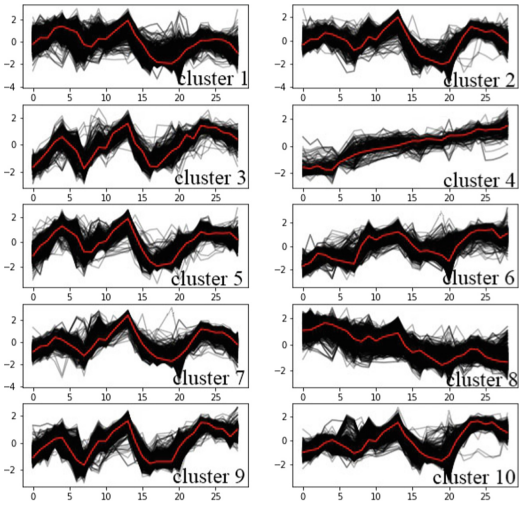

Fig. 11. Number of clusters k = 10.

with correlation, and the model used in this paper is better than other models commonly used for time series and is more suitable for the prediction of the subsidence trend.

4 Conclusions

In this paper, the proposed method of ground subsidence time series prediction model based on soft-DTW's Kmedoids clustering algorithm of DeepAR, first classifies and labels the data well, and embeds the labels into the model when modeling with DeepAR

to establish a unified prediction model. Experiments show that the DeepAR established by classifying the subsidence trend and then predicting it has high accuracy and is able to predict the subsidence sequence, while reducing the requirement for the amount of historical data at a single observation point. The method is relevant for the prediction of subsidence time series of large samples with high density of observation points.

References

1. Berndt DJ, Clifford J (1994) Using dynamic time warping to find patterns in time series. In: KDD workshop, vol 10, no 16, pp 359–370
2. Cuturi M, Blondel M (2017) Soft-DTW: a differentiable loss function for time-series. In: International conference on machine learning, pp 894–903
3. He L, Jiao M, Wang Y et al (2022) Combined model-based prediction and hazard assessment of land subsidence in Tianjin. Water Resour Hydropower Eng 53(1):178–189
4. Kaufman L, Rousseeuw PJ (2009) Finding groups in data: an introduction to cluster analysis. Wiley
5. Kaufman L, Rousseeuw P (1987) Clustering by means of medoids. North-Holland
6. Li W, Wang L, Guo H et al (2021) Effectiveness and countermeasures of land subsidence control in China (07):32–35
7. Lloyd SP (1982) Least squares quantization in pcm. Inf Theory IEEE Trans 28(2):129–137
8. Qin S, Zhang Y, Zhang L et al (2021) Prediction of ground settlement around deep foundation pit based on stacking model fusion. J Jilin Univ (Earth Sci Edn) 51(5):1316–1323
9. Rong J, Wang K, Wang W et al (2021) A settlement prediction method based on improved support vector machine model. Geotech Investig Surv 49(09):46–49+59
10. Salinas D, Flunkert V, Gasthaus J et al (2020) DeepAR: probabilistic forecasting with autoregressive recurrent networks. Int J Forecast 36(3):1181–1191
11. Weng Z, Qiu C, Qiu F et al (2020) An optimized GM(1,1) grey prediction model based on Markov chain and its application. Sci Technol Eng 20(29):12065–12070
12. Xi G, Yue J, Zhou B (2012) The forecasting model of ionospheric delay based on holt-winter. Bull Surv Map (9):7–10
13. Xue T, Shen C, Chen R (2021) Monitoring methods and control measures of land subsidence (22):20–22
14. Yan Q, Gao M, Liang M et al (2021) Combined model for CFRD settlement prediction based on wavelet SR-PSO-ELM and its application. J China Three Gorges-Univ (Nat Sci) 43(5):1–5
15. Yang F, Huang C (2020) Prediction of ground settlement in deep foundation pit excavation based on BAS-BP model. J Geomat 2020:1–6
16. Zhang GP (2003) Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50:159–175