International Journal of Machine Learning and Cybernetics (2025) 16:4151–4170
https://doi.org/10.1007/s13042-024-02503-x

ORIGINAL ARTICLE

Cluster-based prediction for product sales of E-commerce after COVID-19 pandemic

Zhaolin Lv[1] · Hongyue Kang[2] · Zhenyu Gao[2] · Xiaotian Zhuang[2] · Jun Tang[1] · Zhongshuai Wang[2] · Xintian Jiang[2]

Received: 31 January 2024 / Accepted: 6 December 2024 / Published online: 31 December 2024

Abstract
After a public emergency, predicting product sales in e-commerce can help to better understand and respond to market uncertainties and fluctuations. This can be of significant importance for business decision-making and inventory management. Therefore, in this study, we propose a novel sales forecasting model, which combines the dynamic time warping K-means clustering approach with the convolutional neural networks, Long Short-term Memory Networks, and Attention mechanism (called DKCLA) to predict E-commerce sales after the pandemic. Specifically, products with similar sales patterns are clustered. After that, the data within each cluster are used to construct a predictive model. In this case, if a new city experiences an outbreak of COVID-19 and the sales data in the early stage are obtained, the trained predictive model can be employed to predict product sales after lifting the lockdown in the city. Real sales data from a certain E-commerce platform are collected to verify the effectiveness of DKCLA. The results demonstrate that the proposed DKCLA model outperforms the other 36 benchmarks. In addition, the cluster-based prediction algorithm performs better than the non-clustered prediction algorithm in predicting product sales after the pandemic, and the number of clusters directly affects the prediction. And the learning rate and LSTM units exert great influence on the model performance.

Keywords Sales forecasting · Cluster-based prediction · COVID-19 · DTW K-means · CNN-LSTM-Attention

1 Introduction

In 2019, there was a massive coronavirus disease 2019 (COVID-19), a new type of coronavirus that was not previously described. COVID-19 spreads rapidly, infecting 4 million and causing 1976 daily deaths in the US. To curb the outbreak, countries have imposed strict lockdowns. In that case, urban transportation is disrupted, and people are stuck in a supply shortage [1]. Also, online shopping logistics are interrupted, significantly decreasing E-commerce product sales during the pandemic. However, after the pandemic is under control and lockdowns are lifted, there will be a surge in sales of commodities within a short period. Figure 1 shows the sale pattern of two items from January to October 2022 in a city in China. It is observed that online sales grow rapidly after the lockdown is lifted on May 1st, and different items show different request patterns.

Facing market uncertainties, if the E-commerce platform cannot respond in time and does not stock up in advance, it will result in product shortage and profit loss. On the contrary, if we can know the item demand information early and timely adjust inventory management, we can

✉ Hongyue Kang
19112051@bjtu.edu.cn

✉ Jun Tang
tangjun06@nudt.edu.cn

Zhaolin Lv
lvzhaolin@nudt.edu.cn

Zhenyu Gao
gaozhenyu5@jd.com

Xiaotian Zhuang
zhuangxiaotian@jd.com

Zhongshuai Wang
wangzhongshuai3@jd.com

Xintian Jiang
jiangxintian@jd.com

[1] College of Systems Engineering, National University of Defense Technology, Changsha 410073, China

[2] Dept Intelligent Supply Chain, JD Logist, Beijing 100076, China

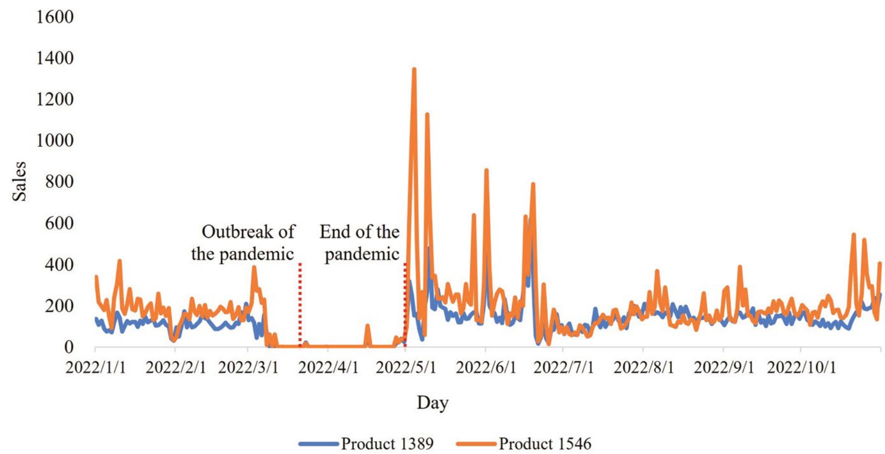


Fig. 1 Comparison of product sales before, during, and after the pandemic. The data is the sale of products 1389 and 1546 from January to October 2022 in Jilin province, which have been locked down since March 1st, and the lockdown is lifted on May 1st. Product 1389 belongs to the beauty & skincare category, while product 1546 belongs to the mother & baby category

respond to market uncertainties promptly. Therefore, it is important to predict product sales after emergencies like COVID-19. To achieve that, two challenges should be considered. On the one hand, faced with market uncertainties and fluctuations caused by emergencies, how to accurately predict sales is a significant challenge for business decision making. On the other hand, there are a wide variety of commodities with different demand patterns, and how to accurately capture different commodities' demand patterns is another challenge.

In existing works, many studies have investigated product sales forecasting. For example, Rohaan et al. [2] applied a supervised machine learning method to predict sales on B2B websites. Miguei et al. [3] proposed a daily fresh fish demand forecasting model using an LSTM network. Weng et al. [4] predicted the E-commerce supply chain sales using a light gradient boosting machine (LightGBM) and an LSTM network. However, although the prediction models utilized in these studies are effective in scenarios where minimal disparities exist between predicted and training data, they face significant challenges when applied to the unique context of post-pandemic sales forecasting. The COVID-19 pandemic has disrupted traditional market dynamics, resulting in drastic changes in sales patterns and consumer behavior that conventional forecasting approaches cannot capture. Therefore, it is not appropriate to rely solely on pre-pandemic data and traditional forecasting methods when predicting post-pandemic sales for a city. Specifically, the profound differences in sales before and after the pandemic render traditional models inadequate for accurately predicting the surge in sales post-pandemic. Moreover, the pandemic has highlighted the heterogeneous nature of product demand patterns. Different products have experienced varying degrees of demand fluctuations during the pandemic. This heterogeneity underscores the importance of adopting a tailored approach that considers the unique characteristics of various product categories. Applying a single model to predict the sales of all products would overlook these critical differences, leading to inaccurate forecasts and ineffective decision-making.

To tackle these challenges, this paper proposes a DTW K-means CNN-LSTM-Attention (DKCLA) model to predict the sales of commodities in a new region after the end of a large pandemic. Namely, the dynamic time warping (DTW) K-means clustering technique, convolutional neural networks (CNN), Long Short-term Memory (LSTM) Networks, and Attention mechanism are used to develop the proposed DKCLA model. It can address the issue of information loss due to excessively long input time series data [5]. Specifically, we collect sales data from other cities that have experienced a pandemic that has ended. Then, to obtain the sales mode of different commodities, the sales data in the early stage of the pandemic are clustered by DTW K-means to obtain K different commodity clusters. The products with similar sales patterns are divided into clusters. After that, the data within each cluster are used to construct a predictive model. In this case, if a new city experiences an outbreak of COVID-19 and the sales data in the early stage are obtained, the trained predictive model can be employed to predict product sales after lifting the lockdown in the city.

To verify the effectiveness of DKCLA, we collect real sales data from a specific E-commerce platform in cities that have already experienced a large-scale pandemic. The experimental results show that DKCLA has a higher prediction accuracy than those of the other benchmarks, including the ARIMA (autoregressive integrated moving average) [6] and the additional 35 clustering-based prediction approaches, which consist of four different clustering algorithms (i.e., K-means, DTW K-means, Softdtw K-means, and K-shape) and six prediction methods (i.e., Time Convolutional Networks (TCN), Gate Recurrent Unit (GRU), CNN-LSTM-Attention (CLA), LSTM, eXtreme Gradient Boosting (XGBoost), support vector regression (SVR)) and Transformer. Therefore, the DKCLA model proposed in this paper is an effective sales forecasting model, which can solve the problem of forecasting sales in response to large fluctuations in commodity sales after the pandemic and provide accurate forecasts for product sales with different demand patterns.

In this paper, three key contributions are made.

(1) We consider a sales prediction problem to address the surge in sales caused by lifting a lockdown in a city. To the best of our knowledge, we are the first to employ a cluster-based prediction model to predict the E-commerce product sales surge after the pandemic.
(2) We design a clustered-based prediction model, DTW K-means CNN-LSTM-Attention, to predict the surge in commodity sales after the COVID-19 pandemic. The proposed DKCLA model can solve the problem of forecasting product sales in a new city after a pandemic. It can deal with market uncertainties and fluctuations, and provide accurate forecasts for different types of commodities.
(3) The experimental results indicate that our proposed DKCLA algorithm outperforms other benchmarks regarding prediction accuracy. Additionally, the cluster-based model performs better than the non-clustered model in predicting product sales after the pandemic. Moreover, further analysis reveals that the number of clusters directly affects the prediction accuracy.

The rest of this paper is organized as follows. We introduce some preliminaries in Sect. 2. Then, we describe the related work on sales forecasting in Sect. 3. The problem to be solved is discussed in Sect. 4. The proposed cluster-based sales forecasting scheme is thoroughly described in Sect. 5. Section 6 presents the experimental results and analysis. Parameter sensitivity analysis was conducted in Sect. 7. The paper is concluded in Sect. 8.

2 Preliminaries

In this section, we introduce some preliminaries to enhance comprehension of the research. Specifically, the time series clustering method is introduced in Sect. 2.1, and the CNN-LSTM-Attention method is presented in Sect. 2.2.

2.1 Time series clustering

Time series clustering refers to the process of grouping time series data without depending on labels. Two main categories of algorithms are used for clustering time series: distance measures specifically designed for time series and extracting features from the time series. Both methods typically rely on traditional clustering algorithms like k-means. Numerous time series distance measurement techniques exist in the literature, with Euclidean distance (ED) and DTW being widely recognized as the most significant ones.

DTW matches the time dimension of another time series by distorting the time dimension of one time series. The optimal alignment is determined by minimizing the cumulative distance between corresponding elements in both time series. The optimal alignment between two given time series using the DTW method is shown in Fig. 2.

The sum of squared error (*SSE*), which represents distances between data points and their cluster centers, has been used to measure the clustering quality [7]. It can be defined by Eq. (1).

$$\min\ SSE = \min \sum_{k=1}^{K} \sum_{\mathbf{Y}\in\gamma_k} D\left(\mathbf{Y}, \Theta_k\right)^2 \tag{1}$$

where K represents the number of clusters. $\mathbf{Y}$ denotes the time series data and γ_k signifies the cluster to which time series $\mathbf{Y}$ belongs. Θ_k refers to the cluster center of cluster γ_k.

The K-means algorithm is commonly utilized in many fields for its simplicity and efficiency [8]. It is a typical distance-based clustering algorithm that uses distance as an evaluation index of similarity. An unsupervised method called DTW K-means is proposed in this paper through this approach. This method leverages the benefits of both the DTW metric and the K-means algorithm for clustering.

2.2 CNN-LSTM-Attention

CNN is made up of three layers: convolutional, pooling, and fully connected. The convolutional layer contains multiple units that use a back-propagation algorithm to optimize their parameters [9]. Its purpose is to extract distinct features from the input data through convolution operations. These operations produce feature maps representing specific neural characteristics within the network's space. The pooling layer

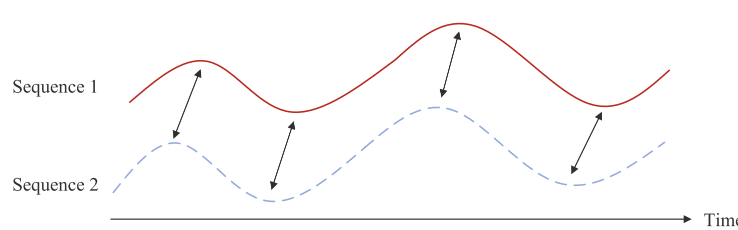


Fig. 2 The optimal alignment between two given time series uses the DTW method

serves the purpose of downsizing the feature map, aiming to decrease the computational load by reducing network parameters and mitigating overfitting. The fully connected layer amalgamates local features into global ones, which are then utilized to compute the final score for each category. In this scenario, the convolution operation involves multiplying corresponding weights with local features and accumulating their sum. On the other hand, pooling operation samples lower-layer extracted features to shrink network size and acquire invariant characteristics from input data.

LSTM incorporates a recurrent neural network with a state value and a control structure called the "gate". The transmission of data information between different units in the hidden layer is regulated by three thresholds: input gate, forget gate, and output gate [10]. The role of the forget gate is to determine whether to reset previous information and multiply it with prior memory information for retention decisions. The input gate governs the intake of current information, where it undergoes multiplication with the input unit to decide whether to store it. Similarly, the output gate manages the release of current memory information by multiplying it with present memory content to determine if it should be outputted.

In Fig. 3, f_t, O_t, and i_t denote the forget gate, output gate, and input gate, respectively. And σ is the activation function. $\widetilde{C}_t$ refers to the immediate state while C_t represents the long-term state. Then, tanh denotes a hyperbolic tangent activation function. X_t stands for the current input information whereas h signifies the output information.

When confronted with a vast amount of data, some key information is filtered and amplified. The attention mechanism is the process of identifying and choosing the most suitable inputs from a large number of available information based on the observed environmental information. It provides a way for models to focus on specific parts of input data and allocate varying degrees of importance to different elements. The attention mechanism has been extensively employed in sequential models with recurrent neural networks and LSTM. Its primary objective is to efficiently process crucial information while inhibiting irrelevant data using limited resources. It is essentially concerned with the allocation of input weights [11].

3 Related work

In this section, some related works are discussed. Specifically, the non-clustered sales forecasting method is introduced in Sect. 3.1, and the cluster-based forecasting method is compared in Sect. 3.2.

In some review literature, sequence models showed significant potential in solving time series forecasting [12]. [13] reviewed the extent to which extreme learning machines have helped companies' sales forecasts based on literature from the past 17 years. Similarly, Mallik et al. [14] reviewed sales prediction models based on machine learning algorithms. Moreover, Wu and Levinson [15] reviewed prediction methods and applications of combined models and data. Many studies have studied different approaches to sales forecasting to improve forecast accuracy.

3.1 Non-clustered sales forecasting method

Sales forecasting is commonly regarded as a time series forecasting problem. From the point of view of forecasting methods, popular time series forecasting methods like

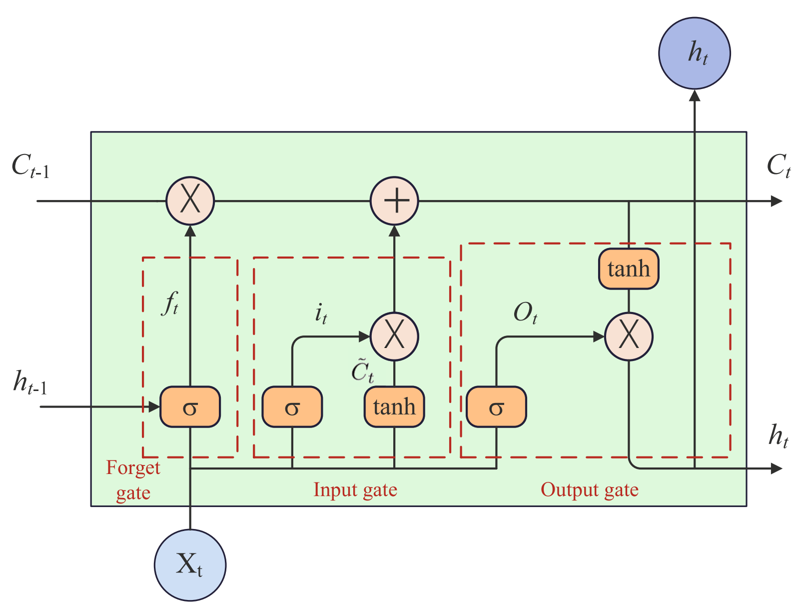


Fig. 3 The structural framework of the LSTM network model. It comprises an input gate, a forget gate, and an output gate

ARIMA, exponential smoothing (ETS) [16], and state space models [17] are widely discussed. However, the drawback of conventional time series approaches is that they utilize only internal variables for the prediction and cannot effectively capture the effects of nonlinear factors.

Therefore, a data-driven machine learning methodology has been suggested in numerous studies to convert the time series prediction problem into a supervised learning approach for handling, among which commonly used approaches include neural network models (for example, CNN and LSTM [18]), regression algorithms (such as SVR [19]), integrated algorithms based on regression trees (such as random forests (RF) [20]), and XGBoost [21]. These algorithms have been employed in many existing works [13, 22]. For example, Xu et al. [23] introduced a sales prediction model for sportswear sales utilizing the multi-layer perceptron (MLP) and the CNN. Chandriah and Naraganahalli [24] constructed a recurrent neural network (RNN)-LSTM model to predict automotive parts demand. Han [25] proposed a combined forecasting method based on the ARIMA model and an LSTM model to predict the sales of drugs. In addition to the LSTM model, XGBoost is another popular choice. Dairu et al. [21] integrated feature engineering and XGBoost models to predict the future sales of Wal-Mart stores. Ampountolas and Legg [26] investigated forecasts of hotel bookings. They constructed a hotel requirement modeling framework by incorporating machine learning methods. A segmented boosting technique utilizing social media was observed to be more accurate in predicting hotel bookings across all study timeframes by forecasting a large hotel chain in the United States.

In recent years, there has been widespread adoption of novel machine learning algorithms across various domains. This has paved the way for the application of neural network models in forecasting time series data. Attention-based models and TCN have demonstrated exceptional efficacy in handling temporal data among these advanced machine learning algorithms. Ekambaram et al. [27] presented a novel multimodal encoder-decoder model with an attention mechanism for forecasting the sales of newly launched items. Empirical results on an extensive fashion dataset demonstrated that the proposed approach enhances prediction accuracy. Li et al. [28] developed a novel model, the composite GRU-Prophet model, which incorporates an attention mechanism to forecast sales. The experimental results demonstrated that this proposed approach outperforms conventional methods such as RNN, LSTM, GRU, Prophet, and ARIMA in terms of prediction accuracy. Gandhi et al. [29] proposed a novel approach for accurately predicting future demand in the e-commerce field by incorporating Graph Neural Network (GNN), LSTM, and TCN. The experimental results demonstrate a significant improvement in prediction accuracy. In [30], an attention mechanism was integrated with LSTM to capture spatial-temporal features and weather characteristics, resulting in superior predictive performance compared to Gradient Boosting Decision Tree (GBDT), RNN, and LSTM. Similarly, Wang and Zhang [31] introduced a novel approach to prediction by integrating an attention mechanism with a TCN. Experimental results on two layered time series datasets, Dairy and Walmart, demonstrated the superior predictive performance of this model compared to other baseline models. Furthermore, Transformer-based methods have achieved remarkable performance in the field of long-term series forecasting. Yang and Lu [32] proposed a Transformer-based prediction model and validate it on three real-world datasets, which outperforms other baseline models. Yu et al.[33] designed a decomposition-based Transformer consisting of three internal modules to enhance the predictability of the Transformer. The model's predictions outperformed the benchmarks in five domains.

The study demonstrated that neural network algorithms have high accuracy in terms of predicting future sales of products. However, these approaches directly build prediction models by using all the training data without considering the degree of correlation between the training data and the predicted data. Therefore, the training data may contain too much noise, making the prediction model's performance poor [34]. To further improve the prediction model performance, some recent studies have proposed a combination of clustering and prediction to build models.

3.2 Cluster-based forecasting method

Time series clustering is a fundamental task in machine learning that finds utility across various domains. In recent years, time series clustering algorithms have developed rapidly. For instance, to effectively measure the similarity between different sequences, researchers have proposed the dynamic time warping (DTW) algorithm [35]. Using a distance matrix between sequences, this technique computes the similarity of each point in a sequence relative to multiple consecutive points in another sequence. On this basis, soft-DTW was developed, a differentiable version of DTW that can be used to compare and align time series data [36]. The k-shape algorithm is also widely employed as a time series clustering method and has demonstrated enhanced efficiency across various domains [37]. The cluster-based forecasting method has been proposed and studied by integrating these clustering algorithms in many fields. Especially in energy load prediction, such as electricity and wind speed prediction, cluster-based prediction methods are widely used. Yang et al. [38] proposed a new k-shape and enhanced memory deep attention convolutional recurrent network to achieve more accurate short-term wind speed prediction. The effectiveness of the proposed model was validated through benchmark testing with a set of widely considered methods.

Hadjout et al. [39] first used K-shape and K-means clustering techniques to identify similarities in customer consumption patterns. Subsequently, they developed a deep learning model using GRU to predict the electricity usage of each cluster. The proposed method was compared with other well-known techniques, such as ARIMA, LSTM, TCN, and two combined models, verifying that the proposed method has better predictive accuracy. Chen et al. [40] employed time series clustering techniques, including K-means (DTW) and kernel K-means, to classify daily energy consumption patterns for short-term load forecasting. The experimental results demonstrate that this approach significantly enhances the performance of the LightGBM model.

The cluster-based method is rarely employed in the domain of commodity sales forecasting. Chen and Lu [41] combined the K-means clustering (KM) technique, extreme learning machines (ELMs) model, and SVR model to construct cluster-based KM-ELM and KM-SVR models for fast fashion retailers. The results showed that the predictive accuracy of the constructed KM-ELM and KM-SVR models was better than those of the ELM and SVR models alone. Van Steenbergen and Mes [42] presented DemandForest, which combined K-means, RF, and quantile regression forests. Their experimental results showed that the DemandForest method can generate better prediction results in new product sales forecasting. Some innovative clustering algorithms have been proposed in several studies, which are subsequently employed by scholars to address the challenge of predicting new products. Malarya et al. [43] employed the DTW technique to detect comparable patterns in historical matching time and utilized it as a leading indicator in regression models for forecasting demand trends during the product introduction phase. The experimental results demonstrated that the proposed model exhibits significantly superior prediction accuracy compared to ARIMA and LSTM. Similarly, when forecasting the lifecycle of new products, the model constructed by Li et al. [44] was composed of clustering and prediction methods. The results revealed that the model based on clustered products generally outperformed models based on non-clustered products. The forecasting models combined with clustering proposed in these studies are essentially similar, which illustrates that the whole dataset is first divided into several sub-datasets using clustering algorithms before making predictions. The data in the dataset are enhanced to reduce noise and obtain similar data structures. After that, the data in each sub-dataset are utilized to train a forecasting model so that the forecasting model obtained is more accurate and stable.

In the existing works, there have been a few studies on product sales forecasting under the influence of the COVID-19 pandemic. Sleiman et al. [45] introduced a novel framework that integrates clustering, classification, and regression to account for the crisis-specific impacts on sales. By leveraging a real-world dataset from a French fashion retailer, the proposed approach demonstrated its effectiveness in capturing the pandemic's effect on consumer behavior. Park et al. [46] focused on analyzing changes in customer preferences and sentiments toward product features before and after the COVID-19 pandemic, utilizing an innovative framework that combines latent Dirichlet allocation, sentiment analysis, and interpretable machine learning. Gimbach et al. [47] quantified the influence of the COVID-19 pandemic on the consumption of attention deficit hyperactivity disorder (ADHD) medicine in 47 countries and regions using a seasonal autoregressive integrated moving average (SARIMA) model. Similarly, Andueza et al. [48] utilized time series models, including ARIMA and SARIMA, to predict the change in cigarette sales in Spain under the influence of COVID-19. Kim et al. [49] used ARIMA and ETS models to study retail data of a shopping center in Seoul, South Korea, during the COVID-19 pandemic. Tudor [50] developed an integrated framework that was able to estimate six machine learning and statistical approaches to forecast the influence of the COVID-19 pandemic on E-commerce retail sales and their share of total retail sales. However, few studies have predicted E-commerce product sales after supply chain recovery from COVID-19.

To solve this problem, in this study, we design a DKCLA model to forecast commodity sales after the supply chain has experienced disruptions and recovery from COVID-19. Our research innovatively develops a sales forecasting framework specifically designed for unique disruptions caused by the pandemic. Unlike traditional sales forecasting methods that rely on historical sales data or statistical data, we comprehensively consider the pandemic's specific impacts, including sales disruptions, shifts in consumer behavior, and supply chain adjustments. By incorporating sales data before and after pandemic, our model can capture these changes more accurately, leading to more ccurate prediction. And we utilize cross-city sales data to expand our dataset and improve generalization, allowing for forecasts in cities with limited pandemic data. It can deal with market uncertainties and fluctuations, and provide accurate forecasts for different types of commodities. To highlight the differences between our work and the existing research, Table 1 provides a comparison of related work on sales forecasting. It can be observed that this study is novel in terms of methods and application scenarios by employing the DTW K-means clustering technique, CNN, LSTM, and Attention mechanism to address the challenge of sales forecasting after logistics interruption.

4 Problem statement

In the system, we consider $I' = \{1, 2, \cdots, I\}$ products in $Q' = \{1, 2, \cdots, Q\}$ cities. Additionally, a time horizon of $T' = \left\{1, 2, \cdots, p, \cdots, p', \cdots, T\right\}$ is considered and describes the early, middle and late stages of the pandemic,

Table 1 Comparison of related works on sales forecasting

Reference	Application field	Cluster-based	Attention mechanism
Gandhi (2021)	E-commerce products sales		
Li(2021)	clothing sales		✓
Ye(2021)	online car-hailing		✓
Chandriah(2021)	automotive parts sales		
Malarya(2021)	demand trends	✓	
Gustriansyah(2022)	retail sales	✓	
Van Steenbergen(2020)	new products sales	✓	
Chen(2021)	fast fashion products sales	✓	
our work	sales after logistics disruption	✓	✓

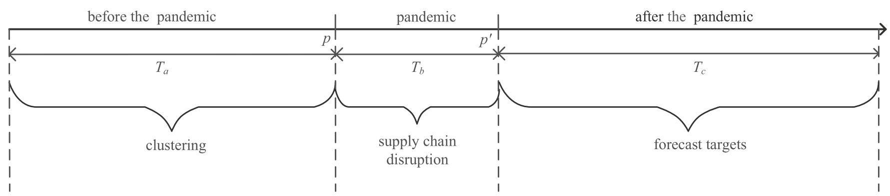


Fig. 4 The timeline of pandemic. A time horizon of $T' = \{1, 2, \cdots, p, \cdots, p', \cdots, T\}$ is considered and describes the early, middle and late stages of the pandemic, in which p and p' denote the moment of pandemic occurred and ended, respectively. T_a and T_c represent the time before and after the pandemic

as shown Fig. 4, in which p and p' denote the moment the pandemic occurred and ended, respectively. T_a and T_c represent the time before and after the pandemic. The meanings of the notations utilized in this paper are shown in Table 2. We need to predict the data of T_c in a new city. The goal of the model is to generate the minimum prediction error. It can be defined by Eq. (2):

$$\min \sqrt{\frac{1}{T - p' + 1} \sum_{t=p'}^{T} \left(\tilde{y}_{\hat{i}t}^{\hat{q}} - y_{\hat{i}t}^{\hat{q}} \right)^2} \tag{2}$$

where $y_{\hat{i}t}^{\hat{q}}$ and $\tilde{y}_{\hat{i}t}^{\hat{q}}$ represent the actual and predicted values of $\hat{i}$ products in city $\hat{q}$ at time t, respectively.

As shown in Fig. 5, the data exhibit a relatively large variation in sales patterns. Product 1 and product 2 have similar sales patterns, while product 3 and product 4 have similar sales patterns. Therefore, the data need to be clustered. When the new product and product 1 have similar sales patterns, the model constructed by the cluster containing product 1 can be used for prediction. If all the training data are used directly without clustering, the degree of correlation between the training data and the predicted data is not considered. The training data may contain too much noise, which will make the performance of the prediction model poor. To improve the prediction model performance,

Table 2 Meanings of the notations

Notation	Definition
$i, j \in I'$	product index
$q, q' \in Q'$	city index
$t \in T'$	time index
$k \in K$	cluster index
$\hat{i}$	product in test data
$\hat{q}$	city in test data
$p \in T'$	the moment of the pandemic occurred
$p' \in T'$	the moment of the pandemic ended
y_{it}^{q}	indicate product sales of i products in city q at time t
$\tilde{y}_{\hat{i}t}^{\hat{q}}$	indicate predicted value of $\hat{i}$ products in city $\hat{q}$ at time t
$\mathbf{Y}_i^q$	indicate vectors containing y_{i1}^q to y_{it}^q
K	number of clusters
θ_{kt}	the value of clustering center of cluster k at time t
Θ_k	indicate vectors containing
m	prediction steps
$\mathbf{Y}_{(T-m-l+1)\times l}$	prediction variable matrix
$\mathbf{Y}_{(T-m-l+1)\times m}$	target variable matrix
$\mathbf{V}_i^q$	indicate prediction variable and target variable matrix
$\mathbf{V}$	indicate vectors containing $\mathbf{V}_1^1$ to $\mathbf{V}_I^Q$

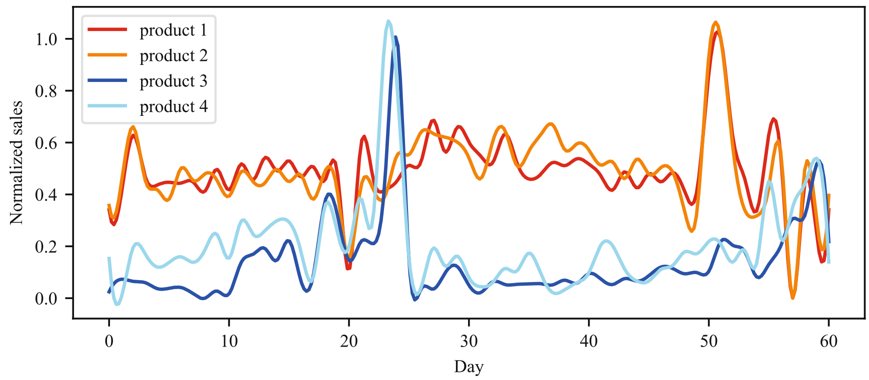


Fig. 5 Comparison of different product sales. The data exhibit a relatively large variation in terms of sales patterns. Product 1 and product 2 have similar sales patterns, while product 3 and product 4 have similar sales patterns. Therefore, the data needs to be clustered

we propose a cluster-based prediction model to forecast the sales of commodities in a new region after the end of a large pandemic. By taking into account the data characteristics of different clusters, our model can better capture the fluctuations in demand and provide more accurate predictions.

5 Proposed cluster-based sales forecasting scheme

This section describes the constructed cluster-based sales prediction model used in this paper. Section 5.1 introduces how product sales data are clustered, and Sect. 5.2 describes the CNN-LSTM-Attention model for sales forecasting. Section 5.3 presents the proposed DKCLA model.

5.1 Sales data clustering

First of all, we collect product sales data for a time length of T in the cities where the pandemic has occurred. The sales data of certain products in a certain city can be regarded as time series data $\mathbf{Y}_i^q = \left[y_{i1}^q,\ y_{i2}^q,\ \cdots\ y_{it}^q\right]$, where y_{it}^q represents the product sales of i products in city q at time t, where $i \in I', q \in Q', t \in T'$. In the clustering stage, we use the data before the pandemic to cluster and classify items with similar sales patterns into one cluster. In practice, it is generally difficult to define the value of clusters K. Therefore, we utilize the elbow approach in this study to determine K [51]. After this, the detailed procedure of clustering products into K clusters follows. Initially, K sales curves are randomly selected from all the data samples as clustering centers Θ_k $(1 \le k \le K)$. The DTW distance between each clustering center and the remaining sales curves is calculated. Each sample is assigned to the cluster whose center is closest to the sample. Then, each cluster center Θ_k is updated as the mean of its constituent data until the cluster centers do not change or a pre-determined number of training epochs is reached. The pseudo-code of DTW K-means is displayed in Algorithm 1.

Algorithm 1 DTW K-means

Input: Dataset $\mathbf{Y}_i^q = [y_{i1}^q, y_{i2}^q, \cdots, y_{it}^q]$
Output: Cluster labels $cluster\_labels$

1: extract data $\mathbf{Y}_i^{'q} = \left[y_{i1}^q, y_{i2}^q, \cdots y_{ip}^q\right]$
2: select K sale curves as the clustering centers $\Theta_K = \{\Theta_1, \Theta_2, \cdots, \Theta_k\}$
3: **for** every product sale **do**
4: calculate the DTW distance between extracted data $\mathbf{Y}_i^{'q}$ and each clustering centers Θ_k
5: DTW distance between $\mathbf{Y}_i^{'q}$ and Θ_k : $\left|\mathbf{Y}_i^{'q}\right| = p$, $|\Theta_k| = r$
6: initialize a matrix D with (n+1)×(m+1), where D[0][0] = 0 and the other elements are infinite
7: make the first row and column of matrix D infinite, indicating that the starting point is unreachable
8: let $D[a][b] = d\left(\mathbf{Y}_i^{'q}[a-1], \Theta_k[b-1]\right) + \min(D[a-1][b], D[a][b-1], D[a-1]$
9: $[b-1])$ for $1 \leq a \leq q, 1 \leq b \leq r$
10: $\mathrm{DTW}(\mathbf{Y}_i^{'q}[i-1], \Theta_k) = D[q][r]$
11: find the nearest clustering center θ_k and the corresponding cluster k
12: **end for**
13: obtain cluster $1, 2, \cdots, K$
14: **for** every cluster **do**
15: calculate the v center of every cluster
16: **end for**
17: obtain the cluster centers and clusters
18: **if** there is no change in the clusters obtained **then**
19: return cluster k
20: **else**
21: $\Theta_k = v$ and go to step 2
22: **end if**
23: **return** Outputs

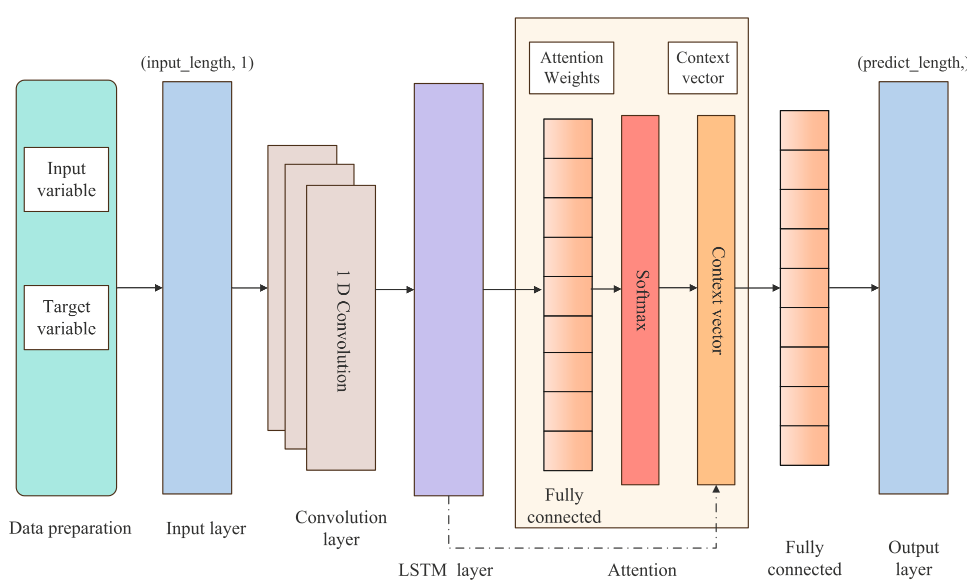


Fig. 6 The structural framework of the proposed CNN-LSTM-Attention neural network model. It is composed of the input layer, the convolution layer, the LSTM layer, the attention layer, the fully connected layer, and the output layer

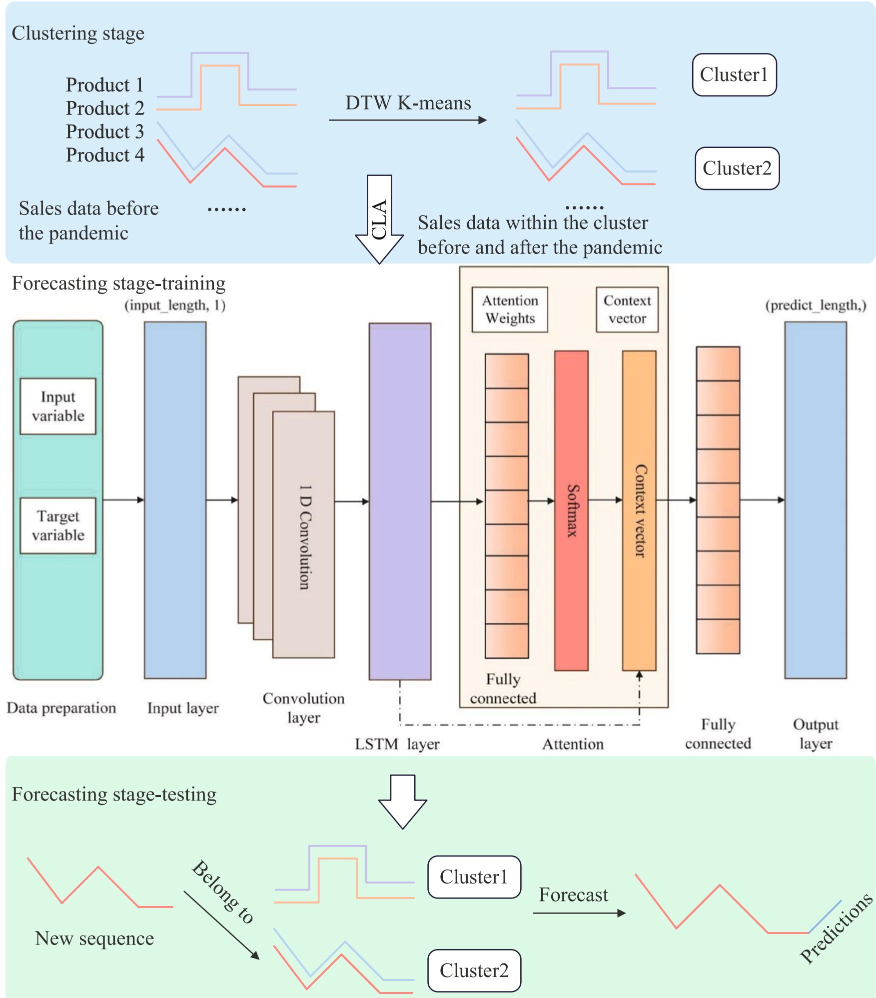


Fig. 7 The structural framework of the proposed DTW K-means CNN-LSTM-Attention method for sales forecasting. It has two stages, namely, the clustering and the forecasting stages, including the training and testing phases

After clustering, we obtain the cluster label and all the sales data of T_a and T_c for the items within each cluster. These data will be further used for sales forecasting.

5.2 CNN-LSTM-Attention for sales forecasting

We utilize the CNN-LSTM-Attention model to train separate prediction models for clusters. The detailed process is organized into the following steps:

Firstly, we extract historical lags using sliding time windows on the raw sales data to construct the prediction and

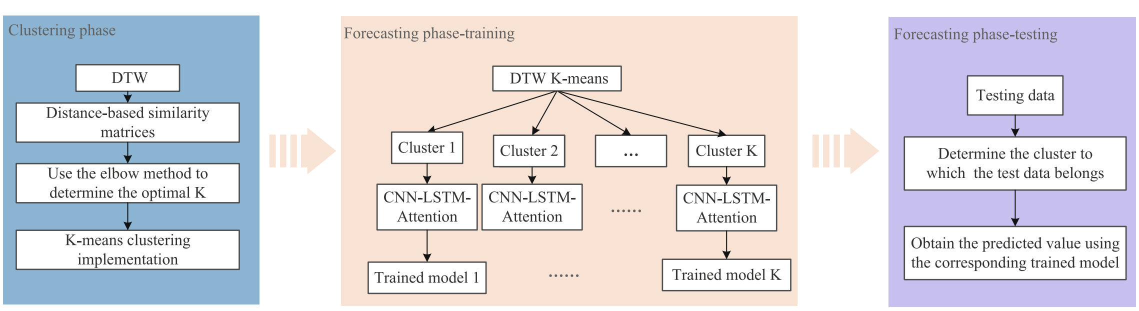


Fig. 8 The flow chart of the proposed DTW K-means CNN-LSTM-Attention method for sales forecasting. It includes how to use the DTW k-means method for clustering and achieve sales forecasting

Table 3 Data description

City	Time range of selected data	Duration of the localized outbreak
Shanghai	Jan–Oct 2022	Mar–June 2022
Shijiazhuang	Nov 2020–Aug 2021	Jan–Feb 2021
Xi'an	Oct 2021–Jul 2022	Dec 2021–Jan 2022
Jilin	Jan–Oct 2022	Mar–May 2022

Table 4 The corresponding relationships of products

Product category	Product code
nutrition and health care	27997 27998 28028
personal care	16756 16757 16813 21489 27505
home cleaning	15908 15933 21488 22291 29136
home textiles	15278
industrial products	14101
medicine	13337 13339 21921 21924
agricultural materials	17420
fresh food	12241 13582 15568 15618 21456
mobile communication	13658
building materials	9921 9925
furniture	30546
healthcare	1514 1517 23042 30914
pet life	7002 11984 13964 17318 17320
kitchenware	6232
household daily necessities	1657 11167
food and beverage	1590 1602 2675 23096
mother and baby	1533 22491 30701
sports and outdoor activities	1474 2691
beauty and skincare	1389
household appliances	760

target variable matrices. Afterward, we build the CNN-LSTM Attention model for predicting post-pandemic sales. The specific process of model construction is as follows. Firstly, the input layer of the model is defined, followed by applying a 1D convolution operation to extract local features from the input data. Subsequently, an LSTM layer is incorporated into the model to further process the convolution layer's output and effectively capture long-term dependencies within sequences. Additionally, we introduce an attention mechanism to enhance focus on different time steps in the input sequence based on their importance. Specifically, attention weight calculation and normalization are performed initially. Then, a weighted summation is conducted using attention weight and LSTM layer output. Finally, a fully connected layer serves as the output layer for generating prediction results. The structural framework is shown in Fig. 6.

5.3 Proposed DKCLA for sales forecasting

Based on the above discussion, a cluster-based predictive model DKCLA is constructed to forecast merchandise sales after the pandemic utilizing the DTW K-means algorithm and CLA algorithm. As Fig. 7 shows, the proposed method comprises two stages, namely, the clustering stage and the forecasting stage.

The construction process of the cluster-based prediction model is as follows. First, we collect sales data from other cities where pandemics have already occurred and ended. Then, in the clustering stage, the sales data of T_a are clustered by DTW K-means to obtain K different commodity clusters. The products with similar sales patterns are clustered. After that, in the forecasting stage, the data of T_a and T_c within each cluster are used to construct the predictive CLA model. In this case, if a new city experiences an outbreak of COVID-19 and the sales data of T_a are obtained, the trained predictive model can be employed to predict the product sales after the city lifts the lockdown. Specifically, we first need to identify the different clusters of sales patterns with the highest similarity to the sales pattern within the test data. Once the most similar cluster is confirmed, the trained CLA prediction model of that cluster is utilized to generate the prediction results. The following steps outline the detailed process of the test phase in the forecasting stage:

(1) We validate the model by using the sales of product $\hat{i}$ in the city $\hat{q}$ (i.e., test data) where a pandemic has occurred. We first obtain the sales data before the pandemic.
(2) After that, we can calculate the DTW distance among the sales data of T_a and the center of each cluster. The cluster with the smallest DTW distance is considered to have the most similar data patterns to the test data. We refer to this cluster as k'. The constructed prediction model of k' is the most appropriate cluster to yield sales prediction results.
(3) Finally, the test data's predicted values are generated using the trained prediction model corresponding to the clustering k'.

Figure 8 illustrates the flow chart of the proposed DKCLA method for sales forecasting. It includes how to use the DTW k-means method for clustering and achieve sales forecasting.

Table 5 Parameter settings

Model	Type	Parameter
CLA	*filters*	64
	kernel size	3
	epochs	50
	batchsize	32
TCN	*filters*	32
	dropout_rate	0.2
	epochs	50
	batchsize	32
GRU	*epochs*	50
	batchsize	32
XGBoost	*n_estimators*	100
	learning_rate	0.01
SVR	*kernel*	*rbf*
	gamma	0.05
	C	12
	epsilon	0.1
LSTM	*epochs*	50
	batchsize	1
Transformer	*u*	2
	d_model	128
	h	8
DLinear	*epochs*	50
	batchsize	32
Pyraformer	*learning_rate*	0.01
	batchsize	32
ARIMA	*w*	1
	z	1

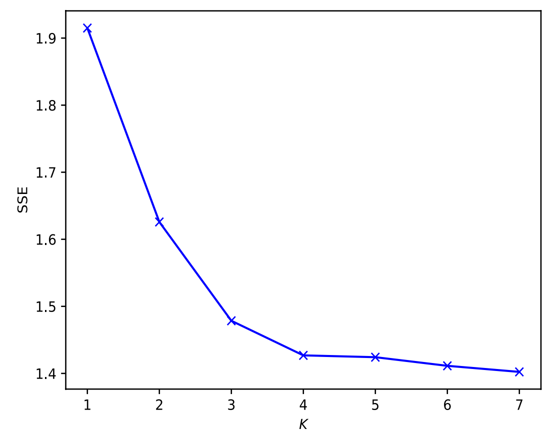


Fig. 9 *SSE* of different *K*. We can find that the curve has an inflection point at $K = 3$, so according to the principle of the elbow method, we choose the number of clusters to be 3

6 Empirical study

In this section, the experimental part is described in detail. Section 6.1 describes the data and Sect. 6.2 demonstrates the experimental setup. The evaluation metrics of the experiment are presented in Sect. 6.3. Section 6.4 shows the clustering results and the evaluation of the prediction results for each model.

6.1 Experimental dataset

In this paper, we collect sales data from a certain E-commerce platform in cities that have already experienced a large-scale pandemic. As shown in Table 3, four typical cities in China (Shanghai, Shijiazhuang, Xi'an, and Jilin) are chosen as examples. Additionally, we select the sales data of more than two hundred commodities in each city, and the time range of the collected data spans the early, middle, and late stages of the pandemic. In total, there are hundreds of instances of sales data, with each instance encompassing 238 days of sales data. The corresponding relationship between the code of a product and its category is shown in Table 4.

On the 60th day, the pandemic broke out. Data from the National Health Commission of the People's Republic of China are applied to select the periods for the experiment. Specifically, the sales data from Shanghai, Xi'an, and Shijiazhuang are used for model training, while the sales data from Jilin are used to construct the test set. The average values of the prediction evaluation metrics are calculated. The data are pre-processed by data cleaning, processing the outliers, and filling missing values.

6.2 Experimental setup

The effective prediction of machine learning and deep learning is significantly influenced by the hyperparameters. To find the optimal hyperparameters, we conduct the cross-validation method encompassing all feasible hyperparameter value combinations to capture the optimal hyperparameters and evaluate their performance using mean square error (MSE) on the training set. The key hyperparameters of the proposed model encompass batch size, number of epochs, filter number, and kernel size. To drive our model, we employed the 'Adam' optimization algorithm and the 'tanh' and 'softmax' activation functions in the attention mechanism.

Our benchmark approach consists of four different clustering algorithms (i.e., K-means, DTW K-means, Softdtw K-means, and K-shape), along with nine prediction methods (i.e., TCN, GRU, CLA, LSTM, XGBoost, SVR, Transformer, Pyraformer, DLinear). We have combined these methods to create an additional 23 clustering-based prediction approaches and ARIMA as the reference models to verify the validity of the proposed model.

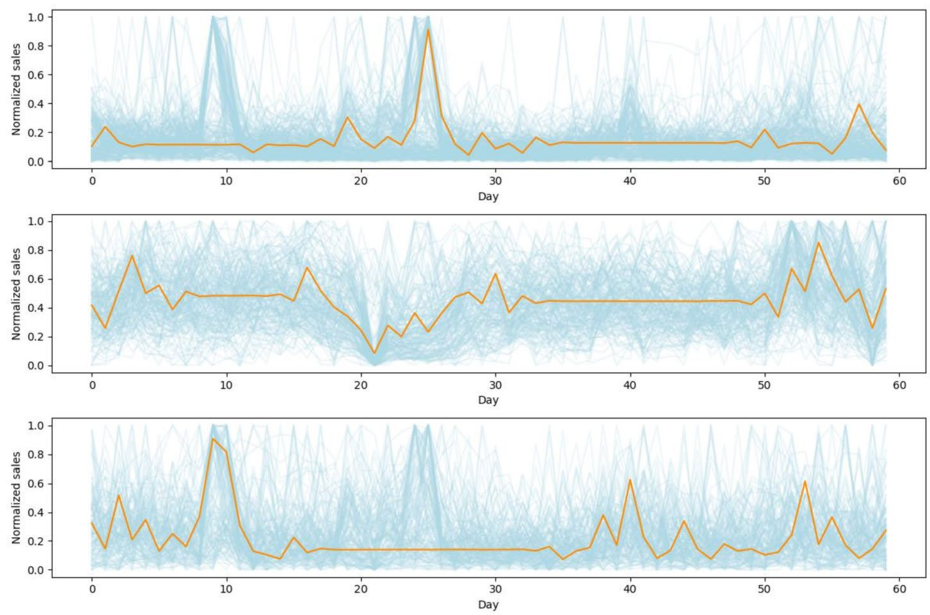


Fig. 10 Line chart of the K-means clustering result. The value of K is set to 3, and the clustering results are shown. The blue lines represent the sales curve of the items, and the orange lines represent the cluster centers

The main parameters are given in Table 5. As shown in Table 5, the *filters* = 64 is used to define the number of filters in the convolutional layer, while the *kernel size* = 3 refers to the size of the convolutional kernel. The parameter *epochs* indicates the number of iterations during the model training, and *batchsize* refers to the number of samples used in each training iteration. The *dropout_rate* parameter controls the dropout rate applied to the Dropout layer within the TCN model. And *n_estimators* in the XGBoost model, representing the number of base models is 100, and the *learning_rate* is 0.01. In SVR, the hyper-parameters are set as follows. *Kernel* represents the kernel function and the radial basis function (*rbf*) is chosen in the experiment. The value of *gamma* is the coefficient of the *rbf* kernel, and its value is 0.05. C is the regularization coefficient, and its value is 12. *Epsilon* is used to define the tolerance of the model for misclassification, and it is set to 0.1. For the Transformer, the dimension of the model *d_model* = 128, and the number of linear projections in the multi-head attention h is 8. In ARIMA, u represents the autoregression order, which is 2, and w represents the moving average order, which is 1. The difference order is z, and its value is 1.

6.3 Evaluation metrics

The root mean square error (*RMSE*), mean absolute error (*MAE*), and mean absolute percentage error (*MAPE*) are often applied to assess the validity of predictions [52]. Nevertheless, the *MAPE* tends to ∞ when the actual value has data equal to 0, which means that the calculation formula of the *MAPE* is not available. To solve this problem, the *SMAPE* was proposed [53]. Therefore, we select the *RMSE*, *MAE*, and *SMAPE* as the evaluation metrics to measure the prediction performance. These metrics are defined as follows:

$$RMSE = \sqrt{\frac{1}{T - p' + 1} \sum_{t=p'}^{T} \left(\tilde{y}_{\hat{i}t}^{\hat{q}} - y_{\hat{i}t}^{\hat{q}} \right)^2} \tag{3}$$

$$MAE = \frac{1}{T - p' + 1} \sum_{t=p'}^{T} \left| \tilde{y}_{\hat{i}t}^{\hat{q}} - y_{\hat{i}t}^{\hat{q}} \right| \tag{4}$$

$$SMAPE = \frac{1}{T - p' + 1} \sum_{t=p'}^{T} \frac{\left| \tilde{y}_{\hat{i}t}^{\hat{q}} - y_{\hat{i}t}^{\hat{q}} \right|}{\left(\tilde{y}_{\hat{i}t}^{\hat{q}} + y_{\hat{i}t}^{\hat{q}} \right)/2} \tag{5}$$

where $y_{\hat{i}t}^{\hat{q}}$ and $\tilde{y}_{\hat{i}t}^{\hat{q}}$ represent the actual and predicted value of $\hat{i}$ products in city $\hat{q}$ at time t, respectively. To mitigate the potential bias of any single metric, a comparative analysis is performed using these four metrics to assess the model performance.

6.4 Experimental result

The results of the experiment are discussed in this subsection. The experiments with all the algorithms using Python version 3.9.13 are performed on a computer equipped with a single NVIDIA GeForce RTX 3070 Ti laptop GPU.

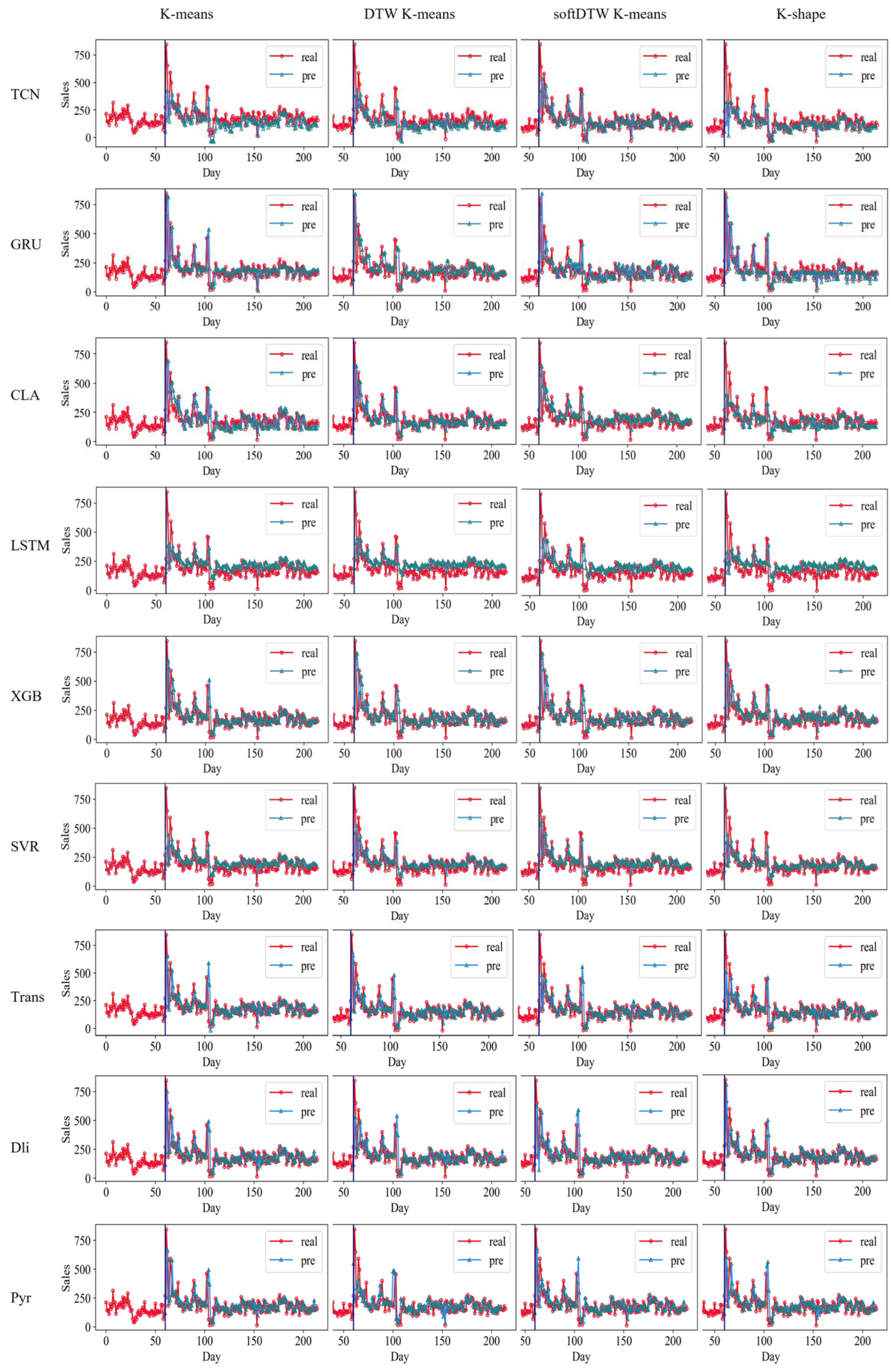
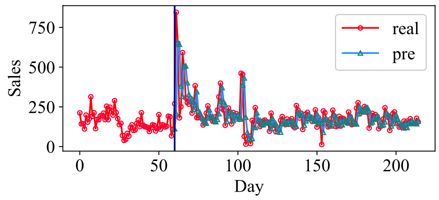


Fig. 11 Prediction results of different models. The prediction results of the proposed method and benchmarks are presented, where the legend real presents real data, and the legend pre presents prediction data. When $t = 60$, the large pandemic ends

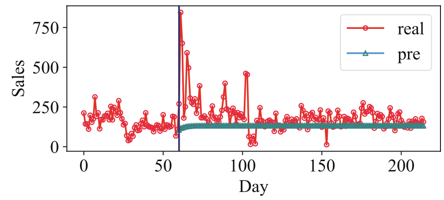
6.4.1 Clustering result

In our system, the *SSE* of clustering is shown in Fig. 9. From Fig. 9, it is observed that there is an inflection point when $K = 3$. Therefore, the clustering number K is determined to be 3.

In the experimental process, the sales data of different products are first normalized before clustering to address the problem of data magnitude differences. After that, the value of K is finally set to 3, and the clustering results are shown in Fig. 10 (line chart). In Fig. 10, the blue lines represent the sales curve of the items, and the orange lines represent the cluster centers. It can be visually observed in Fig. 10 that the product sales curves within the different clusters differ significantly. Additionally, the sale patterns within a cluster are highly similar.



(a) DKCLA



(b) ARIMA

Fig. 12 Prediction results of different models. The results of DKCLA and ARIMA predictions are shown, where the legend real presents real data, and the legend pre presents prediction data. The red line in the graph shows the actual sales data, and the blue line shows the forecast data. When $t = 60$, the large pandemic ends

Table 6 Predicted evaluation metrics of different cluster-based models

	K-means			DTW K-means			softDTW K-means			K-shape		
	RMSE	SMAPE	MAE	RMSE	SMAPE	MAE	RMSE	SMAPE	MAE	RMSE	SMAPE	MAE
Trans	10.19	30.08	59.56	9.91	28.27	56.94	10.08	28.52	58.40	9.74	28.44	56.31
TCN	10.20	39.05	67.79	9.65	28.02	61.79	9.59	28.33	57.94	9.73	27.51	54.84
GRU	9.96	32.11	56.64	10.19	**19.86**	66.97	10.00	24.33	59.79	10.38	30.62	63.62
CLA	9.82	28.51	55.23	**9.44**	**23.73**	**52.77**	9.84	26.02	56.56	9.95	28.48	64.83
LSTM	9.83	33.42	67.77	9.78	33.49	67.79	9.87	32.76	66.95	10.13	37.02	76.01
XGB	10.03	28.47	58.09	10.17	29.76	62.35	10.52	30.27	63.79	10.07	30.19	62.43
SVR	9.81	31.06	62.73	9.61	28.87	57.72	9.69	29.65	59.52	9.62	27.47	54.75
DLi	10.08	29.36	60.15	9.91	28.56	58.56	10.21	29.57	60.59	9.88	28.93	58.72
Pyr	10.05	28.84	58.95	9.46	28.60	54.58	10.10	29.02	58.69	9.96	29.18	58.86

6.4.2 Prediction result

After the parameters of each model have been determined, the prediction results of the DKCLA and other benchmarks are displayed in Figs. 11 and 12, where the legend real presents real data and the legend pre presents prediction data. The red line in the graph shows the actual sales data, and the blue line shows the forecast data. The XGB, Trans, DLi, and Pyr in the figure represent XGBoost, Transformer, DLinear, and Pyraformer, respectively. When $t = 60$, the large pandemic ends. From Fig. 11, it can be seen that all the cluster-based prediction algorithms can capture the surge in sales after the pandemic ends. Differently, our proposed DKCLA achieves more accurate sales predictions than other cluster-based methods.

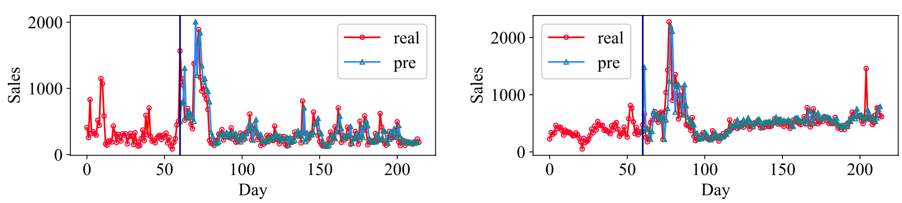
In Fig. 12, we also compare the proposed DKCLA with the ARIMA, a traditional time series prediction method. It is evident that the single ARIMA model generates poor predictions considering all prediction evaluation metrics. Since the ARIMA model relies on sales data for a single product, the predictive information is limited to historical data for that specific product. However, the impact of the pandemic can cause relatively large fluctuations in the data. Therefore, the sales forecasting method utilizing clustering techniques can yield significantly superior results compared to the conventional approach relying solely on internal data. We can conclude that the proposed DKCLA model is more robust to the impact of the pandemic and shows superior predictive performance.

The specific mean values of the experimental data of the prediction evaluation metrics for each model in this study are summarized in Table 6. This table indicates that the forecasting results of the DKCLA model outperform those of the other cluster-based forecasting methods in terms of all prediction evaluation metrics. The results reflected by the forecasting indicators show that DKCLA is significantly more effective than the other models. By the combination of Fig. 12 and Table 6, it can be seen that DKCLA is superior to other models in terms of forecasting effect. For the rapid increase in sales after the pandemic ends, the predicted values of DKCLA can quickly reflect and are closest to the actual sales. The DKCLA model proposed in this paper can be used to handle the situation of a rapid increase in sales after the pandemic.

Table 6 presents the predictive evaluation index values of 36 methodologies achieved by integrating four clustering techniques with nine prediction approaches. Among them, DKCLA exhibits the most favorable performance regarding *RMSE* and *MAE*, yielding values of 9.44 and 52.77, respectively. The method combining DTW K-means and GRU yields the lowest *SMAPE* value, 19.86%, closely followed by DKCLA's corresponding *SMAPE* metric. Overall, based on various evaluation metrics, our proposed DKCLA outperforms other baseline methods significantly. When keeping the prediction method constant, a comparison solely between the four clustering methods reveals that the DTW K-means clustering method exhibits superior effectiveness compared to the other three methods. DTW is a commonly employed method for calculating distances in clustering time series data. The simplicity of computation and clarity of interpretation have contributed to the widespread adoption of ED. However, it fails to capture temporal patterns within time series data. It only focuses on the absolute disparities among data points, disregarding their temporal alignment and shape variances. The experimental results demonstrate that the DTW k-means clustering method outperforms K-shape, indicating the superiority of the DTW distance-based clustering approach over shape-based clustering algorithms in the prediction stage. Similarly, when maintaining the clustering method unchanged, a comparison of the nine prediction methods reveals that CLA>Pyraformer>TCN>Transformer>DLinear>GRU>XGBoost>LSTM>SVR in terms of prediction accuracy. The CNN module can extract local patterns and short-term features more effectively through local feature extraction. Furthermore, it can reduce the noise and instability factors in the sales data. The LSTM layer can effectively capture long-term dependencies and temporal

Table 7 Predicted evaluation metrics of different cluster-based models

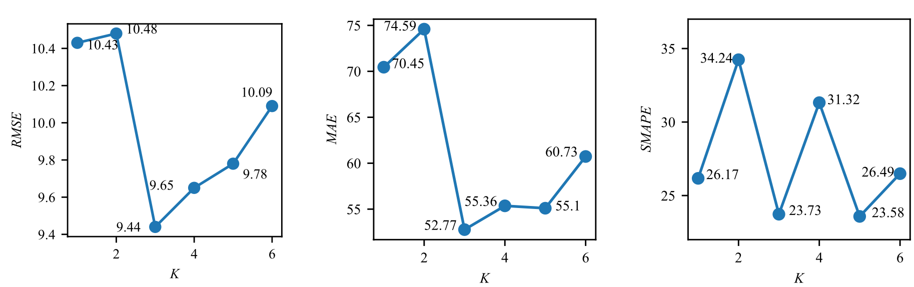
	K-means			DTW K-means			softDTW K-means			K-shape		
	RMSE	SMAPE	MAE	RMSE	SMAPE	MAE	RMSE	SMAPE	MAE	RMSE	SMAPE	MAE
Trans	11.45	30.15	61.33	10.89	27.31	59.32	10.98	29.58	60.57	10.35	30.14	60.96
TCN	11.14	32.98	64.84	10.74	**26.17**	60.15	10.77	29.65	59.91	10.62	29.89	58.88
GRU	10.69	31.93	60.18	10.92	27.86	62.98	11.49	27.20	60.80	11.69	30.41	63.25
CLA	10.74	29.32	59.20	**10.28**	26.35	**57.69**	10.80	28.33	59.62	10.65	29.40	64.08
LSTM	11.00	32.13	64.92	10.60	32.48	64.22	10.55	33.59	63.24	11.51	33.19	67.78
XGB	11.76	29.82	61.59	11.53	29.24	62.13	11.51	29.77	63.18	11.52	29.94	61.93
SVR	10.82	31.14	62.63	10.96	29.41	60.77	10.35	29.15	60.60	10.07	29.65	58.36
Dli	10.51	30.78	61.05	10.67	28.39	60.93	10.83	29.69	61.64	10.81	30.31	61.18
Pyr	10.27	30.12	60.48	10.38	27.82	60.62	10.25	29.35	59.91	10.26	29.38	59.12



(a) Shijiazhuang (b) Shanghai

Fig. 13 Prediction results of products in different cities. Prediction results for products in Shijiazhuang and Shanghai are shown, where the legend real presents real data, and the legend pre presents prediction data. The red line in the graph shows the actual sales data, and the blue line shows the forecast data. When t = 60, the large pandemic ends

Fig. 14 Evaluation metrics under different K values. Figure 14 visualizes the impact of changes in K values on different metrics. It is found that as the value of K increases, different evaluation indicators show a tendency to decrease and later increase, reaching the minimum value when $K = 3$



(a) RMSE (b) MAE (c) SMAPE

dynamics in sales data. It can better handle nonlinear and non-stationary time series data. Through the LSTM layer, the model can learn the complex dynamic changes in sales data before and after the pandemic, such as the surge in sales after the pandemic, which provides a basis for accurately predicting sales after the pandemic. By inputting the output of CNN into LSTM, a refined representation containing both local and global features can be provided. This enhanced feature extraction can bring more robust and informative inputs to the LSTM algorithm, ultimately improving its performance. The attention layer is connected behind the LSTM layer, the sales data can be processed in a weighted sum manner, and different attention weights are added to the sales. It enables the CNN-LSTM model to dynamically allocate its focus to the most relevant parts of the input sequence. By assigning higher weights to important time steps or features in sales data, the model could effectively filter out noise and pay attention to the most informative elements. Therefore, it can better mine and enhance the characteristics and change information of sales data and has higher prediction accuracy. The DKCLA model has both a CNN-LSTM module and a self-attention mechanism, so it can not only effectively extract data features but also comprehensively mine the potential features and discover the rules of sales data. The comparative results demonstrate the superiority of the proposed DKCLA model.

7 Discussion

In this section, we analyzed the stability and flexibility of the proposed model. Specifically, we studied the robustness of the model in Sect. 7.1, and the sensitivity of parameters affecting the model performance was analyzed in Sect. 7.2.

7.1 Robustness of the proposed method

The proposed model achieves satisfactory performance when using data from Shanghai, Xi'an, and Shijiazhuang as training data and Jilin data as testing data. To demonstrate the robustness of the proposed method, we conducted experiments using Shanghai, Xi'an, and Jilin as training data and Shijiazhuang as testing data. The experimental results are shown in Table 7. It shows that DKCLA performs most favorably on RMSE and MAE, with yields of 10.28 and 57.69, respectively. The combination of DTW K-means and TCN methods resulted in the lowest SMAPE value of 26.17%, followed closely by the corresponding SMAPE of DKCLA. Overall, after replacing the test and training sets, our proposed DKCLA still significantly outperforms other baseline methods based on various evaluation metrics.

Figure 13 illustrates the model prediction results using Shijiazhuang as the test set. Similarly, we conduct experiments using products data from Shanghai as the test set and obtained the results shown in Fig. 13. Figure 13 demonstrates that whether using Shijiazhuang as the test set or Shanghai as the test set, the model can capture the surge in sales after the pandemic and obtain favorable prediction results. Therefore, through Table 7 and Fig. 13, we can demonstrate the robustness of the proposed model.

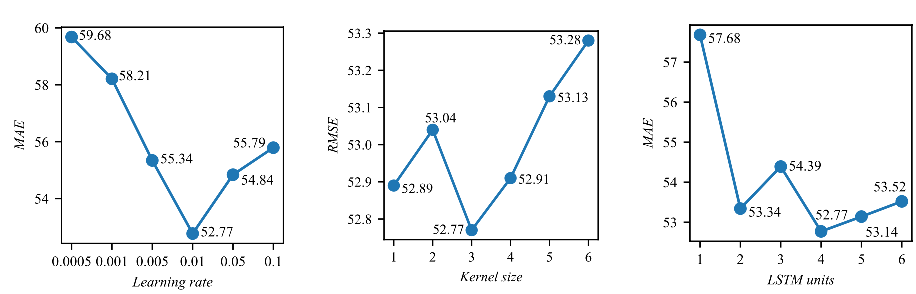
7.2 Sensitivity analysis

7.2.1 Prediction results for different numbers of clusters

The number of clusters K is a key parameter of the DTW K-means method. Therefore, we conducted sensitivity analysis on it. We conducted experiments to verify whether clustering can improve the prediction performance and whether the number of clusters K is a sensitive parameter of the model. For the DKCLA model, $K = \{0, 1, \cdots, 6\}$ are set. Figure 14 visualizes the impact of changes in K values on different metrics.

As given in Fig. 14, it is found that the minimum values of *RMSE* and *MAE* are obtained when $K = 3$. The *SMAPE* at $K = 3$ is slightly greater than 5. Therefore, it is considered that the best predictions can be made when $K = 3$. We found that as K increases, different evaluation indicators tend to decrease and later increase, reaching the minimum value when $K = 4$ or 5. $K = 1$ means there is no clustering. It can be observed that the cluster-based forecasting effects are mostly better than the effect of no clustering ($K = 1$) for

Fig. 15 MAE of different hyperparameters. Sensitivity analysis of hyperparameters is demonstrated, including learning rate, kernel size, and number of LSTM units



(a) Learning rate (b) Kernel size (c) LSTM units most prediction evaluation metrics. For example, the *RMSE* value (9.44) when $K = 3$ is the lowest and lower than the value (10.43) when $K = 1$. It can be observed that the prediction performance is poorest when $K = 2$ and the number of clusters significantly impacts *SMAPE*. This means that the value of clusters K is a crucial parameter in cluster-based forecasting models and directly impacts the final prediction results. The model is sensitive to changes in the number of clusters K. The reason is that if the number of clusters is too small, similar sales patterns of products may not be accurately identified, and thus, data noise reduction cannot be achieved. However, if the number of clusters is too large, each sub-data set will have less data, which can negatively impact the accuracy of the final prediction model. Therefore, a cluster-based prediction model is necessary, and the choice of K is important.

7.2.2 Prediction results for different hyperparameters

By the single control variable method, we systematically conducted a sensitivity analysis aimed at exploring the impact of various hyperparameters on the model's performance. This encompassed learning rate, convolutional layer parameters, specifically, the kernel size, as well as LSTM layer parameters such as the number of LSTM units. The analysis results are presented in Fig. 15.

Figure 15a gives the *MAE* for different learning rates for training the DKCLA model. As the learning rate increases from 0.0005 to 0.01, the *MAE* decreases from 59.68 to 52.77, which indicates that a lower learning rate is not conducive to the model improvement due to the local optimization problem. After that, when the learning rate increases from 0.005 to 0.1, the *MAE* tends to increase, which may be a result of poorer convergence performance at higher learning rates. Among these choices, a learning rate value of 0.01 results in the smallest *MAE* value of 52.77. The large variation in *MAE* indicates that the learning rate is a sensitive parameter for the proposed model. Figure 15b represents the relationship between *MAE* and hyperparameter *kernel size*. We observe that both too large and too small convolution kernels could decrease prediction accuracy, mainly because too small convolution kernels may not represent features well, while larger convolution kernels can increase computational complexity. It is worth noting that the decrease magnitude in *MAE* is less than 0.51, indicating that the size of the convolution kernel has a relatively small impact on *MAE*. Therefore, the size of the convolution kernel can be excluded from sensitive parameters. In Fig. 15c, as the number of *LSTM unit*s increases to 4, the *MAE* fluctuates greatly, decreasing from 57.68 to 52.77. After that, there is a slight increase. At the *LSTM unit* = 4, the model exhibits the best performance. The LSTM layer can effectively capture long-term dependencies and temporal dynamics in time series data. In this case, the underlying pattern of sales data is successfully captured by 4 LSTM units. From the above analysis, it can be seen that the *LSTM unit* is also a sensitive parameter of the proposed model. In summary, the *learning rate* and the number of *LSTM units* play a significant role, while the size of the convolution kernel has less impact on model performance.

8 Conclusion

In this paper, we propose a cluster-based forecasting model, DKCLA, to predict the sales of commodities in a new region after a large pandemic. DTW K-means and CNN-LSTM-Attention are employed to construct DKCLA. In addition, real sales data from a specific E-commerce platform are utilized to verify the effectiveness of the proposed DKCLA model. Specifically, sales data from Shanghai, Xi'an, and Shijiazhuang are used to construct the training set, while sales data from Jilin are used to build the testing set. The experimental results indicate that the proposed DKCLA forecasting model is superior to the ARIMA and the additional 35 clustering-based prediction approaches regarding forecasting accuracy. Moreover, it is observed that the cluster-based prediction model outperforms the non-clustered prediction model in predicting product sales after the pandemic. Furthermore, the number of clusters is found to have a critical impact on the prediction accuracy of the DKCLA model. The *learning rate* and the number of *LSTM units* play a significant role, while the size of the convolution kernel has less impact on model performance. Overall, the DKCLA model proposed in this paper is an effective sales forecasting model that is suitable for product sales forecasting in a new city after the pandemic ends. And it can deal with market uncertainties and fluctuations, and provide accurate forecasts for product sales with different demand patterns. For products with significant and minor fluctuations in sales, we will adopt a unified approach and include it in future work.

Author Contributions All authors contributed to the study conception and design. The first draft of the manuscript was written by Zhaolin Lv and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding The authors did not receive support from any organization for the submitted work.

Data availability The data that support the findings of this study are available from JD Logist but restrictions apply to the availability of these data, which were used under licence for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of JD Logist.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Ethics approval Written informed consent for publication of this paper was obtained from the National University of Defense Technology of Systems Engineering College and all authors.

Consent to participate All authors agree.

Consent for publication All authors agree.

References

1. Branley-Bell D, Talbot C.V (2020) Exploring the impact of the covid-19 pandemic and uk lockdown on individuals with experience of eating disorders. JOURNAL OF EATING DISORDERS **8**(1) https://doi.org/10.1186/s40337-020-00319-y
2. Rohaan D, Topan E, Groothuis-Oudshoorn CGM (2022) Using supervised machine learning for b2b sales forecasting: a case study of spare parts sales forecasting at an after-sales service provider. Expert Syst Appl 188:115925. https://doi.org/10.1016/j.eswa.2021.115925
3. Migueis VL, Pereira A, Pereira J, Figueira G (2022) Reducing fresh fish waste while ensuring availability: Demand forecast using censored data and machine learning. JOURNAL OF CLEANER PRODUCTION 359. https://doi.org/10.1016/j.jclepro.2022.131852
4. Weng T, Liu W, Xiao J (2020) Supply chain sales forecasting based on lightgbm and lstm combination model. Ind Manag Data Syst 120(2, SI):265–279. https://doi.org/10.1108/IMDS-03-2019-0170
5. Wan A, Chang Q, AL-Bukhaiti K, He J, (2023) Short-term power load forecasting for combined heat and power using cnn-lstm enhanced by attention mechanism. ENERGY 282. https://doi.org/10.1016/j.energy.2023.128274
6. Andueza A, Del Arco-Osuna MA, Fornes B, Gonzalez-Crespo R, Martin-Alvarez J-M (2023) Using the statistical machine learning models arima and sarima to measure the impact of covid-19 on official provincial sales of cigarettes in spain. Int J Interactive Multimedia Artificial Intell 8(1):73–87. https://doi.org/10.9781/ijimai.2023.02.010
7. Korkmaz M (2021) A study over the general formula of regression sum of squares in multiple linear regression. Numer Methods Par Differ Eqs 37(1):406–421. https://doi.org/10.1002/num.22533
8. Ikotun AM, Ezugwu AE, Abualigah L, Abuhaija B, Heming J (2023) K-means clustering algorithms: a comprehensive review, variants analysis, and advances in the era of big data. Inform Sci 622:178–210. https://doi.org/10.1016/j.ins.2022.11.139
9. Yang H, Wang L, Xu Y, Liu X (2023) Covidvit: a novel neural network with self-attention mechanism to detect covid-19 through x-ray images. Int J Mach Learn Cybernet 14(3):973–987. https://doi.org/10.1007/s13042-022-01676-7
10. Lin K, Zhao Y, Tian L, Zhao C, Zhang M, Zhou T (2021) Estimation of municipal solid waste amount based on one-dimension convolutional neural network and long short-term memory with attention mechanism model: A case study of shanghai. SCIENCE OF THE TOTAL ENVIRONMENT 791. https://doi.org/10.1016/j.scitotenv.2021.148088
11. Cui B, Liu M, Li S, Jin Z, Zeng Y, Lin X (2023) Deep learning methods for atmospheric pm2.5 prediction: A comparative study of transformer and cnn-lstm-attention. ATMOSPHERIC POLLUTION RESEARCH **14**(9) https://doi.org/10.1016/j.apr.2023.101833
12. Narang R, Singh U.P (2023) Interpretable sequence models for the sales forecasting task: A review. In: 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 858–864 https://doi.org/10.1109/ICICCS56967.2023.10142614
13. Saldaña-Olivas E, Huamán-Tuesta J.R (2021) Extreme learning machine for business sales forecasts: A systematic review. In: Iano, Y., Arthur, R., Saotome, O., Kemper, G., Padilha França, R. (eds.) Proceedings of the 5th Brazilian Technology Symposium, pp. 87–96. Springer, Cham
14. Mallik R.S, Abhiram R, Reddy S.R, Jagadish R.M (2022) A comprehensive survey on sales forecasting models using machine learning algorithms. In: 2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT), pp. 1–6 https://doi.org/10.1109/ICERECT56837.2022.10060168
15. Wu H, Levinson D (2021) The ensemble approach to forecasting: A review and synthesis. Transportation Research Part C: Emerging Technologies 132:103357. https://doi.org/10.1016/j.trc.2021.103357
16. Burinskiene A (2022) Forecasting model: The case of the pharmaceutical retail. FRONTIERS IN MEDICINE 9. https://doi.org/10.3389/fmed.2022.582186
17. Rezende R, Egert K, Marin I, Thompson G (2022) A white-boxed issm approach to estimate uncertainty distributions of walmart sales. Int J Forecasting 38(4, SI):1460–1467. https://doi.org/10.1016/j.ijforecast.2021.11.006
18. Migueis VL, Pereira A, Pereira J, Figueira G (2022) Reducing fresh fish waste while ensuring availability: Demand forecast using censored data and machine learning. J Clean Prod 359. https://doi.org/10.1016/j.jclepro.2022.131852
19. Qu F, Wang Y.-T, Hou W.-H, Zhou X.-Y, Wang X.-K, Li J.-B, Wang J.-Q (2022) Forecasting of automobile sales based on support vector regression optimized by the grey wolf optimizer algorithm. Mathematics **10**(13) https://doi.org/10.3390/math10132234
20. Panda SK, Mohanty SN (2023) Time series forecasting and modeling of food demand supply chain based on regressors analysis. IEEE Access 11:42679–42700. https://doi.org/10.1109/ACCESS.2023.3266275
21. Dairu X, Shilong Z (2021) Machine learning model for sales forecasting by using xgboost. In: 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), pp. 480–483 https://doi.org/10.1109/ICCECE51280.2021.9342304
22. Fildes R, Ma S, Kolassa S (2022) Retail forecasting: Research and practice. International Journal of Forecasting 38(4):1283–1318. https://doi.org/10.1016/j.ijforecast.2019.06.004. Special Issue: M5 competition
23. Xu J, Zhou Y, Zhang L, Wang J, Lefloch D (2021) Sportswear retailing forecast model based on the combination of multi-layer perceptron and convolutional neural network. Textile Res J 91(23–24):2980–2994. https://doi.org/10.1177/00405175211020518
24. Chandriah KK, Naraganahalli RV (2021) Rnn / lstm with modified adam optimizer in deep learning approach for automobile spare parts demand forecasting. Multimedia Tools Appl 80(17):26145–26159. https://doi.org/10.1007/s11042-021-10913-0
25. Han Y (2020) A forecasting method of pharmaceutical sales based on arima-lstm model. In: 2020 5th International Conference on Information Science, Computer Technology and Transportation(ISCTT 2020), pp. 336–339 https://doi.org/10.1109/ISCTT51595.2020.00064

26. Ampountolas A, Legg MP (2021) A segmented machine learning modeling approach of social media for predicting occupancy. Int J Contemporary Hospitality Manage 33(6, SI):2001–2021. https://doi.org/10.1108/IJCHM-06-2020-0611
27. Ekambaram V, Manglik K, Mukherjee S, Sajja S.S.K, Dwivedi S, Raykar V (2020) Attention based multi-modal new product sales time-series forecasting. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. KDD '20, pp. 3110–3118. Association for Computing Machinery, New York, NY, USA https://doi.org/10.1145/3394486.3403362
28. Li Y, Yang Y, Zhu K, Zhang J (2021) Clothing sale forecasting by a composite gru-prophet model with an attention mechanism. IEEE Trans Ind Inform 17(12):8335–8344. https://doi.org/10.1109/TII.2021.3057922
29. Gandhi A, Aakanksha Kaveri S, Chaoji V (2021) Spatio-temporal multi-graph networks for demand forecasting in online marketplaces. In: Dong, Y., Kourtellis, N., Hammer, B., Lozano, J.A. (eds.) Machine Learning and Knowledge Discovery in Databases. Applied Data Science Track, pp. 187–203. Springer, Cham
30. Ye X, Ye Q, Yan X, Wang T, Chen J, Li S (2021) Demand forecasting of online car-hailing with combining lstm + attention approaches. Electronics **10**(20) https://doi.org/10.3390/electronics10202480
31. Wang H, Zhang Z (2023) Hierarchical time series forecasting based on temporal convolution and attention mechanism. In: Zhang S, Zhang Y (eds) Artificial Intelligence Logic and Applications. Springer, Singapore, pp 403–410
32. Yang Y, Lu J (2023) Foreformer: an enhanced transformer-based framework for multivariate time series forecasting. Appl Intell 53(10):12521–12540
33. Yu Y, Ma R, Ma Z (2024) Robformer: a robust decomposition transformer for long-term time series forecasting. Pattern Recognit 153:110552
34. Li Y, Yang Y, Zhu K, Zhang J (2021) Clothing sale forecasting by a composite gru-prophet model with an attention mechanism. IEEE Trans Ind Inform 17(12):8335–8344. https://doi.org/10.1109/TII.2021.3057922
35. El Amouri H, Lampert T, Gancarski P, Mallet C (2023) Constrained dtw preserving shapelets for explainable time-series clustering. Pattern Recognit 143. https://doi.org/10.1016/j.patcog.2023.109804
36. Li Q, Zhang X, Ma T, Liu D, Wang H, Hu W (2022) A multi-step ahead photovoltaic power forecasting model based on timegan, soft dtw-based k-medoids clustering, and a cnn-gru hybrid neural network. Energy Rep 8:10346–10362. https://doi.org/10.1016/j.egyr.2022.08.180
37. Wang X, Song R, Xiao J, Li T, Li X (2023) Accelerating k-shape time series clustering algorithm using gpu. IEEE Trans Parallel Distributed Syst 34(10):2718–2734. https://doi.org/10.1109/TPDS.2023.3298148
38. Yang L, Zhang Z (2022) A deep attention convolutional recurrent network assisted by k-shape clustering and enhanced memory for short term wind speed predictions. IEEE Trans Sustain Energy 13(2):856–867. https://doi.org/10.1109/TSTE.2021.3135278
39. Hadjout D, Sebaa A, Torres JF, Martinez-Alvarez F (2023) Electricity consumption forecasting with outliers handling based on clustering and deep learning with application to the algerian market. EXPERT SYSTEMS WITH APPLICATIONS 227. https://doi.org/10.1016/j.eswa.2023.120123
40. Chen Z, Chen Y, Xiao T, Wang H, Hou P (2021) A novel short-term load forecasting framework based on time-series clustering and early classification algorithm. Energy Build 251. https://doi.org/10.1016/j.enbuild.2021.111375
41. Chen I.-F, Lu C.-J (2021) Demand forecasting for multichannel fashion retailers by integrating clustering and machine learning algorithms. Processes **9**(9) https://doi.org/10.3390/pr9091578
42. Steenbergen RM, Mes MRK (2020) Forecasting demand profiles of new products. Decision Support Syst 139. https://doi.org/10.1016/j.dss.2020.113401
43. Malarya A, Ragunathan K, Kamaraj M.B, Vijayarajan V (2021) Emerging trends demand forecast using dynamic time warping. In: 2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI), pp. 402–407 https://doi.org/10.1109/IRI51335.2021.00063
44. Li X, Yin Y, Manrique DV, Back T (2021) Lifecycle forecast for consumer technology products with limited sales data. Int J Prod Econ239. https://doi.org/10.1016/j.ijpe.2021.108206
45. Sleiman R, Mazyad A, Hamad M, Tran K-P, Thomassey S (2022) Forecasting sales profiles of products in an exceptional context: Covid-19 pandemic. Int J Comput Intell Syst 15(1):99
46. Park S, Lin K, Joung J, Kim H (2023) Investigation of customer preference changes following covid-19 market disruption using online review analysis. Proc Design Soc 3:2375–2384
47. Gimbach S, Vogel D, Fried R, Faraone SV, Banaschewski T, Buitelaar J, Doepfner M, Ammer R (2023) The impact of the covid-19 pandemic on adhd medicine consumption in 47 countries and regions. Euro Neuropsychopharmacol 73:24–35. https://doi.org/10.1016/j.euroneuro.2023.04.008
48. Andueza A, Del Arco-Osuna MA, Fornes B, Gonzalez-Crespo R, Martin-Alvarez J-M (2023) Using the statistical machine learning models arima and sarima to measure the impact of covid-19 on official provincial sales of cigarettes in spain. Int J Interactive Multimedia Artificial Intell 8(1):73–87. https://doi.org/10.9781/ijimai.2023.02.010
49. Kim H.-J, Kim J.-H, Im J.-b (2023) Forecasting offline retail sales in the covid-19 pandemic period: A case study of a complex shopping mall in south korea. Buildings **13**(3) https://doi.org/10.3390/buildings13030627
50. Tudor C (2022) Integrated framework to assess the extent of the pandemic impact on the size and structure of the e-commerce retail sales sector and forecast retail trade e-commerce. Electronics **11**(19) https://doi.org/10.3390/electronics11193194
51. Bagirov AM, Aliguliyev RM, Sultanova N (2023) Finding compact and well-separated clusters: Clustering using silhouette coefficients. Pattern Recognit 135. https://doi.org/10.1016/j.patcog.2022.109144
52. Pan Q, Wang H, Tang J, Lv Z, Wang Z, Wu X, Ruan Y, Yv T, Lao M (2024) Eioa: A computing expectation-based influence evaluation method in weighted hypergraphs. Inform Process Manag 61(6):103856. https://doi.org/10.1016/j.ipm.2024.103856
53. Chicco D, Warrens MJ, Jurman G (2021) The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. PEERJ Comput Sci. https://doi.org/10.7717/peerj-cs.623