https://doi.org/10.1007/s43684-024-00078-6

Autonomous Intelligent Systems

ORIGINAL ARTICLE Open Access

A multi-step regularity assessment and joint prediction system for ordering time series based on entropy and deep learning

Yichen Zhou[1,2], Wenhe Han[1,2*] and Heng Zhou[1,2]

Abstract

Customer maintenance is of vital importance to the enterprise management. Valuable assessment and efficient prediction for customer ordering behavior can offer better decision-making and reduce business costs significantly. According to existing studies about customer behavior regularity segment and demand prediction most focus on e-commerce and other fields with large amount of data, making them not suitable for small enterprises and data features like sparsity and outliers are not mined when doing regularity quantification. Additionally, more and more complex network structures for demand prediction are proposed, which builds on the assumption that all the samples have predictive value, ignoring the fine-grained analysis of different time series regularity with high cost. To deal with the above issues, a multi-step regularity assessment and joint prediction system for ordering time series is proposed. For extracting features, comprehensive assessment of customer regularity based on entropy weight method with the result of predictability quantification using K-Means clustering algorithm, real entropy, LZW algorithm and anomaly detection adopting Isolation Forest algorithm not only gives an objective result to 'how high the regularity of customers is', filling the gap in the field of regularity quantification, but also provides a theoretical basis for demand prediction models selection. Prediction models: Random Forest regression, XGBoost, CNN and LSTM network are experimented with sMAPE and MSLE for performance evaluation to verify the effectiveness of the proposed regularity quantitation method. Moreover, a merged CNN-BiLSTM neural network model is established for predicting those customers with low regularity and difficult to predict by traditional machine leaning algorithms, which performs better on the data set compared to others. Random Forest is still used for prediction of customers with high regularity due to its high training efficiency. Finally, the results of prediction, regularity quantification, and classification are output from the intelligent system, which is capable of providing scientific basis for corporate strategy decision and has highly extendibility in other enterprises and fields for follow-up research.

Keywords: Regularity quantification, Customer assessment, Time series prediction, Machine learning, Deep learning

1 Introduction

Customer is the core asset of an enterprise, whose retention and loss will directly affect the market share [1], so the maintenance of customers is crucial to the bottom line of enterprise profits. However, the resources of an enterprise are often limited, and it is impossible to carry out unified and good management of all customers. Therefore, customer segment and assessment have far-reaching significance for enterprises to increase profits. With the rapid development of the Internet industry and big data technology, many new solutions have been proposed for

*Correspondence: hanwenhe@sribs.com
[1]Science and Technology Development Co., Ltd. of Shanghai Research Institute of Building Sciences, Shanghai 201506, China
[2]Shanghai Engineering Research Center for Safety Intelligent Control of Building Machinery, Shanghai 200032, China

the customer management and the prediction of enterprise market demand. Traditional manufacturing enterprises are also in urgent need of innovation driven by data elements on the original business model to improve efficiency and revenue for enterprises. Although there are many related research methods in customer segmentation, most of them use historical sales trend and related basic statistical data characteristics in customer description. The classification standards are basically focused on the statistical calculation results, which are more subjective and interpretative, focusing on the understand of 'what king of regularity' [2]. While in the trading history of many manufacturing enterprises, customer demand sometimes occurs randomly, accompanied by a large number of zero demand periods and only the same large customers have continuous demand, so it leads to the intermittent and distribution with lump formation characteristics, making it difficult to extract sufficient fluctuation rules from such time sequences, which can affect the prediction performance [3]. How to extract the internal evolution regularity from these time series and answer the problem of 'how regular they are' is an urgent demand in the management of manufacturing enterprises that has highly theoretical research value. The concept of predictability originates from the prediction problem based on dynamical systems [4]. Usually the essential predictability describes the prediction accuracy achieved when the optimal prediction method is used in the system [5, 6]. As can be seen, predictability is tend to give a quantified result of 'upper limit of prediction effect' and reflect how much information in the current time series data is valuable for predicting future state. Specifically, in the time series prediction, the random noise, finiteness, and the approximate characterization ability of models make it unable to achieve completely accurate prediction [7]. It can be known that predictability, independent of the prediction models, characterizes the properties of the data, which is able to measure the quality and improvement potential of different models [8]. Additionally, predictability can be used to pre-process multiple time series before modeling, such as screening out poorly predictable sequences and finding the potential structure of data [9]. At present many scholars used various indicators to quantity the abstract concept of predictability including exponent in dynamical system [10], information theory related predictive information, wavelet entropy energy measure, entropy-based indicators of time series complexity, and so on [11, 12].

Currently time series prediction is mainly based on various prediction methods and data mining models. Statistical learning methods such as exponential smoothing [13] and moving average [14], indicator describing time sequence like Average Demand Interval (ADI) and Coefficient of Variation (CV^2) [15], and hierarchical clustering of sequences are commonly used for time series prediction. With the development of data mining, machine learning and deep learning models to predict customer demand and the long-term sales trend as prevailing tools have been widely researched in various fields, including Support Vector Regression (SVR), Extreme Gradient Boosting (XGBoost), Random Forest Regression (RFR), Long Short-Term Memory neural network (LSTM), and Convolutional Neural network (CNN). By collecting a great number of various customer features, customer demand prediction is established and early waning of customers being lost is given to help enterprises' decision-making in advance.

On the whole, given the circumstances that the collected data in this paper comes from an industrial product manufacturing enterprise, whose customers are mostly concentrated in the construction machinery industry, so the data source is relatively single and it is difficult to refer to the multiple characteristics used for the description of customers in e-commerce, finance, transportation and other industries. Although the related research of the above methods have achieved certain results, there are still some limitations: 1) The predictability methods only consider the statistical characteristics of time series, other data features such as sparsity and outliers could not be mined; 2) Most researches only focus on improving the performance of one prediction model by proposing a complex network structure, which builds on the assumption that all the data have predictive value, ignoring the fine-grained analysis of time series regularity. In fact, many customers only have several requirements in one year and the distribution of time series data is extremely sparse, which has no predictable value, so they should be filtered out through preliminary data analysis method or predicted by simple prediction models. To sum up, the main contributions of this paper are: 1) A comprehensive assessment method combined with predictability quantification and anomaly detection is proposed. This quantitative analysis of customer ordering regularity based on historical ordering data gives another new angle for customer behavior analysis, which can enrich the quantitative analysis method in customer segmentation and provide a theoretical basis for model selection to improve overall prediction efficiency, so it has certain theoretical and practical significance; 2) The CNN-BiLSTM merged network is introduced which has higher prediction accuracy; 3) A multi-step regularity assessment and joint prediction system for ordering time series is proposed which has significantly better efficiency and performance so that the enterprise can make better decisions in customer maintenance; 4) The analysis method adopted in this paper has lower requirements on data sources, which is closer to the actual business of enterprises, so it can be popularized well. The rest of this paper is organized as follows: the method reviewed in Sect. 2. Section 3 introduces the regularity quantification and the prediction model. The experimental study is stated in Sect. 4 and the results are discussed in Sect. 5. Finally, Sect. 6 concludes.

2 Review methodology

At present, there are many studies on the relationship between management of customer value and customer segmentation by domestic and foreign scholars, but most of them focus on the application of e-commerce retail, logistics, transportation, finance and other industries or enterprises. Dwyer proposed the calculation model of Customer Life Value (CLV) [16]. PCI model was used to verify the cluster number through sales of an industrial product, then AHP was used to sort the weights. This method measured customers as three type: superstar customers, typical customers and dormant customers [17]. To evaluate the impact of different categories of coupons on short-term marketing costs and long-term customer lifetime value, a target strategy in the framework of customer segmentation based on customer churn, frequency and loyalty by studying customer data of large supermarkets [18]. Four types of customer value matrices were proposed for customer value segmentation and retention strategy through the study of four combinations current value and potential value of customers [19]. With the application effect of classical probability models represented by Pareto/NBD and BG/NBD as well as machine learning models represented by SVM in the measurement method of noncontract customers' lifetime value, a comprehensive measurement which was more applicable based on classical methods and machine learning algorithm was proposed [20]. A series of countermeasures for score and ranking of customers through a quantitative and qualitative evaluation method of customer value hierarchy. These indicators included three first-level indicators: customer contribution, development potential, and own strength, and twelve second-level indicators of plan implementation rate [21]. Improved RFM model and K-Means algorithm were used to classify customers in the aviation market and corresponding marketing strategies for them was proposed [22]. It can be found that most customer segment and indicator establishing methods still focus on answering the problem 'what king of regularity', lacking specific quantitative methods to achieve objective classification, so predictability is taken into account in this paper for giving the answer of 'how regular customers are'.

Considering that the predictability cannot be calculated by traversal of all prediction models, various methods have been used to quantify this abstract concept, which can mainly be classified as time series complexity and entropy. Complexity is a very common concept in dynamical system and different time series research have different views of characterization [23]. There are many indicators used to characterize the complexity, including Lyapunov exponent, fractal dimension, Lempel-Ziv complexity, similarity index, and Coefficient of Variation (CV) [24]. Different measures are suitable for time series with different characteristics and have different calculation complexities. Boffetta et al. believed that there was a close relationship between predictability and the complexity of time series [25]. Moreover, information measures based on entropy have received much attention in predictability studies [26], such as approximate entropy, sample entropy, permutation entropy, fuzzy entropy, weighted permutation entropy, and so on [27–30]. Garland et al. used weighted permutation entropy as an indicator to measure the time series complexity and four prediction models to predict a variety of different data sets. They analyzed the relationship between weighted permutation entropy and the minimum prediction accuracy of four methods to determine whether the prediction model and time series is matched [31]. Yang et al. proved with an example that approximate entropy can effectively describe the predictability of wind power time series [32]. Xu et al. used the complexity of travel time series with multi-scale entropy to study predictability [33]. Lin et al. adopted information entropy to measure the predictability of cellular flow under different models [34]. As can be seen that the entropy-based methods now have been the mainstream quantitative measure of predictability.

In terms of sales predicting research, most research based on long-term sales trend are a typical time series forecasting problem [35]. Traditional statistical methods are widely used by enterprises because of their simplicity and availability [36], such as moving average, HW method, auto-regressive integrated moving average model, and so on. In order to overcome the problem that traditional models cannot capture nonlinear features, many predicting methods adopting machine learning and deep learning have emerged, such as support vector machine (SVM), random forest (RF), GBDT, LSTM, and so on. Machine learning is a part of artificial intelligence, which is widely used in data analysis, data mining, and prediction. A model based on random forest was proposed to predict the spare parts demand to verify the performance in predicting intermittent time series [37]. Yang et al. used SVR regression model to optimize the regression parameters through the genetic algorithm, which can effectively capture the non-linear characteristics of the short-term passenger flow [38]. SVM-KNN model is used to predict the bus passenger flow with influencing factors [39]. Caigny et al. adopted the combination of decision tree and logistic regression, which ensured the performance of classification and had good interpretability [40]. Deep learning now has become an increasing important part of data mining and prediction research with the development of technology [41]. Recurrent neural network (RNN) is popular for time series data processing. A method based on competitive learning to predict long-term machine health status is proposed by Malhi et al. [42]. LSTM network is an RNN, which improves RNN and is able to solve the problem of long-term memory that RNN cannot solve. An LSTM-based model to predict temperature sensitive products sales has achieved excellent results [43]. Zhang et al. regarded customer behavior as a temporal behavior, using LSTM network to predict user loss. The results were comparable to random forest model and better than logistic regression model [44]. On the other hand, CNN is another classical deep learning network, which is found that can effectively reduce the complexity of a feedback neural network. CNN is always a research hotspot in many fields, especially pattern classification, which has many similarities with the traditional neural network. CNN is successful in reducing the dimension of image recognition which has a great number of data [45]. CNN discovered the spatial correlation between adjacent intersections for its local perception and weight sharing and CNN-LSTM network is capable of extracting the spatial-temporal characteristics of passenger flow [46, 47]. Chen et al. used CNN network to mine the internal correlation between meteorological parameters and used LSTM network to mine historical photovoltaic power timing information, which improved the output power prediction performance [48]. Chen et al. proposed a CNN-LSTM network to process the short-term passenger flows with spatial and temporal characteristics [49]. Shi et al. selected case enterprise and elaborated the risk control activities carried out on the basis of a multi-value chain collaborative RF-BAS-CNN operating risk prediction model [50]. Huang et al. proposed a prediction model of customer attrition based on transformer and graph convolution to enhance the ability of the multi-layer perceptron fitting the sample [51]. At present, generative adversarial networks [52], graph neural networks [53], and multi-task learning are also the mainstream research methods for prediction in many fields.

3 Method

A deep learning-based regularity assessment and joint prediction system with multi-step data processing is proposed to assess customer ordering behavior in terms of regularity and establish customer demand prediction based on collected customer data, which consists of two main stages. Firstly, it is necessary to extract the features of the ordering data, including predictability quantification results and anomaly numbers, which decides the regularity classification of customer ordering behavior. Therefore, K-Means clustering is used to discrete the original data and LZW algorithm is set to calculate the real entropy of the customer ordering time series as the result of customer predictability quantification. Considering that the existence of outliers is a high-frequency phenomenon for real data, which has a great influence on regularity and prediction judgment, iForest(Isolation Forest) algorithm is adopted for anomaly detection that utilizes the separation property of data. Then the regularity classification result can be assessed through entropy weight method with the predictability quantification and outlier numbers as a whole. Before prediction, the original time series data is standardized by Rubust Standardization method which has good robustness to outliers, making the standardized data less susceptible to the influence of outliers. Sliding window and PCA dimensional reduction are used to get input array data. Secondly, according to the customer classification result, two prediction models are merged for customer demand prediction. Random Forest regression model is used for those customers with high regularity and the proposed CNN-BiLSTM network is used for customers with low regularity. Finally, the results of customer regularity quantification, classification and demand forecasting will be output from the system. The framework of the proposed system is shown in Fig. 1.

3.1 Entropy-based quantification of customer ordering regularity

3.1.1 Problem

Definition 1 The customer number is used to record each customer as $C^{(1)}, C^{(2)}, \ldots, C^{(m)}$, and month is chosen as the time granularity to calculate the monthly product order quantity, so the corresponding time series is recorded as $T^{(i)} = (t_1^{(i)}, t_2^{(i)}, \ldots, t_n^{(i)})$, here $t_j^{(i)} (j = 1, 2, \ldots, n)$ represents the product order quantity of customer $C^{(i)}$ in the month j.

Definition 2 (Quantification of customer ordering predictability) Given a time series of a customer's ordering history $T^{(i)}$, the proposed method in this paper aims to represent the ordering predictability in the form of numerical values, so as to provide a new quantitative index for describing the characteristics and classification of customers.

Through the quantification of customer ordering predictability, we can understand the characteristics of customer ordering behavior more intuitively and deeply, which can help solve the problem of comparability of different customer ordering regularity to answer 'how regular it is'.

3.1.2 Customer ordering predictability quantification

The predictability of customer ordering is mainly reflected in its time dependence, while entropy is an important method to measure the state of some material systems, and it can be used as an effective method to represent the system predictability. In general, lower entropy means lower state uncertainty, which means higher predictability. This paper experiments three entropy measures: random entropy $E^{(i)}_{random}$, Shannon entropy $E^{(i)}_{Shannon}$, real entropy $E^{(i)}_{real}$. For a given customer, the corresponding time series $T^{(i)} = (t_1^{(i)}, t_2^{(i)}, \ldots, t_n^{(i)})$ is obtained after discretization. Here, $t_j^{(i)} (j = 1, 2, \ldots, n)$ represents the order size in month j. The calculation methods of the three entropy measures are as follows.

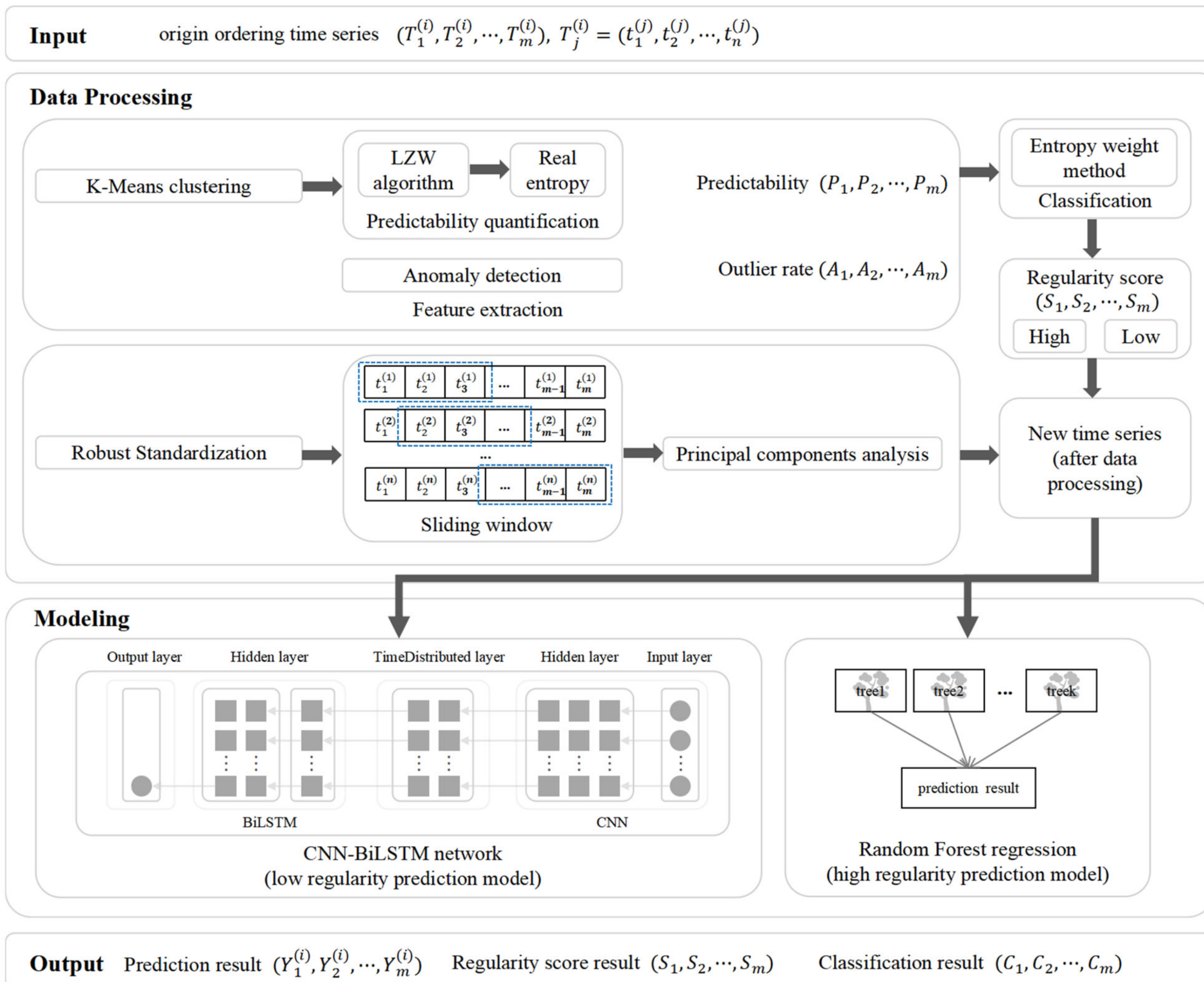

Figure 1 The framework of the proposed system

1) *Random entropy*

$$E_{random}^{(i)} = log_2 N^{(i)}. \tag{1}$$

$N^{(i)}$ represents the number of different categories of the order size in $T^{(i)}$, such as for $T^{(i)} = 2321$, which $N^{(i)} = 3$. Equation (1) shows that random entropy depends on $N^{(i)}$, which means the less a customer's order quantity scale categories, the lower the obtained random entropy, and the higher the ordering predictability.

2) *Shannon entropy* (*Information entropy*)

In 1948, Shannon applied a probability theory and mathematical statistics method to extend the concept of entropy [54], and defined information as "the reduction of entropy", that is "the reduction of system or transaction uncertainty" to reflect the lack of information in the internal configuration of the system or the measurement of the uncertainty of a random event.

$$E_{Shannon}^{(i)} = -\sum_{j=1}^{N^{(I)}} p(t_j^{(i)}) log_2 p(t_j^{(i)}). \tag{2}$$

$p(t_j^{(i)})$ represents the probability of a certain order quantity scale $t_j^{(i)}$ appearing in the customer's time series $T^{(i)}$, which reflects the uncertainty of the state.

Through the calculation formula of random entropy and Shannon entropy, it can be found that they do not take the time sequence into account, that is to say, the random entropy and Shannon entropy obtained by $T^{(i)} = 2321$ and $T^{(i)} = 1322$ is the same.

3) *Real entropy*

$$E_{real}^{(i)} = -\sum_{S_n^{(i)} \subset T^{(i)}} P(T_n^{(i)}) log_2 [P(T_n^{(i)})]. \tag{3}$$

$P(T_n^{(i)})$ represents the probability that a sub-sequence of the particular time series $T_n^{(i)}$ in the customer time series. However, if we directly follow the equation (3) in the calculation of the actual true entropy, the time complexity of finding all subsets under a given set is $O(2^n)$. So the Lempel-Ziv estimator will be used:

$$E_{real}^{\prime(i)} \approx (\frac{1}{n}\sum_j s_j^{\prime(i)})^{-1}\ln n. \tag{4}$$

$s_j^{\prime(i)}$ represents the length of the shortest sub-sequence starting in the month j and not occurring in the time period $1-j-1$.

Unlike random entropy and Shannon entropy, real entropy not only takes the frequency of different order sizes into account, but also their time order.

4) *LZW algorithm*

The LZW algorithm proposed by Lempel-Ziv-Welch which is a compression algorithm whose core idea is to replace repeated strings with tokens and map the repeated ones to tokens so that a longer string can be represented by a shorter code and the shortest sub-sequence is got. The specific algorithm process is shown in Algorithm 1:

The dictionary *code_table* obtained by LZW algorithm is the set of the shortest sub-sequence that never appears in the previous $j-1$ month while calculating the real entropy. For example, the shortest sub-sequence set of time series $T^{(i)} = 112112132$ is: 1,2,3,11,12,21, 121,13,32. Furthermore, the real entropy can be obtained by the equation (4), reflecting the uncertainty of the customer's ordering behavior. This method has been studied [55] that the estimated entropy $E_{real}^{\prime(i)}$ is very close to the real entropy result $E_{real}^{(i)}$.

5) *Predictability*

Entropy $E^{(i)}$ can be used to quantify the uncertainty of a customer's ordering behavior, and it can be transformed into the ordering predictability $P^{(i)}$ by the following [56]:

$$\begin{aligned} E^{(i)} &= -P^{(i)}\log_2 P^{(i)} - (1-P^{(i)})\log_2(1-P^{(i)}) \\ &\quad + (1-P^{(i)})\log_2(N^{(i)}-1), \end{aligned} \tag{5}$$

$E^{(i)}$ represents the entropy of a given customer $C^{(i)}$ and $N^{(i)}$ is the number of different categories of the order size in $T^{(i)}$.

Algorithm 1 The LZW algorithm

Input: customer time series $T^{(i)} = (t_1^{(i)}, t_2^{(i)}, \ldots, t_n^{(i)})$
Output: dictionary *code_table*
Initialization: the dictionary *code_table* only contains all the distinct states $t_j^{(i)}$;

1. Get the next state c reading the new state g
2. Merge g and c into a sub-sequence and look it up in the dictionary;
3. If exist, then $g = g + c$;
4. If not exist, then add $g + c$ to the dictionary and update $g = c$;
5. Repeat steps B-E until the entire sequence is read.

3.2 Anomaly detection

For anomaly detection, the iForest(Isolation Forest) algorithm is adopted, an unsupervised learning algorithm and utilizing the separation property of data to detect outlier number.

3.2.1 Isolation forest algorithm

A certain number of samples whose size is usually much smaller than the original data size are randomly selected from the entire data set as the preparation for building trees. An isolated tree is constructed from each subsampling set. Then the samples are divided into left or right sub-trees based on the segmentation value selected between the maximum and minimum values chosen randomly. When the tree reaches a specified height or the sample number in the nodes reaches a certain number or all selected sample feature values are the same, the above process is end. A specific number of isolated trees will be constructed and gathered into the isolated tree. The anomaly score is:

$$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}}. \tag{6}$$

$E(h(x))$ is the average path length of all trees in the forest and $c(n)$ is the average path length of the tree as:

$$c(n) = 2H(n-1) - \frac{2(n-1)}{n}. \tag{7}$$

Here, $H(i) \approx \ln(i) + 0.5772156649$, n is number of samples.

Through Isolation Forest algorithm, the outlier number of Customer $C^{(i)}$ is got and normalized into outlier rate as $A^{(i)}$ whose value ranges from 0% to 100%.

3.3 Customer regularity assessment

The entropy weight method is used to determine the weight of features assessing the regularity of customer behavior, which is a multi-criteria decision-making method based on information entropy theory. According to the characteristics of entropy, the randomness and disorder degree of a system can be judged by entropy. The greater the dispersion, the higher the indicator weight, which means that the indicator has a greater impact on the comprehensive evaluation. The calculation steps are as follows.

3.3.1 Indicators normalization

For the given customers $C^{(1)}, C^{(2)}, \ldots, C^{(m)}$, m samples and 2 indicators: customer predictability quantification and outlier rate:$(P^{(1)}, P^{(2)}, \ldots, P^{(m)})$, $(A^{(1)}, A^{(2)}, \ldots, A^{(m)})$ are got through the above method. Firstly, because the measurement units of each indicator are not uniform, it is necessary to carry out standardization processing before calculating the comprehensive index to solve the homogenization problem of different quality index values.

Indicators are divided into positive and negative by definition. The higher the positive indicators and the lower the negative indicators, the better. Customer predictability quantification $P^{(i)}$ is a positive indicator and outlier rate $A^{(i)}$ is a negative indicator, so they should be normalized as:

$$P'^{(i)} = \frac{P^{(i)} - \min(P^{(1)}, P^{(2)}, \ldots, P^{(m)})}{\max(P^{(1)}, P^{(2)}, \ldots, P^{(m)}) - \min(P^{(1)}, P^{(2)}, \ldots, P^{(m)})}, \tag{8}$$

$$A'^{(i)} = \frac{max(A^{(1)}, A^{(2)}, \ldots, A^{(m)}) - A^{(i)}}{\max(A^{(1)}, A^{(2)}, \ldots, A^{(m)}) - \min(A^{(1)}, A^{(2)}, \ldots, A^{(m)})}. \tag{9}$$

3.3.2 Indicators weight

Secondly, the entropy value of each indicator is calculated as:

$$e_P = -k \sum_{i=1}^{m} p_P^{(i)} \ln(p_P^{(i)}),$$

$$e_A = -k \sum_{i=1}^{m} p_A^{(i)} \ln(p_A^{(i)}). \tag{10}$$

Here, $p_P^{(i)} = P'^{(i)} / \sum_{i=1}^{m} P'^{(i)}$, $k = 1/\ln(m) > 0$.

Thirdly, the information entropy redundancy is calculated as:

$$d_P = 1 - e_P, \qquad d_A = 1 - e_A. \tag{11}$$

Then the weight is calculated as:

$$w_P = \frac{d_P}{d_P + d_A}, \qquad w_A = \frac{d_A}{d_P + d_A}. \tag{12}$$

Finally, the comprehensive assessment score can be got as:

$$S_i = \sum_{i=1}^{m} w_P \cdot P^{(i)} + w_A \cdot A^{(i)}. \tag{13}$$

3.4 Relating prediction model

Firstly, the following representative time series and customer ordering demand prediction models including Random Forest, XGBoost, CNN model, LSTM model, and the proposed CNN-BiLSTM network are selected for the experiment to verify the effectiveness of the proposed regularity quantitative method. Furthermore, the proposed Random Forest and CNN-BiLSTM network are used for the final prediction in the system.

3.4.1 Machine learning

1) *Random forest regression*

Random Forest Regression Model (RFR) is a commonly used ensemble regression algorithm, which is a collection of multiple decision trees, usually using bagging strategy, and different decision trees are trained independently. The final prediction result of random forest is:

$$f(x) = \frac{1}{g} \sum_{i=1}^{g} f_i(x). \tag{14}$$

g is the number of decision trees and $f_i(x)$ is the prediction result of model i. The advantages of random forest regression model include good interpretation, small calculation amount, high training efficiency and balanced error for unbalanced data sets. However, it is also prone to overfitting caused by abnormal data or noise in the data set.

1) *XGBOOST*

eXreme Gradient Boosting (XGBoost) is a variant of GBDT algorithm, which is a supervised learning algorithm. Unlike random forest regression model, XGBoost uses boosting strategy, a common and important ensemble learning technique. By adjusting the weight of the samples, the key classification features are selected many times, and the weak classifiers are trained step by step, while their weights are adjusted. These weak classifiers are combined to form a stronger classifier.

3.4.2 Neural network

The development of customer demand prediction models based on deep learning has been diverse. Here CNN, LSTM, BiLSTM model are selected as representative models of predicting time series for experiments.

1) *CNN network*

The Convolutional Neural Networks, a kind of Feedforward Neural Networks with deep structure including convolution calculation, one of the representative algorithms of deep learning, which has achieved remarkable results in image classification, natural language processing, and other fields [57, 58]. CNN is mainly used to deal with grid data, such as images, time series, and so on. The flow chart of basic one-dimensional CNN is shown in Fig. 2, which contains 3 special neural network layers:

Convolutional layer: It is used to extract features of the input data, which contains multiple convolution kernels whose element corresponds to a weight coefficient and a bias vector, similar to the neuron of a feedforward neural

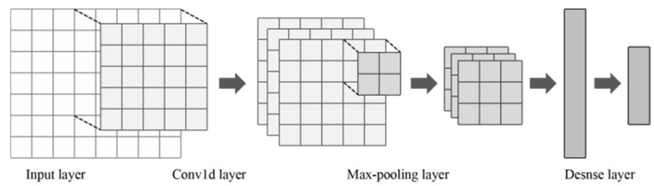

Figure 2 The flow of 1D convolution procession

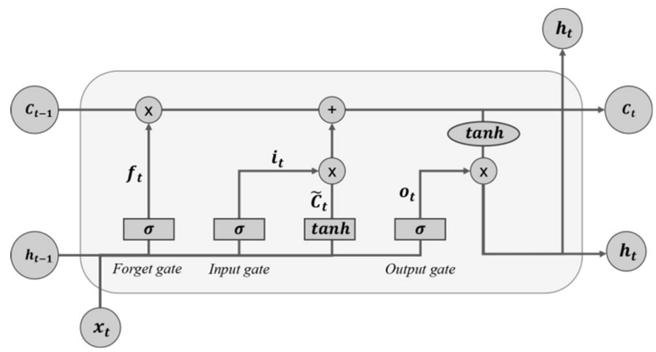

Figure 3 The schematic diagram of a LSTM cell

network.

$$x_{l+1} = f(\sum(x_l \times w_l + b_l)). \tag{15}$$

x_l represents the input data; x_{l+1} is the output data after convolution; f is the nonlinear activation function; w_l is the weight value in this layer; b_l is the bias value.

Pooling layer: It is used to carry out secondary sampling of the local features obtained by convolutional layer, which is controlled by the pooling size, step size and padding. Moreover, parameters are not needed to be retained, thus reducing the computation.

$$x_{l+1} = pooling(x_l). \tag{16}$$

pooling represents pooling operation; x_{l+1} is the output after secondary sampling.

Fully-connected layer (Dense layer): It is usually the last part of hidden layer of a convolutional neural network and only passes signals to other fully-connected layer. The feature maps lose their spatial topology and are expanded into vectors through activation function.

2) *LSTM network*

The Long Short-Term Memory model is designed to solve the long-term dependence problem of RNN because compared to the hidden layer with only one state, the hidden layer of LSTM has an additional long-term memory cell to store long-term state [59]. Moreover, the LSTM network has three gates: the input gate, the forget gate, and the output gate to control the long-term state. So It is suitable for LSTM model to process and predict important events with long intervals and delays in time series. As a nonlinear deep learning model, it can also be used as a complex nonlinear unit to construct a more complex large deep learning neural network. The core of LSTM is cell state and the schematic diagram of an LSTM cell is shown in Fig. 3.

Forget gate: The first step of LSTM is to decide what information the cell needs to discard, which is handled by the activation function called forget gate.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f). \tag{17}$$

It uses the information of h_{t-1} and x_t to output a vector between 0 and 1, whose value indicates what information is keep or discarded.

Input gate: To update the cell state, the input gate is needed. The previous hidden state and the current input is passed to the activation function, which determines what should be updated by converting the values to the range from 0 to 1. Also they should be input to the tanh function, compressing them between –1 to 1 to help regulate the network.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \tag{18}$$

$$\tilde{C}_t = tanh(W_C \cdot [h_{t-1}, x_t] + b_C). \tag{19}$$

Output gate: The output gate determines what the next hidden state is, which is also used for prediction. The pre-

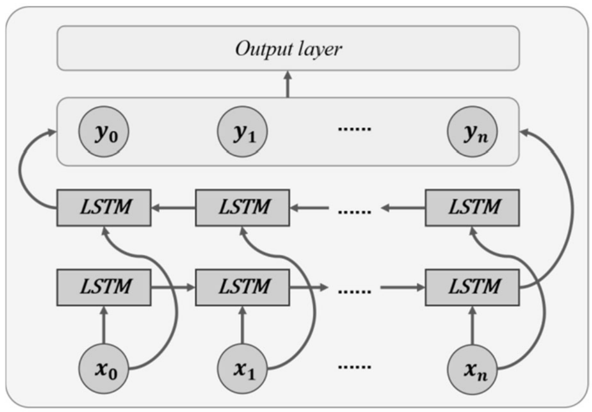

Figure 4 The structure diagram of BiLSTM

vious hidden state and the current input are input to the activation function as above. Then the new cell state is passed to the tanh function. The output is multiplied with them, deciding the information carried, which will be passed to the next time step.

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t, \tag{20}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \tag{21}$$

$$h_t = o_t \times tanh(C_t). \tag{22}$$

3) *BiLSTM network*

The Bidirectional Long Short-Term Memory Network is a variant of RNN network. Through processing the input sequence in two directions, forward and backward, BiLSTM can capture the context information before and after the current position at the same time. It also contains memory units, input gate, forget gate, and output gate. BiLSTM is trained in an end-to-end manner, which means it can learn the mapping between input and output directly from raw data without handcrafting features or rules. Overall, BiLSTM combines the utilization of bidirectional processing, long sequence dependency modeling, and contextual information, allowing it to predict effectively. The structure diagram of BiLSTM is shown in Fig. 4. The single-layer BiLSTM is composed of two LSTM networks, one is to process the input sequence forward and the other is in the forward direction. Then they are concatenated and the final output is obtained after all the time steps have been calculated.

4) *CNN-BiLSTM network*

The system proposed a merged CNN-BiLSTM neural network model to improve the prediction accuracy, whose network structure is shown in Fig. 5. The network aims to use CNN layers for feature extraction on the input data, LSTM layers and BiLSTM layers are combined to support data sequence prediction. The reason for this combination is that CNN is good at extracting local features

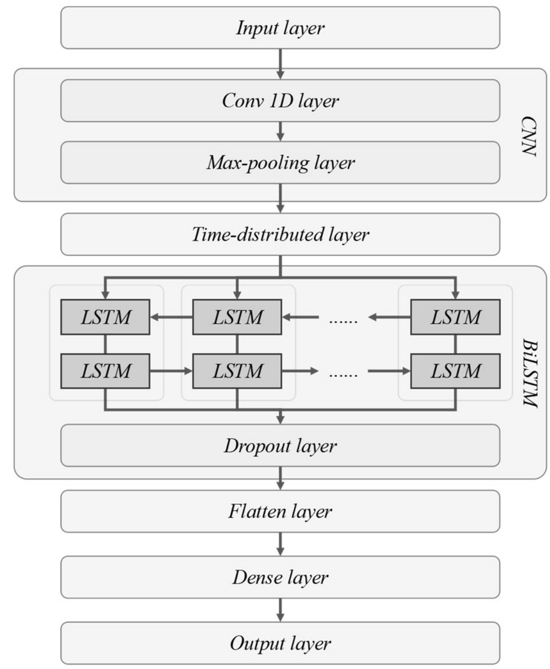

Figure 5 The network structure diagram of CNN-BiLSTM

of 2D image data and BiLSTM is proficient in capturing long-term dependencies in sequences. Through combining the two network, the merged neural network can be considered both temporal and spatial information of the input data, and thus parameter over-fitting risks can be reduced, which provides more accurate predictions, better model performance, and higher training efficiency. A convolutional network needs to be defined as the comprised of Conv1D and MaxPooling1D layers arranged to form stack of the required depth in an orderly manner. However, only CNN network is not capable of handling the time sequence data, so it is added to the BiLSTM network with another time-distributed layer.

4 Experimental step

The experimental data are collected from historical transaction. An efficient and accurate ordering assessment and prediction system can provide insights for enterprise decision-making. Firstly, it is worthwhile to introduce the background of data collection offered by a manufacturing enterprise. Although the time period of the data set covers 10 years, other related useful features are difficult to obtain. So, the data pro-processing is the first main step, which includes two parallel stages in order to extract more diverse data features of customer ordering behavior and provide input data basis for network prediction. It is introduced in Sect. 4.2 in detail. Secondly, five prediction models of machine learning and deep learning are experimented to verify the effectiveness of the data quantita-

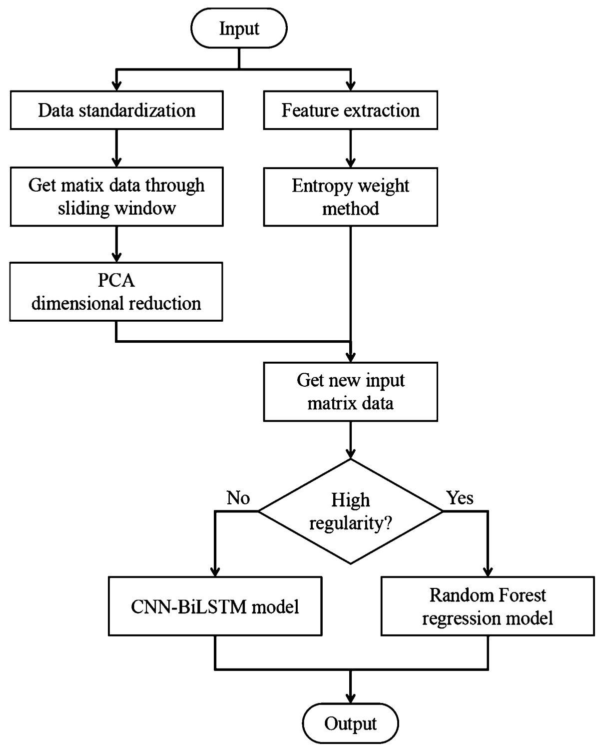

Figure 6 The flow chart of the proposed system

tive method proposed. Finally, Random Forest and CNN-BiLSTM network are established separately to predict ordering demand of the customers with high and low regularity. The flow chart of the proposed system is shown in Fig. 6.

4.1 Data set

The original data set used comes from a total of 36,764 historical transaction data of a manufacturing enterprise, covering the period from July 18, 2014 to January 30, 2024. The attributes included are customer number, order number, customer name, customer type, customer nature, sales quantity, unit price (yuan), total price (yuan), product model, and delivery date. Totally, there are 4,482 transaction customers. The customer types are divided into four categories: large customer, important customer, small customer and sluggish customer, of which 9 are large customers, 104 are important customers, 3,912 are small customers and 457 are sluggish customers by the sales staff according to the empirical evaluation of the customer's trading volume, enterprise scale and other factors.

Secondly, the original data should be calculated and processed to obtain $t_j^{(i)}$ of customer $C^{(i)}$ according to the definition of time series $T^{(i)}$. In order to make the experimental analysis more representative and practical, the remaining 4,473 customers are selected and those with scattered transaction data are cleared. The screening criteria are as follows: the total number of historical transactions and the number of months in which transactions took place. Table 1 intuitively shows that the degree of dispersion of orders among customers is very high, and at least 75% of customers have small orders in total, so their analytical value to the enterprise is not high. In the end, 572 customers whose total number of historical transactions more than 15 and the number of transactions more than 4 months are obtained. Among them, there are 8 big customers, 90 important customers, 441 small customers and 33 sluggish customers.

Table 1 Screening criteria data distribution statistics

Screening criteria	Historical transaction total number	Transaction month number
Number	3009	3009
Average	132.82	4.92
Standard deviation	1892.97	11.48
Minimum	0	0
25% quantile	2	1
50% quantile	4	1
75% quantile	15	4
Maximum	74747	114

4.2 Data processing

For customer ordering predictability quantification, firstly the clustering algorithm: K-Means clustering is used to discretize the monthly product order quantity. Here, 5 clusters are selected and the actual order quantity which is 0 is marked as the sixth class. The similar data points are divided into the same class through clustering, which provides the basis for the subsequent quantitative calculation method of ordering predictability as well as alleviates the trend influence of the market on the change of customer order quantity to a certain extent, making the quantitative results more objective. Figure 7 shows the time series line chart of a large customer before and after discretization. It can be seen that more obvious periodic characteristics are showed in the processed data compared with the original data. At the same time, with the change of time and market, the customer's order quantity shows obvious trend change and the discretization process also partly detrend the time series. Secondly, the result of ordering predictability quantification is calculated by the LZW algorithm which represents the real entropy $E'^{(i)}_{real}$, and the predictability $P^{(i)}$ is got. Thirdly, Isolation Forest is used to calculate the outlier numbers $A^{(i)}$ of customer for anomaly detection. Then the result of customer assessment, regularity score $S^{(i)}$ is calculated with $P^{(i)}$ and $A^{(i)}$ by entropy weight method.

Besides data feature extracting, original data need to be standardized and pre-processed before training the network. Robust standardization which has good robustness

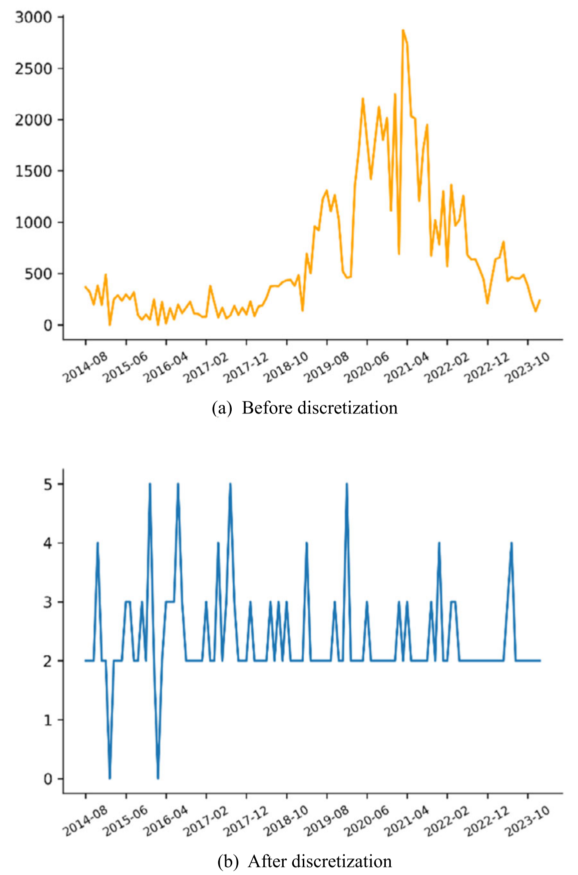

Figure 7 Time series diagram of a large customer before and after discretization

to outliers is used as:

$$x_{norm} = \frac{x - median}{IQR}, \tag{23}$$

median represents the median of data, *IQR* is the interquartile of data, the upper quartile minus the lower quartile.

4.3 Model setup

In the model setup stage, there are five machine learning and deep learning algorithms including Random Forest regression, XGBoost, CNN network, LSTM network, and CNN-BiLSTM network. All these models will be trained with preprocessed data by using Robust standardization, sliding window, and principal components analysis. The total data set is divided into training data set and testing data set in an 8:2 ratio. For the CNN network, the input data should be reshaped, which is $\mathrm{N} \times \mathrm{M} \times 1$ as the first layer. Next, Conv1D layer, Max-pooling layer, Flatten layer, Dense layer and activation function are set in the hidden layers. The size of convolutional kernel is set 5×1. Max-pooling is set 2×1 to help improve model efficiency and prevent over-fitting. For establishing LSTM network, three layers: input layer, hidden layer and output layer need to be set. More, the Dropout layer is also set to prevent over-fitting. The learning rate is set as 0.001 to make the network train positively. Furthermore, the type of layers and the number of nodes and layers should be considered so as to balance the computational power of calculation and the prediction accuracy [60]. The internal network structure of the proposed CNN-BiLSTM model is shown in Fig. 8. For the above three deep learning networks, Rectified Linear Unit (ReLU) and Tahn function are chosen as the activation function:

$$\mathrm{output} = \mathrm{ReLU}(\mathrm{x}) = \max(0, \mathrm{x}), \tag{24}$$

$$\mathrm{output} = \tanh(\mathrm{x}) = \frac{\mathrm{e}^{\mathrm{x}} - \mathrm{e}^{-\mathrm{x}}}{\mathrm{e}^{\mathrm{x}} + \mathrm{e}^{-\mathrm{x}}}. \tag{25}$$

RMSprop algorithm is selected as the optimizer, which draws on Adagrad's ideas that modifies its gradient accumulation to an exponentially weighted moving average so that RMSprop can work better in the non-convex setting. Due to the weighted average, it improves the problem of the excessive swing amplitude in the update of loss function and the learning rate can have adaptive adjustment:

$$E[g^2]_t = \alpha E[g^2]_{t-1} + (1-\alpha) g_t^2, \tag{26}$$

$$W_{t+1} = W_t - \frac{\eta_0}{\sqrt{\mathrm{E}[\mathrm{g}^2]_\mathrm{t} + \epsilon}} \odot g_t, \tag{27}$$

$$g_t = \Delta J(W_t). \tag{28}$$

W_t represents the parameters of the model at time t, g_t is the gradient of the loss with respect to W, $E[g^2]_t$ is the mean of the first time t squared gradients, where α is the impetus. η_0 is the global initial learning rate, and ϵ is a small number, usually $1\mathrm{e}^{-8}$, to avoid having a 0 denominator.

4.4 Performance evaluation

Symmetric Mean Absolute Percentage Error (sMAPE) E_1 and Mean Squared Logarithmic Error (MSLE) E_2 that is also selected as the loss function in the deep learning networks are used to evaluate the performance of the prediction model, which are calculated as:

$$E_1 = \frac{1}{n} \sum_{i=1}^{n} \frac{|\, y_i - y_i^{pred}\,|}{(y_i + y_i^{pred})/2 + 1}, \tag{29}$$

$$E_2 = \frac{1}{n} \sum_{i=1}^{n} (\log(1 + y_i) - \log(1 + y_i^{pred}))^2. \tag{30}$$

y_i is true value, y_i^{pred} is predicted value, $\overline{y}$ is the sample average, n is the total sample number.

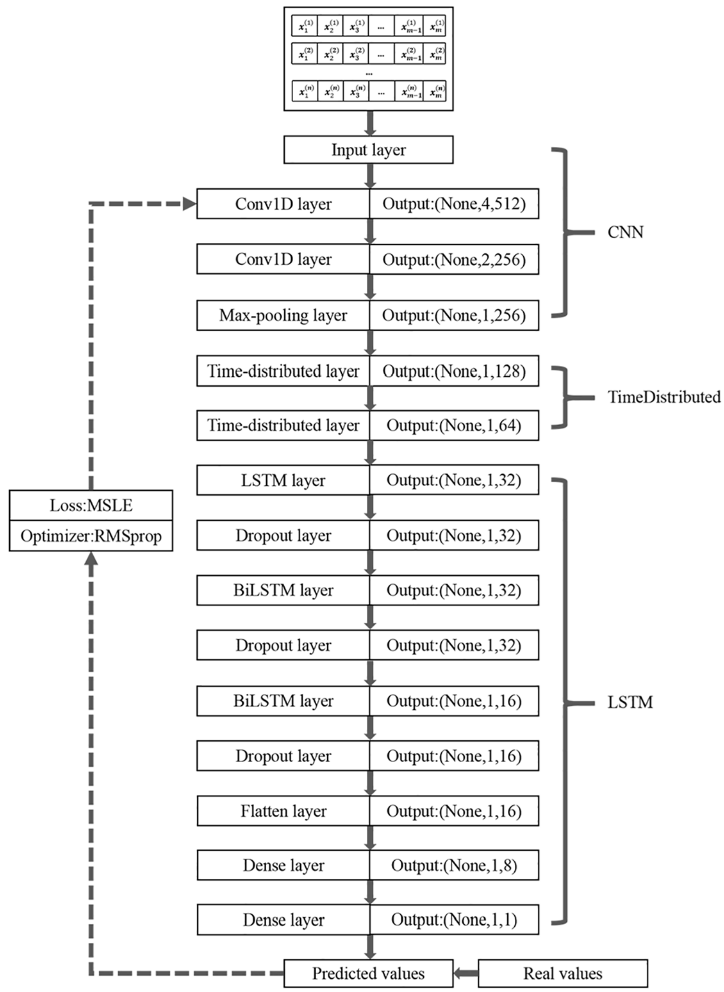

Figure 8 Internal network structure of CNN-BiLSTM

5 Experiment results and discussion

5.1 Feature extracting

It is shown in Fig. 9 that the range of customer ordering predictability $P^{(i)}$ varies from 0.49 to 0.9, which is a wide range. Therefore, this quantification method has a certain role in distinguishing the ordering predictability of customers, and the number of customers with the quantitative value of regularity in the range of 0.7 to 0.8 is the most.

The statistical results of dispersion customer predictability after classification according to the four customer types are shown in Fig. 10. It can be intuitively found that the predictability of large customers is relatively concentrated, whose lower boundary line (the minimum value except outliers) is the highest. Important customers' average predictability and upper boundary line (the maximum value except outliers) are both the highest and no outlier exists. The number of small customers is large and their ordering behaviors are often different, and their predictability variation is the widest, whose anomaly is the most. How-

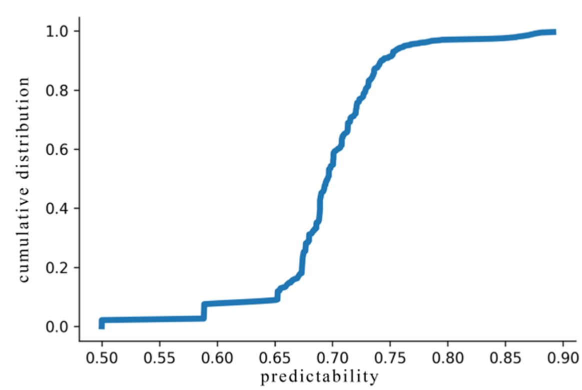

Figure 9 The cumulative distribution of the customer number over ordering predictability

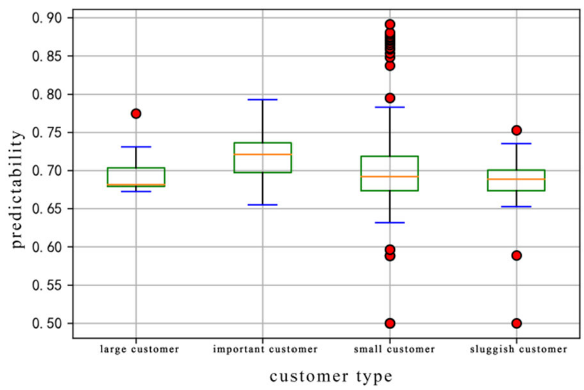

Figure 10 Box plots of predictability over customer types

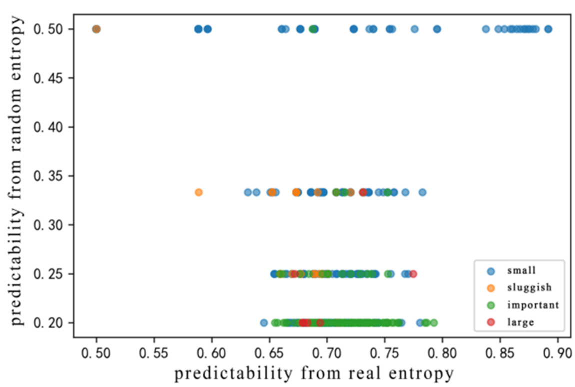

Figure 11 Predictability distributions of real and random entropy

ever, the rankings of various data indicators of sluggish customers are relatively stable.

Figure 11 and Fig. 12 show the data distribution relationship among the predictability results of random entropy, Shannon entropy and real entropy, which can be found that the random entropy quantification results have little differentiation to the customer ordering predictability. However, the reason for the high predictability from random entropy quantification results of sluggish and small customers stems from the fact that their ordering scale is generally concentrated in 0, 1, 2, leading to the small number of categories. On the contrary, the category number of important and large customer is more, so the random entropy obtained will be higher and the predictability will be lower. Except for a small number of outliers, the predictability from Shannon entropy and real entropy quantization shows a relatively obvious positive correlation trend. Since the real entropy calculation takes not only the scale categories number, but also the time order of their appearance into account, the results of real entropy quantization are selected for subsequent experiments.

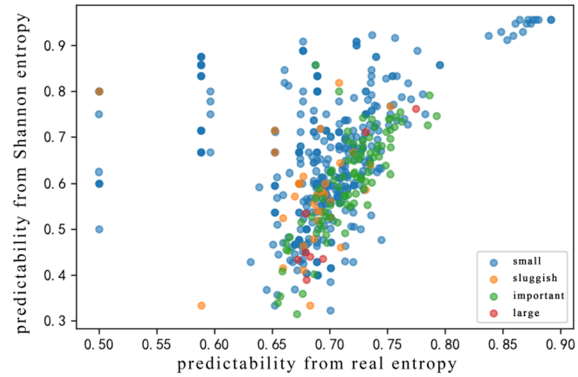

Figure 12 Predictability distributions of real and Shannon entropy

Furthermore, the anomaly detection of customers is shown in Fig. 13. Obviously, whether it's the mean or the minimum or the median, the results of large customer are all the lowest, followed by important customers. Small customers and sluggish customers perform similarly. Such a distinction is equally consistent with experimental expectations and regular experience.

Finally, the result of comprehensive assessment scores $S^{(i)}$ is shown in Fig. 14 to Fig. 16. The trend of statistical value change based on customer types is similar to the result of anomaly detection. Figure 15 shows that the range of scores varies from 0 to 85 and the number of customers who score between 20 and 40 is the most, which can also be found in Fig. 16.

5.2 Model prediction

The relationship between the customer comprehensive assessment score obtained by the above method and the performance of five predicting models is analyzed further so as to verify the effectiveness of the quantitative method proposed, providing relevant basis for the design and selection of customer demand predicting models.

Figure 17 illustrates the change of sMAPE of five prediction models with the enhancement of regularity score. It can be seen that LSTM model has the best prediction effect. When the customer order regularity is higher, the performance of RFR and LSTM model is significantly better. However, it can be seen that three deep learning based networks have better prediction for those customers having low regularity than the two machine learning models, indicating that deep learning networks are able to extract deeper features and overcome the difficulty of prediction caused by low regularity. Likewise, the change of MSLE with the increase of regularity is shown in Fig. 18 and the trend is almost the same. It can be seen that when selecting the method of customer ordering demand prediction, those models based on machine learning with simpler, smaller cost and stronger interpretability can be se-

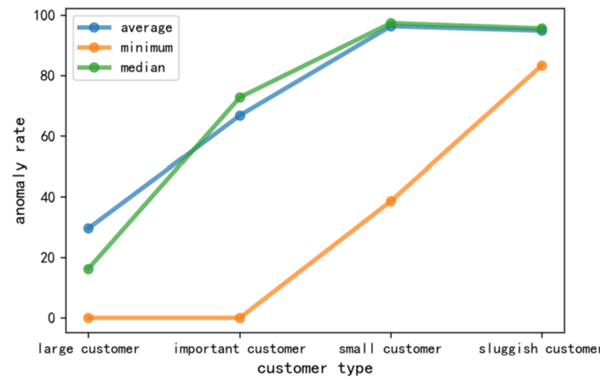

Figure 13 Anomaly rate over customer types

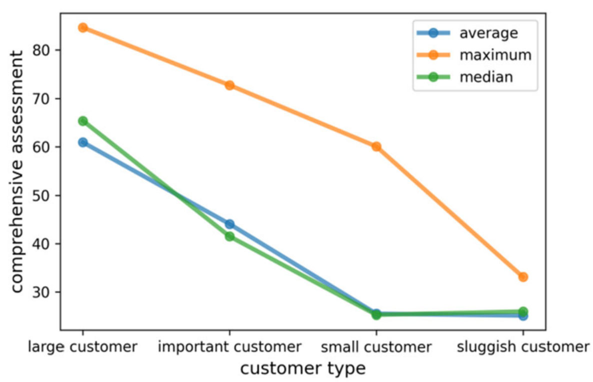

Figure 14 Comprehensive assessment over customer types

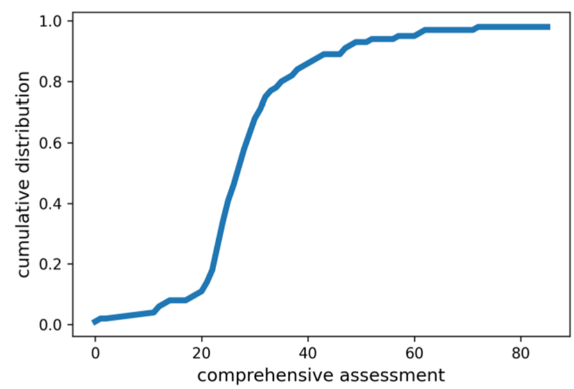

Figure 16 The cumulative distribution of the customer number over comprehensive assessment

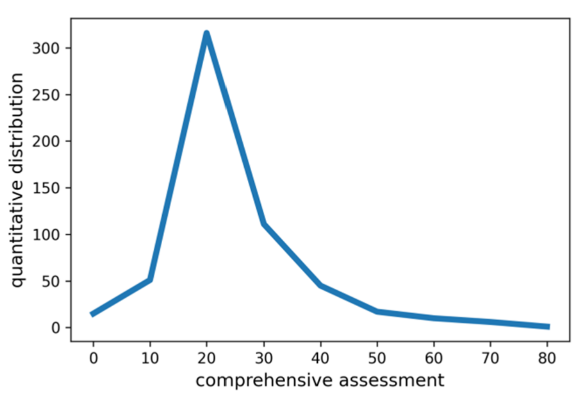

Figure 15 The quantitative distribution of the customer number over comprehensive assessment

lected for the customer who has higher ordering regularity. However, it is necessary to dig out more deep nonlinear features and select more complex models to improve the prediction accuracy for those customers whose ordering regularity are lower. At the same time, the reliability of the quantitative regularity method proposed can also be confirmed. From Fig. 17 and Fig. 18, the performance of CNN-BiLSTM network proposed is not obviously excellent, sometimes worse than RFR and LSTM model. The reason for this situation will be discussed in detail later.

In the scenario using CNN-BiLSTM network to predict the all customers, it can be seen in Table 2 that the performance of the CNN-BiLSTM network is totally better than the other models. Comparing the performance of train and test data set for RFR and LSTM, it is apparent that they have a certain amount of over-fitting, for the performance of training is much better than the testing. Because of the ratio of training and testing, so the average loss of RFR and LSTM is lower than CNN-BiLSTM network, but it does not mean their overall prediction is better. Above all, the performance of deep learning networks is more stable and better than machine leaning models. Finally, the sMAPE of the system proposed is 0.446 and the MSLE is 0.471, which is significantly the best.

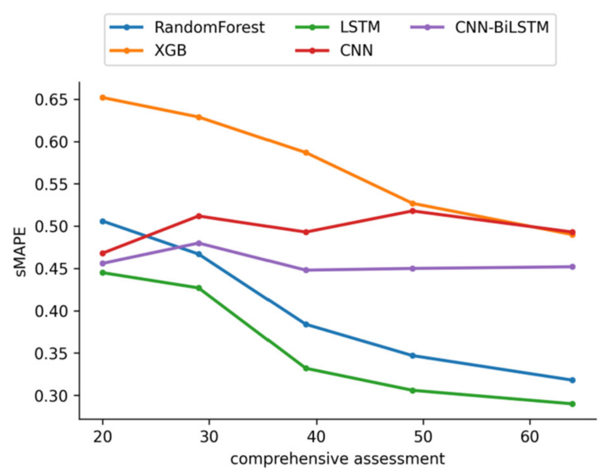

Figure 17 Predicting Model's sMAPE and comprehensive assessment

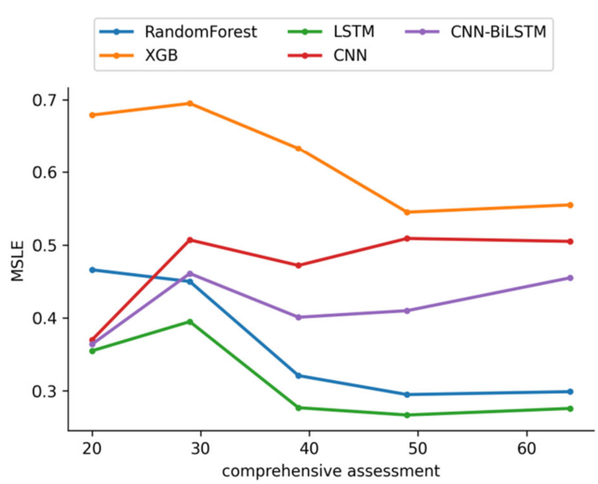

Figure 18 Predicting Model's MSLE and comprehensive assessment

Table 2 The prediction results of different models

Models	Train sMAPE	**Test sMAPE**	Train MSLE	**Test MSLE**
Random Forest	0.359	**0.602**	0.285	**0.711**
XGBoost	0.574	**0.619**	0.614	**0.703**
CNN	0.486	**0.540**	0.444	**0.596**
LSTM	0.314	**0.567**	0.227	**0.678**
CNN-BiLSTM	0.438	**0.537**	0.371	**0.523**
System	**0.446**		**0.471**	

Through the experiment result and cost of the five models, as the key research in the current data mining and prediction field, deep learning has more accurate prediction than most traditional machine learning methods, but their network structural complexity is greatly higher than machine learning models, which rely on higher computer configuration and investment costs. Furthermore, data pre-processing of data which especially comes from real enterprises is of vital importance that can decide the upper limit of model prediction accuracy to a certain extent.

6 Conclusion

The maintenance of customers is essential to enterprises, so a valuable assessment and efficient prediction of customer behavior can provide tangible benefits to the enterprise. In this paper, feature extraction and data mining including entropy-based customer predictability quantification, anomaly detection using Isolation Forest algorithm, and comprehensive assessment method is the first main research point. The second is the machine learning and deep learning modelling for customer demand prediction, including including the proposed Random Forest, XGBoost, CNN model, LSTM model, and CNN-BiLSTM network. The main contribution of this paper is to propose a multi-step regularity assessment and joint prediction system for ordering time series based on entropy and deep learning. In the feature extraction, an entropy-based method to quantify customer ordering predictability is proposed with the analysis of result rationality according to the enterprise experience. Isolation Forest is used for anomaly detection and entropy weight method is adopted to give the final regularity assessment score for customers. In the data pre-processing stage, Robust standardization, sliding window, and principal components analysis are implemented to extract the key features. Subsequently, five prediction models using different parameters are experimented and the performance evaluation of sMAPE and MSLE are selected. The relationship between the prediction performance of the order quantity model and regularity score is compared. Furthermore, a merged CNN-BiLSTM network is established to predict the customers with low regularity more accurately and Random Forest Regression is adopted for customers with high regularity to save cost and improve efficiency. The results show that the distribution and statistics of the quantitative predictability results are consistent with the customer classification made by the enterprise based on historical experience. The research that has practical value for it not only gives an objective result to answer how high the regularity of customers is, filling the gap in the field of customer regularity quantification, but also provides a theoretical basis for the choosing of customer demand prediction models. In terms of model prediction performance's change, the trend of two indicators shows that machine learning models generally perform better in the prediction of customers with higher regularity and perform poorly in predicting those with lower regularity, and the performance of three deep learning networks is quite stable. Finally, the performance of the CNN-BiLSTM network is totally better than the other models and the intelligent system has significantly accurate prediction of customer demand as a whole. According to the output of the customer assessment and prediction system, managers of enterprises can have a more profound and comprehensive understanding of the regular distribution of customer ordering behavior, so sales can work more specifically dealing with customers having different ordering regularity, maintaining consistent high-quality service to those with high regularity and being more aggressive in getting customers with low regularity. Furthermore, the prediction of customer demand and performance of different models is more interpretable with the result of regularity quantification. Due to low requirements for data sources of the proposed system which provides intelligent customer prediction and assessment, it is highly scalable and can be widely used in other enterprises or fields for follow-up research.

In this experiment, the effectiveness of customer regularity quantization method and accuracy of customer prediction with Random Forest regression and the proposed CNN-BiLSTM from the joint system are mainly verified. Due to the features such as small size and high sparsity, the current sample data set has certain limitations, and it may be necessary to increase the data set size or use other sample enterprise data to continue the future study. In the later stage, it may improve the structure of deep learning networks if the configuration of computer allows to improve the accuracy of models with acceptable time and cost.

Acknowledgements
The authors are grateful for the efforts and suggest of our colleagues in the Science and Technology Development Co., Ltd. of Shanghai Research Institute of Building Sciences and Professor Wang at Shanghai Engineering Research Center for Safety Intelligent Control of Building Machinery.

Author contributions
Yichen Zhou was responsible for the data processing, modeling and writing. Wenhe Han led the primary target and Heng Zhou contributed to the implementation of the research. All authors read and approved the final manuscript.

Funding
No funding.

Data availability
Enterprise data is not available.

Declarations

Competing interests
The authors declare no competing interests.

Received: 31 July 2024 Revised: 24 August 2024
Accepted: 11 September 2024 Published online: 25 October 2024

References

1. L. Li, J.C. Zhen, Customer churn prediction based on the integration of meta-learning network of the forest. Telecommun. Sci., 1–14[2024-08-22]
2. D. Wang, D.C. Xie, C. Wang, et al., Regularity and predictability of individual spatiotemporal behavior: a case study of residents working day activities in Shanghai. Progr. Geogr. **40**(3), 433–440 (2021)
3. W.P. Lang, W.T. Mao, T.J. Luo, et al., Predictability evaluation and joint forecasting method for intermittent time series. J. Comput. Appl. **42**(9), 2722–2731 (2022)
4. W.T. Chen, X.L. Pan, X.Y. Zhang, Exploring Time Series Predictability using Permutation Entropy. J. Syst. Sci. Math. Sci., 1–21[2024-08-22]
5. B. Beckage, L.J. Gross, S. Kauffman, The limits to prediction in ecological systems. Ecosphere **2**(11), 1–12 (2011)
6. F. Pennekamp, A.C. Iles, J. Garland, et al., The intrinsic predictability of ecological time series and its potential to guide forecasting. Ecol. Monogr. **89**(2), e01359 (2019)
7. J. Garland, R. James, E. Bradley, Model-free quantification of time-series predictability. Phys. Rev. E **90**(5), 052910 (2014)
8. E. Xu, T. Zhou, Z.W. Yu, et al., Equivalence between time series predictability and Bayes error rate. Europhys. Lett. **6**(141), 61003 (2022)
9. Y.P. Lang, W.T. Mao, T.J. Luo, et al., Predictability evaluation and joint forecasting method for intermittent time series. J. Comput. Appl. **42**(9), 2722–2731 (2022)
10. G. Boffetta, M. Cencini, M. Falcioni, et al., Predictability: a way to characterize complexity. Phys. Rep. **356**(6), 367–474 (2002)
11. W. Bialek, I. Nemenman, N. Tishby, Predictability, complexity, and learning. Neural Comput. **13**(11), 2409–2463 (2001)
12. R.K. Guntu, P.K. Yeditha, M. Rathinasamy, et al., Wavelet entropy-based evaluation of intrinsic predictability of time series. Chaos, Interdiscip. J. Nonlinear Sci. **30**(3), 033117 (2020)
13. J.D. Croston, Forecasting and stock control for intermittent demands. J. Oper. Res. Soc. **23**(3), 289–303 (1972)
14. Y. Chen, H. Zhao, L. Yu, Demand forecasting in automotive aftermarket based on ARMA model, in *Proceedings of the 2010 International Conference on Management and Service Science* (IEEE, Piscataway, 2010), pp. 1–4
15. R.S. Mor, J. Nagar, A. Bhardwaj, A comparative study of forecasting methods for sporadic demand in an auto service station. Int. J. Bus. Forecast. Mark. Intell. **5**(1), 56–70 (2019)
16. F.R. Dywer, Customer lifetime valuation to support marketing decision making. J. Direct Mark. **3**(4), 63–67 (1989)
17. S. Monalisa, P. Nadya, R. Novita, Analysis for customer lifetime value categorization with RFM model. Proc. Comput. Sci. **161**, 834–840 (2019)
18. B. von Mutius, H. Arnd, Customized targeting strategies for category coupons to maximize CLV and minimize cost. J. Retail. **97**(4), 764–779 (2021)
19. M.L. Chen, H.Z. Li, Study on value segmentation and retention strategies of customer. Group Technol. Prod. Modernization **4**, 23–27 (2001)
20. D. Chen, Y.L. Sun, W. Xue, Robust CLV measurement in non-contractual settings: a study of CLV measurement combining probability models and machine learning algorithms. Manag. Rev. **31**(4), 83–98 (2019)
21. X. Nie, J.H. Tian, P.B. Hu, et al., Research on comprehensive evaluation system of customer value based on AHP. Chinese & Foreign Corporate Culture **5**, 76–78 (2022)
22. L. Yang, Y.G. Kou, Z. Bai, et al., An empirical research on the influence of improved RFM model on customer segmentation. Math. Pract. Theory **51**(1), 33–39 (2021)
23. P. Grassberger, Toward a quantitative theory of self-generated complexity. Int. J. Theor. Phys. **25**(9), 907–938 (1986)
24. R.K. Guntu, P.K. Yeditha, M. Rathinasamy, et al., Wavelet entropy-based evaluation of intrinsic predictability of time series. Chaos, Interdiscip. J. Nonlinear Sci. **30**(3), 033117 (2020)
25. G. Boffetta, M. Cencini, M. Falcioni, et al., Predictability: a way to characterize complexity. Phys. Rep. **356**(6), 367–474 (2002)
26. S.V. Scarpino, G. Petri, On the predictability of infectious disease outbreaks. Nat. Commun. **10**(1), 898 (2019)
27. S.M. Pincus, Approximate entropy as a measure of system complexity. Proc. Natl. Acad. Sci. **88**(6), 2297–2301 (1991)
28. D.B. Zhang, B.T. Zhang, L.W. Ling, et al., Carbon price forecasting based on secondary decomposition and aggregation strategy. J. Syst. Sci. Math. Sci. **42**(11), 3094–3106 (2022)
29. B. Fadlallah, B. Chen, A. Keil, et al., Weighted-permutation entropy: a complexity measure for time series incorporating amplitude information. Phys. Rev. E **87**(2), 022911 (2013)
30. M. Xu, L.F. Shen, Analysis of financial abnormal volatility and market efficiency based on symbolic time series method. J. Appl. Stat. Manag. **34**(2), 357–366 (2015)
31. J. Garland, R. James, E. Bradley, Model-free quantification of time-series predictability. Phys. Rev. E **90**(5), 052910 (2014)
32. M. Yang, J.C. Dong, Y. Qi, Predictability of wind power based on approximate entropy. Acta Energ. Solaris Sin. **37**(10), 2710–2718 (2016)

33. T. Xu, X.R. Xu, Y.J. Hu, An entropy-based approach for evaluating travel time predictability based on vehicle trajectory data. Entropy **19**(4), 165 (2017)
34. C.M. Chou, Complexity analysis of rainfall and runoff time series based on sample entropy in different temporal scales. Stoch. Environ. Res. Risk Assess. **28**(6), 1401–1408 (2014)
35. J.Z. Huo, J. Xu, M.Z. Chen, Muti-step sales forecasting of retail merchandise based on EEMD-HW-GBDT model. J. Ind. Eng. Manag. **29**(1), 133–141 (2024)
36. N. Liu, S. Ren, T.-M. Choi, et al., Sales forecasting for fashion retailing service industry: a review. Math. Probl. Eng. **2013**, 1–9 (2013)
37. G.X. Huang, X.H. Cao, X.F. Qian, Research on spare parts prediction model based on random forest. Oper. Res. Manag. Sci. **30**(10), 165–168 (2021)
38. Y. Zhenqi, Highway traffic accident prediction based on SVR trained by genetic algorithm. Adv. Mater. Res. **433**(440), 5886–5889 (2012)
39. L. Xintong, H. Xiaolong, X. Binglei, A model of short-term forecast of passenger flow of buses based on SVM-KNN under rainy conditions. J. Transp. Inf. Saf. **36**(5), 117–123 (2018)
40. A. De Caigny, K. Coussement, K.W. De Bock, A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. Eur. J. Oper. Res. **269**(2), 760–772 (2018)
41. A. Garg, S. Deshmukh, Maintenance management: literature review and directions. J. Qual. Maint. Eng. **12**(3), 205–238 (2006)
42. A. Malhi, R. Yan, R.X. Gao, Prognosis of defect propagation based on recurrent neural networks. IEEE Trans. Instrum. Meas. **60**(3), 703–711 (2011)
43. J.-K. Hong, LSTM-based sales forecasting model. KSII Trans. Int. Inf. Syst. **15**(4), 1232–1245 (2021)
44. T.Y. Zhang, S. Moro, R.F. Ramos, A data-driven approach to improve customer churn prediction based on telecom customer segmentation. Future Internet **14**(3), 94 (2022)
45. M.Z. Alom, T.M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M.S. Nasrin, M. Hasan, B.C. Van Essen, A.A. Awwal, V.K. Asari, A state-of-the-art survey on deep learning theory and architectures. Electronics **8**(3), 292 (2019)
46. C. Miaomiao, V.O.K. Li, V.W.S. Chan, A CNN-LSTM model for traffic speed prediction, in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)* (IEEE, Piscataway, 2020), pp. 1–5
47. T. Bogaerts, A.D. Masegosa, J.S. Angarita Zapata, et al., A graph CNN-LSTM neural network for short and long-term traffic forecasting based on trajectory data. Transp. Res., Part C, Emerg. Technol. **112**, 62–77 (2020)
48. C. Zhenxiang, L. Peijie, C. Shuying, et al., Photovoltaic power prediction based on K-means++ and hybrid deep learning. Electr. Eng. **22**(9), 7–13, 33 (2021)
49. C. Jing, Z. Zhaochong, W. Linkai, et al., Short-term bus passenger flow prediction based on convolutional long-short-term memory network. J. Syst. Simul. **36**(2), 476–486 (2024)
50. Q. Shi, Q.Y. Li, Research on the risk management and control of enterprise muti-value chain cooperative operation based on RF-BAS-CNN model. Commun. Finance Account. **22**, 117–121+170 (2023)
51. Z.X. Huang, R.H. Xia, X.T. Zhang, Prediction of telecom customer churn based on attention mechanism and graph convolution. Comput. Eng. Des. **44**(6), 1685–1691 (2023)
52. Y.X. Zhang, S.Z. Wang, B. Chen, et al., Traffic GAN: network-scale deep traffic prediction with generate adversarial nets. IEEE Trans. Intell. Transp. Syst. **22**(1), 219–230 (2021)
53. N. Feng, S.N. Guo, C. Song, et al., Multi-component spatial-temporal graph convolution networks for traffic flow forecasting. J. Softw. **30**(3), 759–769 (2019)
54. C.E. Shannon, A mathematical theory of communication. Bell Syst. Tech. J. **27**(3), 379–423 (1948)
55. I. Kontoyiannis, P.H. Algoet, Y.M. Suhov, et al., Non-parametric entropy estimation for stationary processes and random fields, with applications to English text. IEEE Trans. Inf. Theory **44**(3), 1319–1327 (1998)
56. K. Zhao, D. Khryashchev, J. Freire, et al., *Predicting Taxi Demand at High Spatial Resolution: Approaching the Limit of Predictability*. IEEE International Conference on Big Data (IEEE 2017)
57. J. Gu, Z. Wang, J. Kuen, et al., Recent Advances in Convolutional Neural Networks. Pattern Recognit. (2015)
58. C. Ken, C. Fei, L. Baisheng, et al., Dynamic spatio-temporal graph-based CNNs for traffic flow prediction. IEEE Access **8**, 185136–185145 (2020)
59. Y. Zhang, R. Xiong, H. He, M.G. Pecht, Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries. IEEE Trans. Veh. Technol. **67**(7), 5695–5705 (2018)
60. S. Wang, J. Wan, D. Zhang, D. Li, C. Zhang, Towards smart factory for industry 4.0: a self-organized multi-agent system with big data based feedback and coordination. Comput. Netw. **101**, 158–168 (2016)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.