



Research article

A strategy for predicting waste production and planning recycling paths in e-logistics based on improved EMD-LSTM

Shujuan Liu, Hui Jin* and Yanbiao Di

School of Logistics, Liaoning Vocational University of Technology, Jinzhou 121007, China

* **Correspondence:** Email: jinhui5868@edu.email.cn; Tel: +8613614165647.

Abstract: With the rapid development of e-commerce, express delivery has been chosen and accepted by consumers, and a large number of express packages have resulted in serious waste of resources and environmental pollution. Because of the irregularity of online goods purchases by users in real life, logistics parks are unable to accurately judge the recycling needs of various regions. In order to solve this problem, we propose an improved empirical mode decomposition (IEMD) algorithm combined with a long-short-term memory (LSTM) network to deal with the addresses and categories in logistics data, analyze the distribution of recyclable logistics waste in the logistics park service area and in the express recycling station within the logistics park, judge the value of recyclable logistics waste, optimize the best path for recycling vehicles and improve the success rate of logistics waste recycling. In order to better research and verify the IEMD-LSTM prediction model, we model and simulate the algorithm behavior of the express waste packaging recycling prediction model system, and compare it with other classification methods through specific logistics data experiments. The prediction accuracy, stability and advantages of the four algorithms are analyzed and compared, and the application reliability of the algorithm proposed in this paper to the logistics waste recycling process is verified. The application in the actual express logistics packaging recycling case shows the feasibility and effectiveness of the waste recycling scheme proposed in this paper.

Keywords: waste recycling; e-logistics; data-driven; long-short-term memory network; recycling path planning

1. Introduction

In today's growing e-commerce, the increase in express-package recycling has become an important factor affecting the green environment. Most of this express packaging can be recycled and reused, except for a small amount of fresh product packaging, disposable packaging and hazardous chemicals, which are polluting and non-reusable. Many logistics pickup points have set up recycling stations, but the use of recycling stations does not achieve satisfactory results for the convenience of users. Therefore, many logistics parks send out recycling vehicles to collect express-package recycling from the service area for centralized processing and secondary use. However, due to the effect of irregular purchase frequency by users, the vehicles are not economical to use, resulting in the recycling process being not environment-friendly. Therefore, we focus on the collection and transportation process of logistics waste, and propose a reasonable method to establish a logistics waste prediction model through the historical data of logistics waste recycling stations, to realize reasonable route planning and scheduling of logistics waste recycling vehicles, to reduce the costs in the collection and transportation process and to promote an efficient waste recycling mechanism and an advanced sustainable management method of logistics waste recycling.

In 1997, Hochreiter and Schmidhuber first proposed the LSTM model [1]. The LSTM network is one of the most successful RNN architectures. Different from a traditional recurrent neural network (RNN), LSTM is very good at learning time series of arbitrary length and making predictions. In addition, the problem of vanishing gradients in RNNs can be solved to some extent by keeping temporal information about the storage unit at all times. LSTM introduces the storage unit, which is a computational unit rather than a traditional artificial neuron. Through the LSTM memory unit, the neural network can connect memory and remote input, so as to dynamically master the time series data structure with high predictive ability [2]. Evidence shows that LSTM is more effective than traditional RNNs [3,4]. LSTM has achieved excellent results in many fields, such as machine translation and image generation, and has been widely used. McDougall [5] used life cycle inventory tools to conduct life cycle assessment on the whole process of municipal solid waste (MSW) treatment, and Arena [6] carried out life cycle assessment research on waste logistics. In addition, Bautista and Pereira [7] studied the optimization algorithm of collection point selection by taking MSW as a recycling logistic problem, and analyzed the optimization and decision-making of MSW logistics, such as the optimization of the design of waste management planning [8]. In 2011, Polat and Savas [9] proposed a data preprocessing method combining a subtraction clustering attribute weighting method and a classifier algorithm. In 2012, Sun and Genton [10] used a functional box diagram to detect abnormal data based on the data obtained by a road detector. In 2014, Chiou et al. [11] used a functional principal component analysis to estimate missing traffic data based on the data obtained using induction coils. In 2015, Jin et al. [12] proposed an abnormal data identification method for the data acquired using a microwave detector and introduced data denoising to improve it. In 2016, Deb and Liew [13] proposed a new data repair method, which uses the correlation within and between records to repair missing data. In 2018, Deb and Liew [14] proposed a new algorithm called Noise Cleaner for identifying abnormal data. Mao, et al. [15] proposed a robust low-rank representation method, which combines time prior information to estimate the missing data and improves the global correlation characteristics of data. Amzal et al. [16] proposed a widely used method in pattern recognition and machine learning, i.e., feature selection optimization based on the mutual information (MI) method, and the maximum feature term frequency mutual information (MTF-MI) method. In other words, a distributed feature selection

method combining feature term frequency and mutual information technology is applied in the experiment to improve the quality of the feature subset used for text classification. Finally, according to the experimental results, the proposed method is improved in both macroscopic F1 and microscopic F1. Di Sarli [17] proposed a completely untrained model. In this case, after testing whether the sentences generated recursively can be dynamically embedded, a simple machine learning algorithm is used to classify the text. Currently, it is obvious that the model cannot only greatly reduce the training time, but also that the classification effect of the model has advantages. According to the experimental results, the proposed model has certain advantages compared with other machine learning algorithm models. Mehta et al. [18] combined different feature selection methods and selected two different levels of aggregator for exploration, namely univariate and multivariate aggregators. The experiment proved that the method was stable in the practical application. Fiok [19] proposed a method called Text Guide for text truncation in order to reduce the length of the original text and reduce the length to the predefined shortest length. This method not only reduced the computational complexity of text classification, but also ensured the accuracy of text classification.

The data-driven based research methods have achieved a wide range of applications in the fields of uncertainty modeling, inverse problems and image or signal processing. In uncertainty modeling, data-driven methods mainly rely on large amounts of experimental data to reveal the intrinsic structure and laws of the system, thus avoiding many assumptions and approximations in traditional physical modeling methods [20,21]. For example, deep learning models are able to learn the dynamic behavior of a system directly from data rather than based on priori physical knowledge. The data-driven approaches also play an important role in the field of inverse problems. While traditional inverse problem approaches often need to be based on some physical model, data-driven approaches can recover unknown inputs or states directly from observed data, which greatly simplifies the problem-solving process. A neural network model can be trained as an inverse operator that estimates the input signal directly from the output data. In the field of image and signal processing, data-driven approaches have become mainstream [22]. The deep learning models such as convolutional neural networks have achieved excellent results in tasks such as image classification and semantic segmentation. In the field of signal processing, models, such as RNN and LSTM, have also demonstrated strong capabilities in applications, such as speech recognition and time series prediction. The data-driven methods in discrete tomography can provide more accurate and faster image reconstruction techniques. While traditional image reconstruction methods often require complex algorithms and a large amount of computation time, deep-learning-based methods can accomplish image reconstruction in a short period of time and with higher image quality [23,24]. Hybrid systems and state switching are important directions in the study of dynamical systems. These systems often contain both continuous and discrete components, making both their modeling and control very complex [25,26]. The data-driven approaches provide a new perspective, allowing researchers to learn the behavior of systems directly from data, rather than relying on complex mathematical models. Research in networks and systems has also benefited from data-driven approaches. For example, in social network analysis, machine learning methods can be used to identify important nodes or community structures in the network, while in the modeling and optimization of complex systems, data-driven methods provide a more flexible and efficient means [27,28]. In short, whether it is an old research method or a new technology, data-driven is providing new ideas and tools for scientific research and engineering applications. With the development of big data technology, we have reason to believe that more breakthroughs and progress will be made in the future based on data-driven research methods.

All of the above studies have classified and analyzed logistics data for improving logistics services, but there is no research related to the use of e-logistics information to improve the recycling rate of express-package recycling. Similar to the research in this paper, research on medical waste [29] management and sustainability started to emerge in the context of the popularity of COVID-19 [30,31], but more attention was paid to the process of waste disposal, while the logistic process of recycling was not sufficiently researched. In this paper, based on the previous methods related to data classification and analysis, we propose the LSTM based on an improved empirical mode decomposition (IEMD) algorithm to process addresses and categories in logistics data, study the concentration range of logistics dispatches and the distribution of recyclable express package recycling in the service area of logistics parks and optimize the routes of recycling trucks by judging the value of recyclable express package recycling, in order to improve the success rate of express package re-cycling.

The major contributions of this paper are:

- An advanced IEMD-LSTM algorithm is proposed in order to suppress logistic waste recycling data with nonlinearity and uncertainty;
- Aiming at the problem of uneven data volume due to too little logistics data in certain regions, a corresponding data expansion method is proposed;
- Experiments on real logistics data have proven that the method proposed in this paper has an excellent role in logistics waste recycling data management;
- The simulation results of logistics waste recycling vehicles in real cases show that the success rate of waste recycling can be improved by using the recycling path planning method proposed in this paper.

2. IEMD-LSTM method

This section describes the methodology structure and improvement strategies used in this study. First, the basic structure of LSTM is described, and then the derivation of improved EMD is explained.

2.1. The LSTM prediction model

A recursive neural network is a kind of deep learning network particularly used for time series problems. Compared with the traditional deep neural network, it has significant advantages in nonlinear time series prediction. Too many storage layers in a recurrent neural network (RNN) will affect the speed of network training, to a certain extent, and will lead to gradient disappearance and gradient explosion. The gradient disappearance or explosion will cause the shallow weight not to be updated, which makes the RNN lack the ability to remember long-term input time information. In short, the traditional RNN model has the defect of long-term dependence when dealing with long-term sequence problems. With regard to this phenomenon, Hochreiter and Schmidhuber proposed an LSTM network structure model [1]. This model is a new network model based on RNN. LSTM is equipped with an input gate, a forgetting gate and an output gate. Through the gate connection, a hysteresis connection is established between the input and the feedback, and continuous flow error is forced to be maintained in the circulating neuron. Finally, the gradient disappearance or explosion caused by derivative multiplication is reduced, and the memory ability of the deep learning network is effectively improved.

For a time series, $x = (x_1, x_2, \dots, x_n)$, traditional RNN neurons output a state sequence, $h = (h_1, h_2, \dots, h_n)$. The calculation process can be expressed as:

$$h_n = f(W_{xh}x_n + W_{hh}h_{n-1} + b_h) \quad (1)$$

In Eq (1), W and b are the weights and deviations of RNN neurons, respectively, and f represents the nonlinear activation function. Different from the general RNN model, the structure of the LSTM model in the self-circulation part is modified. These modified complex structures can help the model return to the model input of previous nodes, which can be called long-term memory. Its structure is shown in Figure 1.

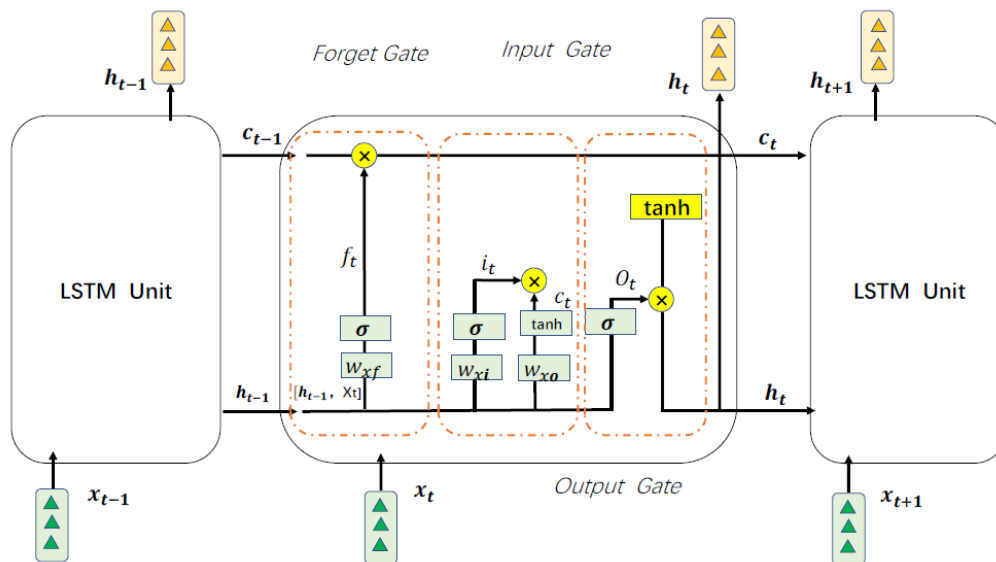


Figure 1. Structure diagram of LSTM.

In Figure 1, the meanings of i , f , o and c are, respectively, the input gate, forgetting gate, output gate and state function of the LSTM model and \tanh as the activation function. The structure of the input gate, forgetting gate and output gate can be written in the forms shown in Eqs (2)–(4) [1]:

$$i_n = \sigma(W_{xi}x_n + W_{hi}h_{n-1} + b_i), \quad (2)$$

$$f_n = \sigma(W_{xf}x_n + W_{hf}h_{n-1} + b_f), \quad (3)$$

$$o_n = \sigma(W_{xo}x_n + W_{ho}h_{n-1} + b_o). \quad (4)$$

The state of the neuron of the neural network in the LSTM structure is represented by c , which can be calculated by the following equation.

$$c_n = f_n c_{n-1} + i_n \tanh(W_{xc}x_n + W_{hc}h_{n-1} + b_c). \quad (5)$$

where W is the weight matrix of the corresponding neural network structure, and b is the data deviation in the corresponding structure.

The LSTM can handle sequential data more efficiently, and its design principles allows it to memorize information over long periods of time and to deal with gradient vanishing and gradient explosion problems. Therefore, the LSTM is a good choice for applications that need to deal with long sequential data such as logistics data or applications that need to capture long-term dependencies. The key structure of LSTM is the “gate” mechanism, which allows LSTM to precisely control the flow of

information in a cell. This flexible approach to information management allows the LSTM to capture complex patterns and long-term dependencies. Therefore, when faced with logistics data with long time intervals and poor continuity or with complex structural patterns in the data, the LSTM has an advantage over the traditional RNN.

2.2. The IEMD algorithm

The precondition of the Fourier transformation method for data processing, which is widely used in engineering, is that the signal must be stable and linear. However, due to the complexity of the supply chain, logistics data do not meet these conditions. For logistics data with non-stationary characteristics, the frequency spectrum obtained based on Fourier transformation does not contain time information. Therefore, in the process of Fourier transformation spectrum analysis, it is impossible to effectively identify the change in frequency with time, and a lot of information will be lost, resulting in distortion of the analysis results. Therefore, it is necessary to apply the EMD algorithm, which is equivalent to starting processing in the original time series. The core content is to calculate the sum of the intrinsic mode function component and a trend function component, which can smoothen the original time series.

Traditional EMD often has serious errors in logistics data prediction. In transportation, especially in terminal logistics transportation, there are often specific situations that are difficult to consider using the EMD model. This specific situation is often a mixture of multiple factors, such as weather, traffic, quarantine, natural disasters and other single reasons, when long-distance transportation is considered. However, factors affecting terminal transportation often occur in combination. For example, weather problems will lead to traffic jams and natural disasters. Such mixed factors often determine the delay time. This leads to large fitting errors. These expected large errors will lead to poor recovery efficiency of logistics waste, which has negative effects on both low-carbon environmental protection and the economy. Therefore, an improved EMD algorithm is proposed to reduce the fitting error of the model.

$$h_n = o_n \tanh(c_n). \quad (6)$$

A component of $h_1(t)$ is defined as Eq (7).

$$h_1(t) = x(t) - m_1(t). \quad (7)$$

$h_1(t)$ is decomposed k times until $h_{1i}(t)$ is the intrinsic mode function (IMF).

$$\begin{cases} h_{11} = h_{10} - m_{11} \\ h_{12} = h_{11} - m_{12} \\ \vdots \\ h_{1i} = h_{1(i-1)} - m_{1i} \end{cases}. \quad (8)$$

The first IMF, $c_1(t)$, is defined as $c_1(t) = h_{1i}(t)$. Generally, the time-characteristic scale of the first IMF is small, i.e., the frequency of the first IMF is the highest. Therefore, if the first IMF is separated from the original signal, $x(t)$, the remaining $r_1(t)$ is written as:

$$r_1(t) = x(t) - c_1(t). \quad (9)$$

To separate the nonlinear part of $x(t)$, the above process is repeated until the n -th IMF meets the requirement that the remaining $r_n(t)$ is an approximate monotonic function.

$$\begin{cases} r_1(t) - c_2(t) = r_2(t) \\ r_2(t) - c_3(t) = r_3(t) \\ \vdots \\ r_{N-1}(t) - c_N(t) = r_N(t) \end{cases} \quad (10)$$

The original output of regional logistics big data can be defined as

$$x(t) = \sum_1^N c_i(t) + r_N(t), i = 1, 2, \dots, N. \quad (11)$$

However, in the decomposition process, the envelope obtained by connecting the extreme points does not usually represent the exact extreme points. Therefore, the upper and lower envelopes at the endpoints will diverge, which will introduce large fitting errors in the decomposition. As the IMF decomposes, errors will accumulate. Therefore, an EMD extreme point prediction strategy based on extreme point correlation is proposed. The calculation steps are as follows:

- 1) Calculate each maximum, $q_{\max}(i)$, and minimum, $q_{\min}(i)$. $q_{\max}(\text{end})$ and $q_{\min}(\text{end})$ are extreme points on the far right. Define the error set as

$$E = \{ |q_{\max}(i) - q_{\max}(\text{end})|, |q_{\min}(i) - q_{\min}(\text{end})| \} i = 1, 2 \dots n. \quad (12)$$

- 2) Calculate the left monotonicity of the rightmost endpoint and compare it with the left uniqueness of each point in the error set. If the monotonicity is different, delete the corresponding points.
- 3) Use the remaining m maximum points to form Eq (13).

$$\begin{cases} A_m q_{\max}(j_1) + A_{m-1} q_{\max}(j_1 - 1) + \dots + A_0 q_{\max}(j_1 - m) = q_{\max}(j_1 + 1) \\ A_m q_{\max}(j_2) + A_{m-1} q_{\max}(j_2 - 1) + \dots + A_0 q_{\max}(j_2 - m) = q_{\max}(j_2 + 1) \\ A_m q_{\max}(j_m) + A_{m-1} q_{\max}(j_m - 1) + \dots + A_0 q_{\max}(j_m - m) = q_{\max}(j_m + 1) \end{cases} \quad (13)$$

- 4) Predict the next maximum point of the rightmost point, $q_{\max}(\text{end} + 1)$, according to Eq (14).

$$A_m q_{\max}(\text{end}) + \dots + A_0 q_{\max}(\text{end} - m) = q_{\max}(\text{end} + 1). \quad (14)$$

Based on the same strategy, $q_{\min}(\text{end} + 1)$, $q_{\max}(\text{start} - 1)$ and $q_{\min}(\text{start} - 1)$ can be calculated. Therefore, the key aim of the proposed strategy is to predict the extreme points of the data.

3. Sample expansion method

The data in this paper come from an enterprise with its own logistics data server. Due to different distributions, the logistics data samples collected from open source datasets cannot be applied to actual scenarios. Therefore, the model needs to be trained and used to detect the real-world delivery logistics data. Thus, it is very important to label the data in the actual data environment and obtain a training data set that meets the needs. As data tagging is a relatively complex work, a small amount of tag data is obtained through manual methods, particularly for malicious samples.

Next, K Nearest Neighbors (KNN) and k-means are used to expand the samples of the entire dataset on the basis of these few labeled data. Then, extended samples highly similar to manually labeled data were used for the subsequent analysis.

A network electronic logistics data feature extraction algorithm network for electronic logistics data has the characteristics of broadcast, so there are a lot of similar network electronic logistics data in the electronic logistics data server. They may come from the same IP or the same address. A feature

extraction algorithm for electronic logistics data based on a seven-tuple set is proposed, and the electronic logistics data are marked with this algorithm. The features mostly consist of the following two parts:

Title features: These features include the real source IP, the real sender address and the consistency between the real address and the displayed sender address. Since the displayed mailing address can be forged, the true information about the recipient is very important and should be considered. This part can be obtained from the *.eml file in the electronic logistics data server.

Content function of e-logistics data: This part includes the e-logistics data title, e-logistics data attachment name, e-logistics data attachment suffix and whether e-logistics data include the goods type. If the cargo type is included, it is first determined whether the cargo type is long or short, and then whether the cargo type is correctly judged by the cargo type, because the sender often uses this method in the electronic logistics data.

The e-logistics data samples can be vectorized based on the e-logistics data feature extraction algorithm, and then the e-logistics data clustered to obtain an accurate marking training data set, so that the e-logistics data detection algorithm can accurately and efficiently identify the e-logistics data.

3.1. Improved levenshtein distance

The edit distance represents the minimum number of times a single character needs to be deleted, inserted, or replaced from s to t . For two strings a and b , their lengths are $|a|$ and $|b|$, their Levenshtein Distance, $L_{a,b}(|a|, |b|)$, is defined as

$$L_{a,b}(|a|, |b|) \begin{cases} \max(i, j) & \text{if } \min(i, j) = 0 \\ \min \begin{cases} L_{a,b}(i-1, j) + 1 \\ L_{a,b}(i, j-1) + 1 \\ L_{a,b}(i-1, j-1) + 1(a_i \neq b_j) \end{cases} & \text{otherwise} \end{cases} \quad (15)$$

when $a_i = b_j$ is 0 or, otherwise, is 1. $L_{a,b}(i, j)$ is the edit distance between the first i characters of a and the first j characters of b . The similarity of a and b can be expressed as $S_{a,b}$,

$$S_{a,b} = 1 - (L_{a,b}(|a|, |b|) / \max(|a|, |b|)). \quad (16)$$

Through this string distance, e-logistics data can be clustered, eliminating the problem of feature loss and accurately grouping the e-logistics data.

3.2. Sample labeling algorithm

The k-means algorithm [32] is a classical partition-based clustering algorithm. The core idea is that all data are clustered around k points in the space, and all clusters will update their center values iteratively until the best clustering result is obtained.

The central idea of the KNN algorithm [33] is that, when the data in the training set and their labels are known, input the test data, compare the characteristics of the test data with the corresponding characteristics in the training set, and find the first k data that are most similar to the training set. The test data category is the most common category of k data.

The KNN is a theoretically mature classification algorithm. It is an algorithm based on the idea

of template matching, which is simple but effective and is still being used today on a number of simple problems. However, KNN is quite time-consuming because it is quite computationally intensive, so it is necessary to use the feature of k-means, an unsupervised clustering technique that does not require the use of training data for learning, to preprocess the data once and reduce the complexity of model prediction.

After the sandbox detects all the sample data, it is divided into recyclable areas and insufficient recovery areas. Owing to the high false alarm rate of sandboxes, sandboxes are classified as insufficient recovery areas containing normal samples, whereas data classified as recyclable areas contain insufficient recovery areas.

There is a certain amount of similar data in the logistics data server, so it can be effectively clustered by a clustering algorithm. The characteristics of these invalid data are displayed in the form of string, such as address area and cargo name, so the above string distance is used as the distance of the clustering algorithm. A k-means algorithm and a KNN algorithm are used to cluster and reclassify the sandbox results.

The results are defined by the following rules: the first character represents the logistics data type judged by the sandbox; “*p*” represents logistics data from non-recyclable logistics waste, and “*n*” represents logistics data from recyclable logistics waste. The second character represents the logistics data type judged by the algorithm; “*p*” represents logistics data that are classified as non-recyclable logistics waste and “*n*” represents logistics data that are judged as recyclable logistics waste. The third character represents the algorithm used. “1” is the result of the k-means algorithm, and “2” is the result of the KNN algorithm. For example, “*pp*₁” means that the logistics data are non-recyclable logistics waste, that is classified as non-recyclable logistics waste by the k-means algorithm. To ensure the size and reliability of the dataset, the Eqs (17) and (18) are used to obtain the extended dataset. Non-recyclable samples in the dataset are calculated as follows:

$$\text{Non-recyclable} = pp_1 \& pp_2 + np_1 \& np_2. \quad (17)$$

Recyclable logistics samples in the dataset are calculated as follows:

$$\text{Recyclable} = pn_1 \& pn_2 + pn_1 \& nn_2. \quad (18)$$

In Eqs (21) and (22), “*a* and *b*” denote the intersection of *a* and *b*, “*a* + *b*” denotes the union of *a* and *b*.

4. Results and discussion

The experiments in this paper are divided into two parts. In Section 4.2, the validation of the proposed IEMD-LSTM method for analyzing logistics waste recycling data is discussed, and in Section 4.3, the results of using the predictive model built from the analyzed data in an actual logistics waste recycling scheduling application are discussed.

4.1. Experimental platform and data source

In this experiment, first real-time logistics data have been collected from the logistics park server and supplemented the data from some logistics distribution points and logistics delivery points. Email data were collected from January 2021 to December 2021. The experiment used MATLAB 2020a as

the neural network framework to build the network. The computer of the simulation platform was an Intel Core, i7-7700 @ 4.20 GHz, 8 GB RAM.

4.2. Experimental result

In the experiment, the experimental results were evaluated by four parameters, namely the Accuracy (*Acc*), Precision (*P*), Recall (*R*) and F1 score (*F1*). These four parameters are defined as follows:

$$Acc = \left(1 - \frac{error}{sum}\right) \times 100\%, \quad (19)$$

$$P = \frac{T_P}{T_P + F_P} \times 100\%, \quad (20)$$

$$R = \frac{T_P}{T_P + F_N} \times 100\%, \quad (21)$$

$$F1 = \frac{2 \times P \times R}{P + R} \times 100\%. \quad (22)$$

In Eq (19), “error” represents the number of samples with wrong classification, and “sum” represents the total number of samples. In Eqs (20)–(22), T_P is true, represents the amount of waste that cannot be recycled; F_P and F_N are the number of false negatives and false positives; R represents the recall rate, i.e., represents the non-recyclable quantity of waste correctly classified by the model; the $F1$ score is based on the harmonic average of accuracy and recall to comprehensively evaluate the performance of the model.

The express delivery category is selected in a delivery area in the logistics data, used 50 weeks of data extracted in a year for classification, and the recyclable logistics waste was divided into five standards. In order to show the superiority of the classification accuracy of the IEMD-LSTM method proposed in this paper, it was compared with three popular classification algorithms, RNN [34], Bi-LSTM [35] and GRU [36].

From Figure 2, the classification accuracy of the two LSTM methods is high, and excellent compared to the traditional RNN and GRU methods. However, one of them, Bi-LSTM, had the problem of a sudden drop in accuracy and did not have the advantage of the IEMD-LSTM proposed in this paper with regard to stability. The classification accuracies of the RNN, Bi-LSTM, GRU and IEMD-LSTM were 81.82%, 87.88%, 79.8% and 91.92%, respectively.

As can be seen in the prediction of the global recovery necessity index, the prediction accuracy of the four methods slowly deteriorates with the increase in disturbing logistics information, as shown in Figure 3. Among them, it is obvious that the stability of Bi-LSTM is worst among the four methods, but it has certain advantages in prediction accuracy. From Table 1, it can be seen that the IEMD-LSTM proposed in this paper has obvious advantages in prediction accuracy and stability compared to the other three methods, which could provide powerful help in improving the success rate and efficiency of logistics scrap recycling. It is worth noting that the recovery costs evaluation index of the target data is evaluated according to the supervisory judgment of the recovery personnel, which is used as the evaluation standard in the experiment.

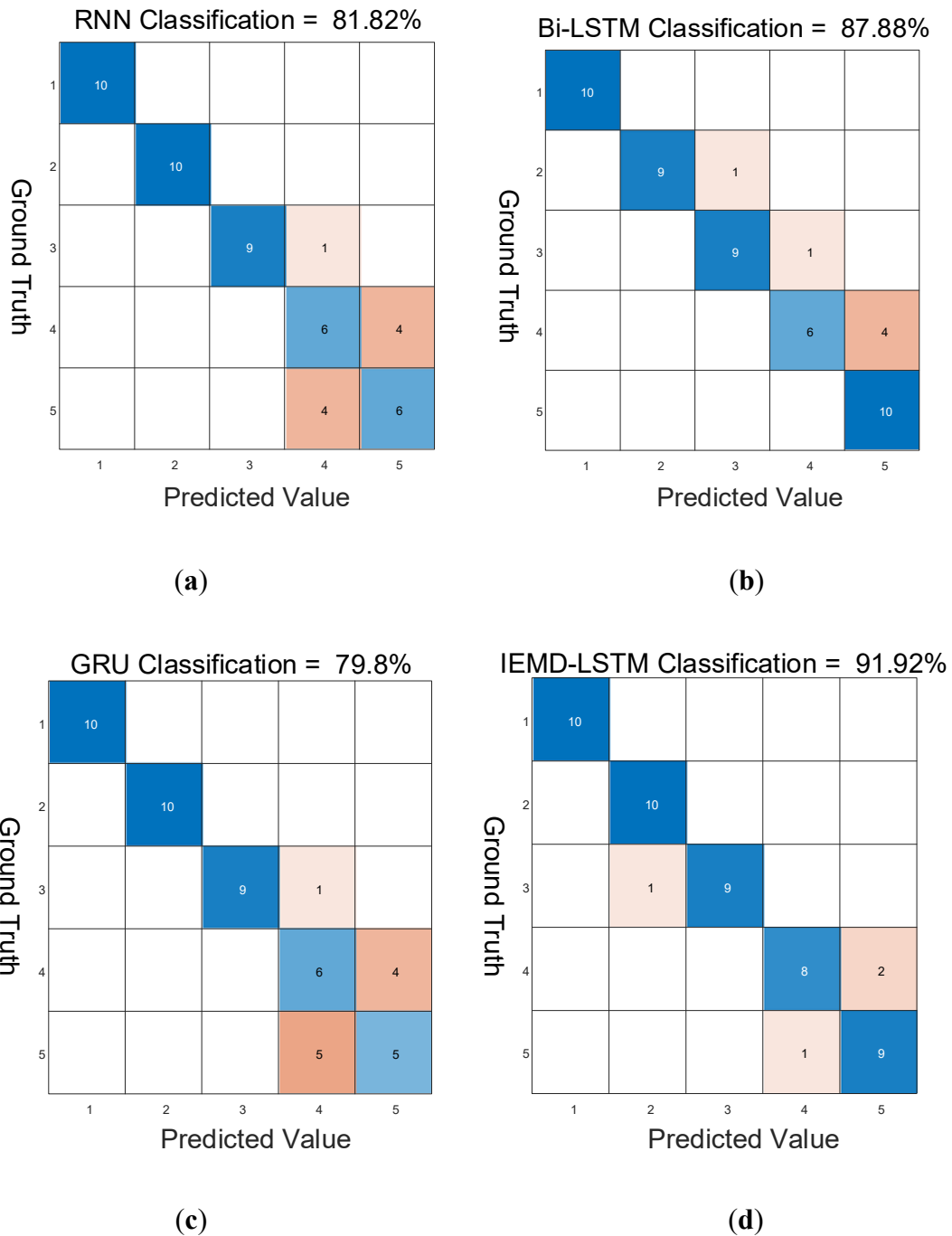


Figure 2. Logistics waste recycling classification results. (a) Results using RNN; (b) Results using Bi-LSTM; (c) Results using GRU; (d) Results using IEMD-LSTM.

Table 1. Evaluation of accuracy of classification results of four methods.

Method	MSE	RMSE	MAE	Std
RNN	0.15775	0.39718	-6.44e-16	0.40121
Bi-LSTM	0.13713	0.37032	-0.22915	0.29386
GRU	0.15264	0.39069	0.039681	0.39262
IEMD-LSTM	0.10141	0.31845	0.072915	0.18419

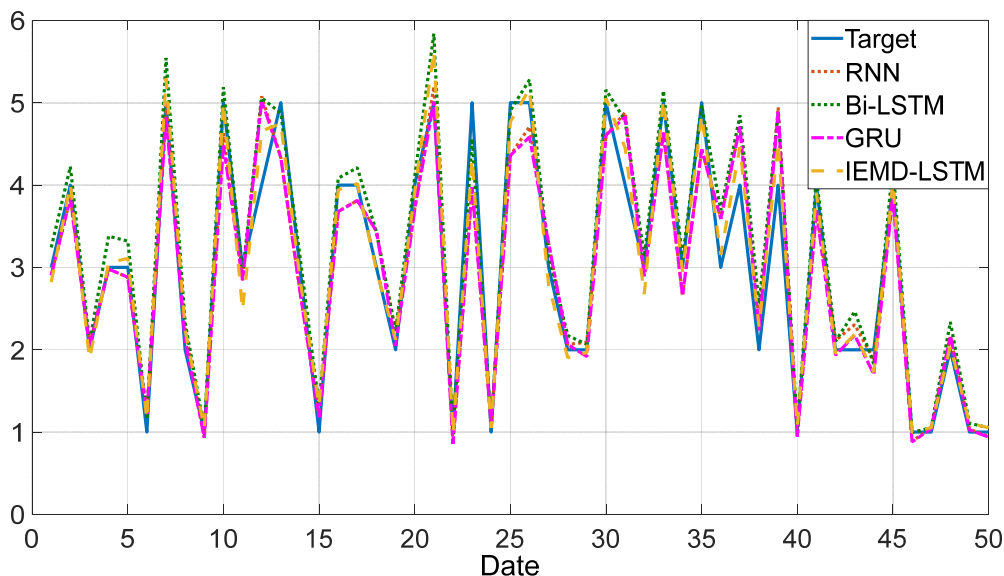


Figure 3. Data prediction results.

Figure 4 shows the performance of the IEMD-LSTM when using various activation functions. The traditional sigmoid activation function in general provides better accuracy, recall and F1 score in LSTM than relu, selu, softmax and softplus activation functions. However, in the case of this paper, the softplus activation function in LSTM has better accuracy. Figure 5 shows the performance of the EMD-LSTM using Adam, sgd, Nadam, Rmsprop and Adagrad optimizers. The Adam optimizer performs better than the other optimizers used in the LSTM.

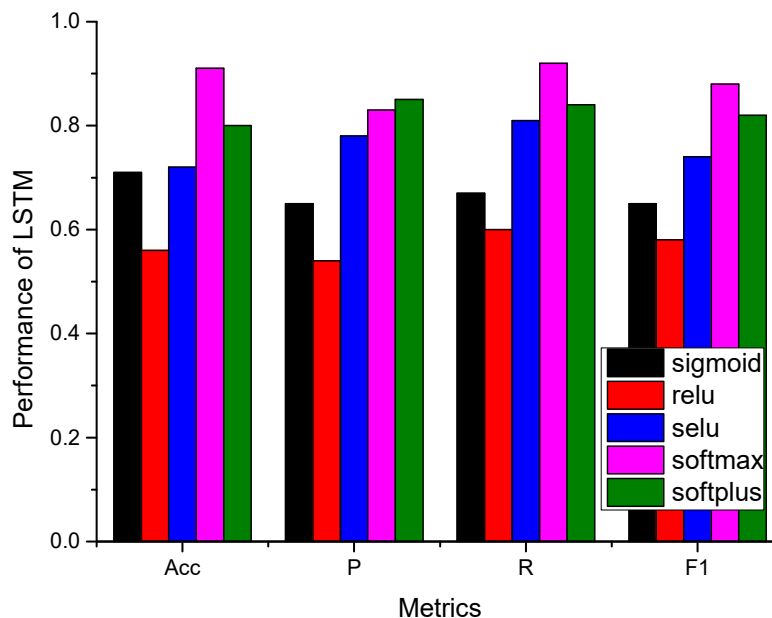


Figure 4. IEMD-LSTM with different activation functions.

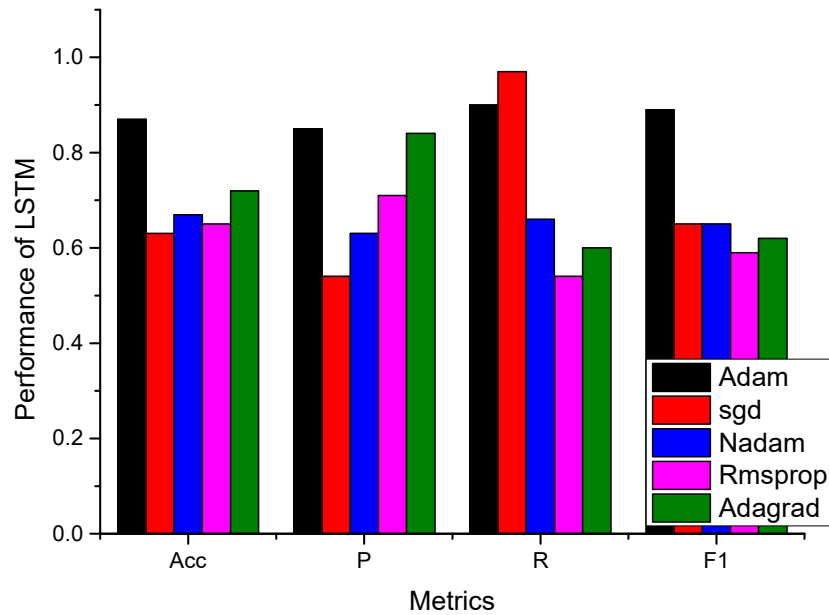


Figure 5. IEMD-LSTM with different optimizers.

Through the above analysis, the IEMD-LSTM proposed in this paper has obvious advantages in classification accuracy and prediction accuracy can be inferred. Moreover, to better apply the method, the performance of the proposed IEMD-LSTM with different activation functions and optimizers are compared. The experimental results demonstrate that the soft plus activation function combined with the Adam optimizer can achieve satisfactory results. It can improve the efficiency of logistics scrap recycling.

4.3. Waste recycling strategy

The classification and prediction accuracy of the IEMD-LSTM proposed in this paper on logistics waste recycling addresses have already been discussed. Hereafter, what advantages the method proposed in this paper has in performing the recycling process will be illustrated through an example of logistics waste recycling in a logistics park.

The logistics park shown in Figure 6 was used in the experiment. A logistics park in Liaoning province is chosen, which has 15 major recycling points in its service area. First, the express packaging recycling data of the 15 logistics waste recycling points were used in the experiment. The yellow five-pointed star was the logistics distribution station. The distribution of the 15 recycling points involved residential areas, campuses and industrial parks. According the accessibility of the major roads, the 15 logistics waste collection points are divided into 6 collection areas, which can be considered as the recycling vehicles without additional transportation costs for recycling within the zones, and the long-distance transportation costs will be calculated in case of transfer between the collection areas. Their data characteristics were different in different time periods. It is difficult to process such strongly nonlinear data with the general data analysis method, Therefore, the method proposed in this paper is adopted in the experiment, and the corresponding path planning method is selected to improve the logistics waste recovery scheme proposed in this paper.

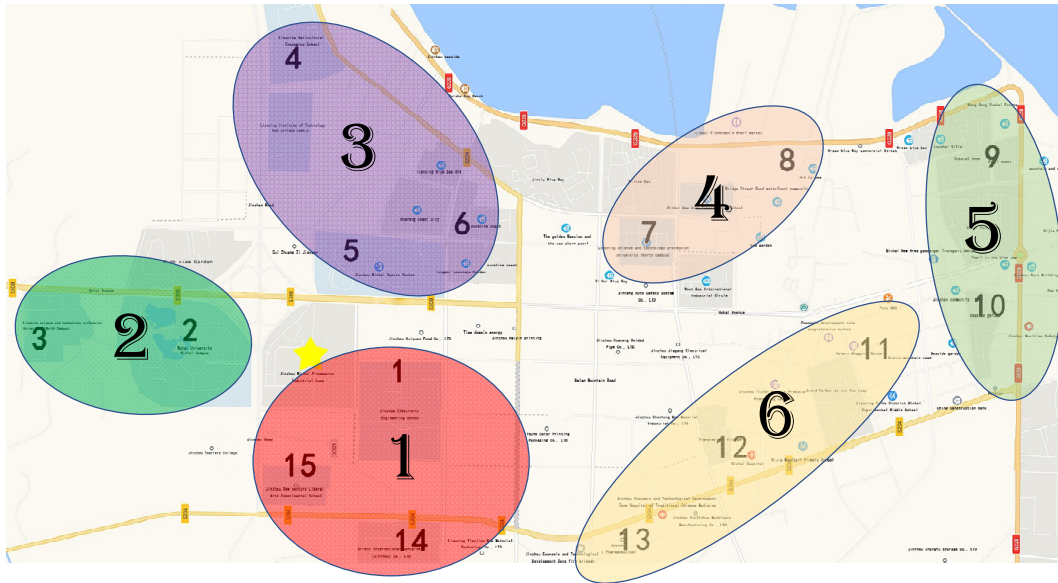


Figure 6. Map of logistics park in the case.

Two new intelligent swarm optimization algorithms are used, Golden eagle optimizer (GEO) [37] and Reptile Search Algorithm (RSA) [38], to find the optimal recycling path in the range and compare the speed of finding it with that of the classical PSO algorithm. The results are shown in Figure 7.

As can be seen in Figure 7, the RSA algorithm has a clear advantage over the other two algorithms in terms of path finding efficiency and optimization degree for logistics waste recycling. Of course, the path finding effect can be also improved by other improvement algorithms and by introducing improvement mechanisms. However, our main research goal was to classify and predict the logistics data for the purpose of waste recycling, not to optimize the recycling path, and the advantages of the RSA algorithm are given in this paper mainly to provide a complete system of reference for related enterprises and logistics parks.

The valid data for seven days in February, June, August and November were randomly selected from the historical data of the 15 logistics waste collection sites in 2021 as the basis for the experiments, where the data in June were used to train the model and the rest of the data were used to validate the illustration of the cost utilization of the logistics waste collection vehicles operating in the six collection areas. In addition to this, it is also compared with other prediction methods to demonstrate the positive effect of the proposed IEMD-LSTM in logistics waste recycling. As shown in Figure 8, the success rate of recycling using the proposed method in this paper is 100% in the regional recycling sites numbered 1, 2, 3 and 5, and the success rate in Regions 4 and 6 is also above 90% and much higher than the other three methods, although there are a number of failures. This shows that the classification and prediction of logistics data using the method proposed in this paper can provide effective support for logistics waste recycling, avoiding the need to traverse all regions to find recyclable logistics waste each time. Using the method in this paper not only improves the economy, but also reduces the driving distance of recycling vehicles, contributing to green and low carbon goals.

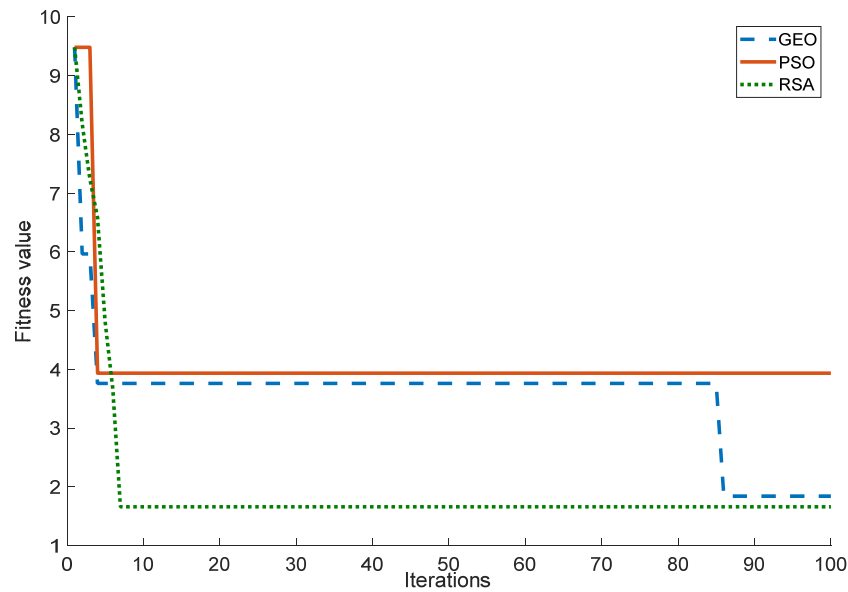


Figure 7. Comparison of different algorithms in recycling path optimization efficiency.

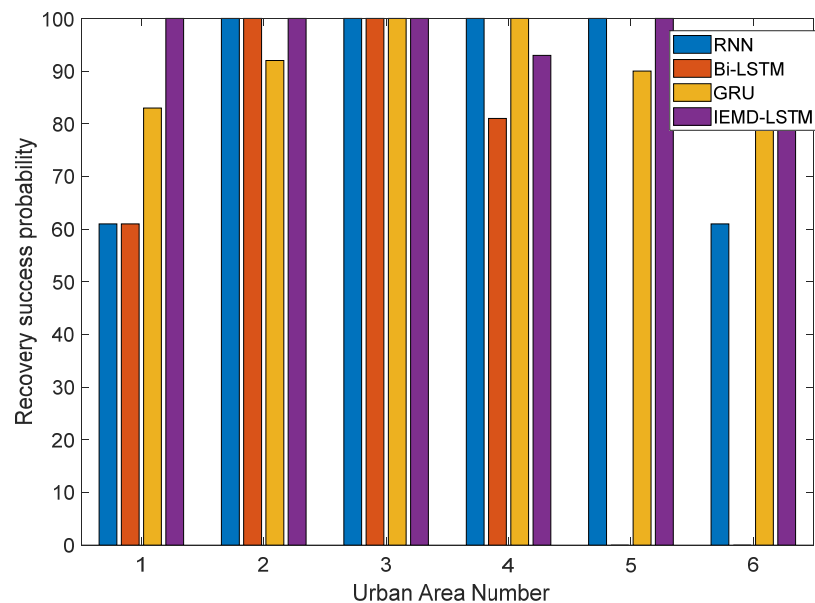


Figure 8. Comparison of the recovery success rate of different classification prediction methods.

The items used in the historical data will be used for costs calculation as the criteria for evaluating the costs, as shown in Table 2. Since some of the costs vary over time such as fuel consumption, the average is curved and rounded to the nearest whole number to facilitate the calculation, and the absolute error between the results of the data processing and the actual data is less than 10%.

Table 2. Logistics waste recycling costs.

Costs items	Costs value
Logistics waste recycling	150 USD/t
Logistics waste accumulation	-80 USD/day
Fuel consumption of recovery vehicles	-1 USD/km
Hours of work for recycling staff	-8 USD/h

Table 3. Comparison of the earnings of the optimal recovery strategy based on four methods.

Method	Waste recycling strategy	Earnings (USD/week)
RNN	2→(3)→1→6→5→4	-796
Bi-LSTM	1→[6]→5→4→3→2	-1410
GRU	(3)→2→1→6→(5)→4	-484
IEMD-LSTM	(2)→3→4→(5)→6→1	-112

As shown in Table 3, the application gains the optimal recovery strategy planned on the basis of the predictive model developed based on the four methods. “(·)” means recycling every two days and “[·]” means recycling every three days. The experimental results show that recycling logistics waste will not bring profit, but accurate predictions of the stock of logistics waste in the recycling area can avoid duplicate paths and reduce the working time of the staff, thus effectively saving costs and avoiding the secondary waste caused by the behavior of logistics waste recycling and management decision errors. It can be seen that applying the method proposed in this paper to establish a prediction model for the amount of logistics waste, and then planning the process of urban logistics waste recycling management, is an effective way to reduce logistics waste recycling costs. The effectiveness and feasibility of the method in this field are proven.

5. Conclusions

The recycling of express packaging plays a vital role in protecting the environment and saving resources. Based on the design and construction of logistics waste recycling path scheduling, we establish a data-driven courier packaging recycling system and propose an IEMD-LSTM prediction model with higher prediction accuracy and better stability. This method is used to conduct modeling and simulation research on the relevant logistics data of each node participating in the express recycling behavior. It provides a basis for finding the most mileage-saving route for recycling vehicles and a new method and idea for solving similar problems, such as making the recycling of express packaging more economical in scale.

In this paper, via collection and processing of the logistics data from the specific distribution center network, the output results under different simulation schemes have been comprehensively analyzed, and the prediction accuracy, stability, degree of influence and express-package recovery distribution have been discussed. Conclusions have been drawn about the probability of express package recycling activities, the total amount of recycling and the operating costs and benefits of the recycling system under different models, so as to provide valuable suggestions for decision-making about express package recycling. However, owing to limited research time, there are still some issues in this paper to be improved and perfected. When constructing the IEMD-LSTM model of express

packaging recycling, although the analysis of individual dependent variables had a theoretical basis, it is still subjective to some extent. It would be beneficial to carry out subsequent research from the perspectives of economics, sociology, psychology and so on. It is hoped that future research can expand the scope of this investigation and the number of samples, and conduct more-accurate and in-depth analysis of the IEMD-LSTM prediction model, in order to make the research results more valuable for reference. In addition, the method proposed in this paper can be extended to supply chain networks with broader applications [39,40], providing a new method and idea for research related to the accurate establishment of complex networks.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This research received funding supported by the Basic Scientific Research Project of Colleges and Universities of Education Department of Liaoning Province (grant number LJKMR20221890).

Conflict of interest

The authors declare no conflict of interest.

References

1. S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput.*, **9** (1997) ,1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
2. M. Roondiwala, H. Patel, S. Varma, Predicting stock prices using LSTM, *Int. J. Sci. Res.*, **6** (2017), 1754–1756.
3. H. Palangi, L. Deng, Y. Shen, J. Gao, X. He, J. Chen, et al., Deep sentence embedding using long short-term memory networks: Analysis and application to information retrieval, *IEEE/ACM Trans. Audio Speech Lang. Process.*, **24** (2016), 694–707. <https://doi.org/10.1109/TASLP.2016.2520371>
4. H. Palangi, R. Ward, L. Deng, Distributed compressive sensing: A deep learning approach, *IEEE Trans. Signal Process.*, **64** (2016), 4504–4518, 2016. <https://doi.org/10.1109/TSP.2016.2557301>
5. F. R. McDougall, J. P. Hruska, Report: The use of Life Cycle Inventory tools to support an integrated approach to solid waste management, *Waste Manage. Res.*, **18** (2000), 590–594. <https://doi.org/10.1034/j.1399-3070.2000.00159.x>
6. U. Arena, M. L. Mastellone, F. Perugini, The environmental performance of alternative solid waste management options: A life cycle assessment study, *Chem. Eng. J.*, **96** (2003), 207–222. <https://doi.org/10.1016/j.cej.2003.08.019>
7. J. Bautista, J. Pereira, Modeling the problem of locating collection areas for urban waste management: An application to the metropolitan area of Barcelona, *Omega*, **34** (2006), 617–629. <https://doi.org/10.1016/j.omega.2005.01.013>

8. O. Eriksson, M. C. Reich, B. Frostell, A. Björklund, G. Assefa, J. O. Sundqvist, et al., Municipal solid waste management from a systems perspective, *J. Cleaner Prod.*, **13** (2005), 241–252. <https://doi.org/10.1016/j.jclepro.2004.02.018>
9. K. Polat, S. S. Durduran, Subtractive clustering attribute weighting (SCAW) to discriminate the traffic accidents on Konya-Afyonkarahisar highway in Turkey with the help of GIS: A case study, *Adv. Eng. Software*, **42** (2011), 491–500. <https://doi.org/10.1016/j.advengsoft.2011.04.001>
10. Y. Sun, M. G. Genton, Adjusted functional boxplots for spatio-temporal data visualization and outlier detection, *Environmetrics*, **23** (2012), 54–64. <https://doi.org/10.1002/env.1136>
11. J. M. Chiou, Y. C. Zhang, W. H. Chen, C. W. Chang, A functional data approach to missing value imputation and outlier detection for traffic flow data, *Transportmetrica B: Transport Dyn.*, **2** (2014), 106–129. <https://doi.org/10.1080/21680566.2014.892847>
12. P. Jin, S. Parker, J. Fang, B. Ran, C. M. Walton, Freeway recurrent bottleneck identification algorithms considering detector data quality issues, *J. Transp. Eng.*, **138** (2012), 1205–1214. [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000424](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000424)
13. R. Deb, A. W. C. Liew, Missing value imputation for the analysis of incomplete traffic accident data, *Inf. Sci.*, **339** (2016), 274–289. <https://doi.org/10.1016/j.ins.2016.01.018>
14. R. Deb, A. W. C. Liew, Noisy values detection and correction of traffic accident data, *Inf. Sci.*, **476** (2019), 132–146. <https://doi.org/10.1016/j.ins.2018.10.002>
15. R. Mao, Z. Chen, G. Hu, Robust temporal low-rank representation for traffic data recovery via fused lasso, *IET Intel. Transport Syst.*, **15** (2021), 175–186. <https://doi.org/10.1049/itr2.12010>
16. H. Amazal, M. Kissi, A new big data feature selection approach for text classification, *Sci. Program.*, **2021** (2021), 6645345. <https://doi.org/10.1155/2021/6645345>
17. D. Di Sarli, C. Gallicchio, A. Micheli, Text classification by untrained sentence embeddings, *Intell. Artif.*, **14** (2020), 245–259. <https://doi.org/10.3233/IA-200053>
18. P. Mehta, S. Chandra., Robustness and predictive performance of homogeneous ensemble feature selection in text classification, *Int. J. Inf. Retr. Res.*, **11** (2021), 75–89. <https://doi.org/10.4018/IJRR.2021010104>
19. K. Fiok, W. Karwowski, E. Gutierrez-Franco, M. R. Davahli, M. Wilamowski, T. Ahram, et al., Text guide: Improving the quality of long text classification by a text selection method based on feature importance, *IEEE Access*, **9** (2021), 105439–105450. <https://doi.org/10.1109/ACCESS.2021.3099758>
20. E. Savku, G. W. Weber, A stochastic maximum principle for a markov regime-switching jump-diffusion model with delay and an application to finance, *J. Optim. Theory Appl.*, **179** (2018), 696–721. <https://doi.org/10.1007/s10957-017-1159-3>
21. S. Kuter, Z. Akyurek, G. W. Weber, Retrieval of fractional snow covered area from MODIS data by multivariate adaptive regression splines, *Remote Sens. Environ.*, **205** (2018), 236–252. <https://doi.org/10.1016/j.rse.2017.11.021>
22. A. Özmen, E. Kropat, G. W. Weber, Robust optimization in spline regression models for multi-model regulatory networks under polyhedral uncertainty, *Optimization*, **66** (2017), 2135–2155. <https://doi.org/10.1080/02331934.2016.1209672>
23. Ö. N. Onak, Y. S. Dogrusoz, G. W. Weber, Evaluation of multivariate adaptive non-parametric reduced-order model for solving the inverse electrocardiography problem: A simulation study, *Med. Biol. Eng. Comput.*, **57** (2019), 967–993. <https://doi.org/10.1007/s11517-018-1934-9>

24. A. Cevik, G. W. Weber, B. M. Eyuboglu, K. K. Oguz, Voxel-MARS: A method for early detection of Alzheimer's disease by classification of structural brain MRI, *Ann. Oper. Res.*, **258** (2017), 31–57. <https://doi.org/10.1007/s10479-017-2405-7>
25. B. Kalaycı, A. Özmen, G. W. Weber, Mutual relevance of investor sentiment and finance by modeling coupled stochastic systems with MARS, *Ann. Oper. Res.*, **295** (2020), 183–206. <https://doi.org/10.1007/s10479-020-03757-8>
26. P. Taylan, F. Yerlikaya-zkurt, B. Bilgi Uak, G. W. Weber, A new outlier detection method based on convex optimization: Application to diagnosis of Parkinson's disease, *J. Appl. Stat.*, **48** (2021), 2421–2440. <https://doi.org/10.1080/02664763.2020.1864815>
27. R. Lotfi, Z. Yadegari, S. H. Hosseini, A. H. Khameneh, E. B. Tirkolae, G. W. Weber, A robust time-cost-quality-energy-environment trade-off with resource-constrained in project management: a case study for a bridge construction project, *J. Ind. Manage. Optim.*, **18** (2020), 375–396. <http://doi.org/10.3934/jimo.2020158>
28. S. Özögür-Akyüz, B. C. Otar, P. K. Atas, Ensemble cluster pruning via convex-concave programming, *Comput. Intell.*, **36** (2020), 297–319. <https://doi.org/10.1111/coin.12267>
29. E. B. Tirkolae, A. Goli, S. Mirjalili, Circular economy application in designing sustainable medical waste management systems, *Environ. Sci. Pollut. Res. Int.*, **29** (2022), 79667–79668. <https://doi.org/10.1007/s11356-022-20740-x>
30. R. Lotfi, K. Kheiri, A. Sadeghi, E. Babae Tirkolae, An extended robust mathematical model to project the course of COVID-19 epidemic in Iran, *Ann. Oper. Res.*, **2022** (2022), 1–25. <https://doi.org/10.1007/s10479-021-04490-6>
31. M. Khalili, M. Karamouzian, N. Nasiri, S. Javadi, H. Sharifi, Epidemiological characteristics of COVID-19: A systematic review and meta-analysis, *Epidemiol. Infect.*, **148** (2020), e130. <https://doi.org/10.1017/S0950268820001430>
32. P. Wang, X. Chen, X. Zhang, Research on location of logistics distribution center based on k-means clustering algorithm, *Secur. Commun. Netw.*, **2022** (2022), 2546429. <https://doi.org/10.1155/2022/2546429>
33. R. Bobbili, V. Madhu, A machine learning model for failure of perforated plates under impact, *Mech. Based Des. Struct. Mach.*, **50** (2022), 2582–2590. <https://doi.org/10.1080/15397734.2020.1763184>
34. C. Li, Construction of the reverse resource recovery system of e-waste based on DLRNN, *Comput. Intell. Neurosci.*, **2021** (2021), 2143235. <https://doi.org/10.1155/2021/2143235>
35. M. K. Lim, Y. Li, X. Song, Exploring customer satisfaction in cold chain logistics using a text mining approach, *Ind. Manage. Data. Syst.*, **121** (2021), 2426–2449. <https://doi.org/10.1108/IMDS-05-2021-0283>
36. K. Cao, A machine learning-based approach to railway logistics transport path optimization, *Math. Probl. Eng.*, **2022** (2022), 1691215. <https://doi.org/10.1155/2022/1691215>
37. A. Mohammadi-Balani, M. D. Nayeri, A. Azar, M. Taghizadeh-Yazdi, Golden eagle optimizer: A nature-inspired metaheuristic algorithm, *Comput. Ind. Eng.*, **152** (2021), 107050. <https://doi.org/10.1016/j.cie.2020.107050>
38. L. Abualigah, M. Abd Elaziz, P. Sumari, Z. W. Geem, A. H. Gandomi, Reptile Search Algorithm (RSA): A nature-inspired meta-heuristic optimizer, *Expert Syst. Appl.*, **191** (2022), 116158. <https://doi.org/10.1016/j.eswa.2021.116158>

39. J. W. Wang, W. H. Ip, R. R. Muddada, J. L. Huang, W. J. Zhang, On Petri net implementation of proactive resilient holistic supply chain networks, *Int. J. Adv. Manuf. Technol.*, **69** (2013), 427–437. <https://doi.org/10.1007/s00170-013-5022-x>
40. J. Wang, R. Dou, R. R. Muddada, W. Zhang, Management of a holistic supply chain network for proactive resilience: Theory and case study, *Comput. Ind. Eng.*, **125** (2018), 668–677. <https://doi.org/10.1016/j.cie.2017.12.021>



AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)