A Novel Neural Network Model for Demand Prediction of Bike-Sharing

Fan Wu[1](✉), Si Hong[1], Wei Zhao[2], Xiao Zheng[2], Xun Shao[3], and Wen Qiu[3]

[1] School of Management Science and Engineering, Anhui University of Technology, Ma'anshan 243032, China
dragonwufan@126.com
[2] School of Computer Science and Technology, Anhui University of Technology, Ma'anshan 243032, China
[3] School of Regional Innovation and Social Design Engineering, Kitami Institute of Technology, Kitami 090-8507, Japan

Abstract. Accurate demand prediction of bike-sharing is a prerequisite to reduce the cost of scheduling and improve the users' satisfaction. However, it is very difficult to make the prediction absolutely accurate due to the stochasticity and nonlinearity in the bike-sharing system. In this paper, a model called pseudo-double hidden layer feedforward neural network is proposed to approximatively predict the practical demand of bike-sharing. In this neural network, an algorithm called improved particle swarm optimization in extreme learning machine is proposed to define its learning rule. On the basis of fully mining the massive operational data of "Shedd Aquarium" bike-sharing station in Chicago (USA), the demand of this station is predicted by the model proposed in this paper.

Keywords: Demand prediction · Bike-sharing · Pseudo-double hidden layer feedforward neural network · Extreme learning machine · Improved particle swarm optimization

1 Introduction

With the development of sharing economy, bike-sharing systems have rapidly emerged in major cities all over the world. Bike-sharing can be described as a short-term bicycle rental service for inner-city transportation providing bikes at unattended stations. It has become one of the most important low-carbon travel ways. Compared with traditional rental service, bike-sharing won't be limited by the boxes at bike-stations. It provides more convenient service, but generates more complicated problems. For instance, the layout of bike-sharing stations is more flexible and the capacities of stations will no longer be fixed, leading to big fluctuant demands for the stations. Some new characteristics are exhibited, such as the uneven distribution of users' demand in time and space. It makes the prediction of bike-sharing more complicated.

S. W. Loke et al. (Eds.): MONAMI 2020, LNICST 338, pp. 19–29, 2020.
https://doi.org/10.1007/978-3-030-64002-6_2

Accurate demand prediction of bike-sharing can effectively improve user experience and enhance brand's competence, which was elaborated in Xu et al. (2019). Several interesting factors affecting demand have been studied. We refer to El-Assi et al. (2017) and Ermagun et al. (2018) for a survey on the main problem and methods arising in bike-sharing systems. The key is to have an effective forecasting method. The existing demand prediction methods can be mainly divided into two types. The traditional one is based on statistical analysis. ZH et al. (2019) used a statistical physics method to define demand fluctuation and established different bike-sharing systems in different periods. Based on the ordinary least square, geographically weighted regression (GWR) and semi-parametric geographically weighted regression methods, Yang et al. (2019) proposed a methodology to estimate a shared-bike trip using location-based social network data and conducted a case study in Nanjing, China. Negahban (2019) proposed a novel methodology combining simulation, bootstrapping, and subset selection that harnessed useful partial information in every bike drop-off observation (even if it is subject to censoring). It estimated true demands in situations where data cleaning approaches commonly used in the bike-sharing literature failed due to lack of valid data. Cheng et al. (2019) devised a trip advisor that recommended bike check-in and check-out stations with joint consideration of service quality and bicycle utilization. Then, it predicted user demands of each station to obtain the success rate of rental and return in future. The other is based on artificial neural networks. Among these literatures, we mention Yang et al. (2018) and Yi et al. (2019) which proposed a deep learning approach using convolutional neural networks to predict the daily bicycle pickups at both city and station levels. Lin et al. (2018) proposed a novel graph convolutional neural network with data-driven graph filter (GCNN-DDGF) model. It can learn hidden heterogeneous pairwise correlations between stations to predict station-level hourly demands in a large-scale bike-sharing network. Xu et al. (2018) developed a dynamic demand forecasting model for bike-sharing by deep learning. The comparison results suggested that the LSTM NNs provide better prediction accuracy than both conventional statistical models and advanced machine learning methods for different time intervals. Chang et al. (2019) developed a novel prediction framework integrating AIS and artificial neural network forecasting techniques. The prediction performance is verified compared with other models. Feng et al. (2017) studied the Markov chain population model to predict bicycles demands among different travel stations and verified their effectiveness. Kim (2018) studied the influence of weather conditions and time characteristics on demands for bike-sharing. Furthermore, deep learning and its combination with a variety of new heuristic algorithms have been applied in various engineering practices (Benkedjouh et al. 2015; Wu et al. 2018; Cao et al. 2018; Hu et al. 2020), but rarely applied in demand prediction of bike-sharing.

In addition, those methods have some limitations as follows.

(1) In order to improve the accuracy of prediction by means of deep learning, most works mainly achieved their goals by adding the number of hidden layers. It means to increase the structure complexity of neural network, which would exponentially increase the running time of the method and amplify signal noise in data.
(2) Current researches on deep learning pay too much attention to the accuracy and convergence rate of learning targets by some methods, such as early stopping, adding noise to gradient (e.g. adjusting batch size and learning rate) or constantly adding

new regulators to train targets for improving the generalization of the samples. It is worth mentioning that accelerating gradient descent sometimes leads to worse generalization.

Considering these limitations, a novel neural network model for demand prediction of bike-sharing is proposed, which is called pseudo-double hidden layer feedforward neural network. In this model, an algorithm called improved particle swarm optimization in extreme learning machine is proposed to optimize initial weights and bias of the neural network and improve the prediction accuracy of this model.

2 Pseudo-double Hidden Layer Feedforward Neural Network

2.1 Network Structure

Generally speaking, pseudo-double hidden layer feedforward neural network (PDLFN) is a biologically inspired computational model, which consists of processing elements (called neurons) and connections between them with coefficients. The structure of PDLFN is different from the single hidden layer feedforward neural network (SLFN) and the double hidden layer feedforward neural network (DLFN). As shown in Fig. 1, it includes one input layer, "two" hidden layers and one output layer. The hidden layers consist of V layer and B layer. Compared with SLFN and DLFN, V layer is a special hidden layer. By means of V layer, PDLFN can directly process original data (e.g. multivariate time series) to produce the final results.

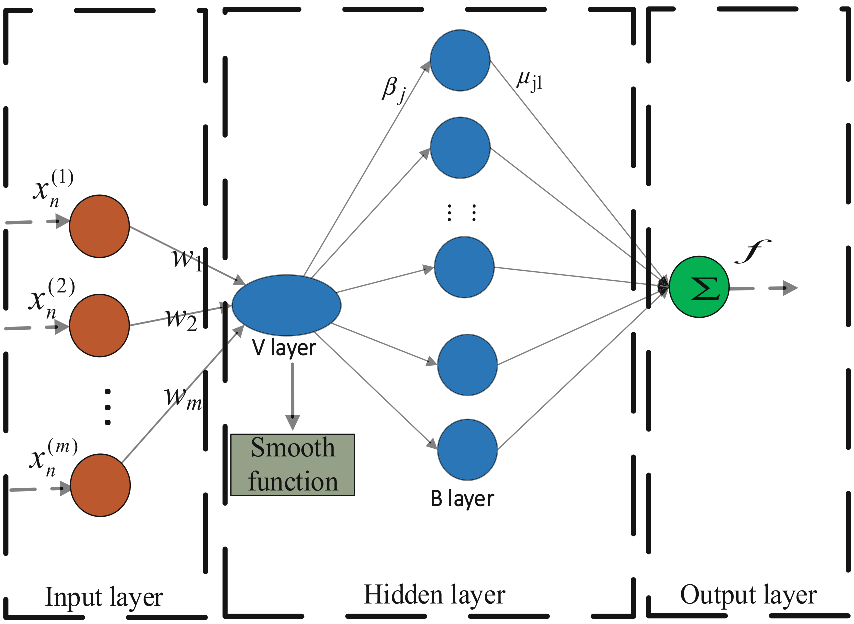

Fig. 1. The structure of PDLFN

Compared with DLFN, the special neuron of V layer in PDLFN is no longer with a bias value and an activation function in traditional sense, but with a smooth function. The smooth process referred to as V layer smooth is defined as follows. Assume that there are N samples. The n-th ($1 \le n \le N$) sample is denoted as $X_n = (x_n^{(1)}, x_n^{(2)}, \ldots, x_n^{(m)})^{\mathrm{T}}$, where $x_n^{(i)}$ denotes the i-th component of X_n, corresponding to the i-th neuron in the input layer ($1 \le i \le m$). The weight of the input layer to V layer is denoted as $W_i = (w_1, w_2, \ldots, w_m)^{\mathrm{T}}$. V layer smoothing is denoted as $f_V(X_n)$.

$$f_V(X_n) = X_n^{'} = \frac{\sum_{i=1}^{m} w_i \bullet x_n^{(i)}}{\sum_{i=1}^{m} w_i} \tag{1}$$

Except V layer, the other parts of PDLFN are similar to the SLFN. With the gradient descent method, the weights and biases are dynamically modified to achieve the expected learning effect.

The traditional BP learning algorithm has some limitations. For example, the network structure can't be determined easily, and the learning speed is too slow. To overcome these limitations, Huang (2015) proposed an improved learning algorithm called extreme learning machine (ELM). With ELM, we only need to determine the number of neurons in the hidden layers without considering the structure.

The outputs of PDLFN can be represented as follows.

$$f_{J\_out}(X_n) = \sum_{j=1}^{J} \mu_j \cdot T(f_V(X_n) \cdot \beta_j + B_j) \tag{2}$$

In this formula, β_j is the j-th ($1 \le j \le J$) input weight of neurons between V layer and B layer. B_j is the bias value of the j-th neuron in B layer. μ_j is the output weight of the j-th neuron in B layer. $T(x)$ is the excitation function, which can be set as "Sig", "Sin" or "Hardlim", etc. In this paper, the excitation function is uniformly set as "tansig". It means the excitation function is a hyperbolic tangent function in the "Sig" functions. The formula is shown in Eq. (3).

$$T(x) = \frac{2}{1 + e^{-x}} - 1 \qquad (-1 < T(x) < 1) \tag{3}$$

2.2 Learning Rule

Improved particle swarm optimization in ELM (IPSO-ELM) is adopted for learning mechanism of pseudo-double hidden layer feedforward neural networks. The influence of initial random bias and weight on prediction accuracy is reduced by improving particle swarm method to optimize the initial threshold and weight of extreme learning machine.

2.2.1 Improved Particle Swarm Optimization

Particle swarm optimization is a swarm intelligence algorithm that simulates the regularity of a bird population. It is based on the concepts of population and evolution, through the cooperation and competition among individuals. And the search for the optimal solution of complex space is realized. Suppose that in a dimensional search space *D*, the total number of particles is *N*, where the *i*-th particle is represented as a *D* dimensional vector $X_i = (x_{i1}, x_{i2}, \ldots, x_{iD}), i = 1, 2, \ldots, N$. The velocity of the *i*-th particle is also a *D*-dimension vector $V\mathrm{i} = (v_{i1}, v_{i2}, \ldots, v_{iD}),\ i = 1, 2, \ldots, N$. The individual extremum searched before the *i*-th particle is $Pbest = (p_{i1}, p_{i2}, \ldots, p_{iD}), i = 1, 2, \ldots, N$. And the global extreme value of particle swarm is $gbest = (g_1, g_2, \ldots, g_D)$.

In standard PSO, particles update their velocity and position according to the following formulas:

$$v_{ij}(t+1) = w * v_{ij}(t) + c_1 * r_1(t)[p_{ij}(t) - x_{ij}(t)] + c_2 * r_2(t)[p_{gj}(t) - x_{ij}(t)] \quad (4)$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \quad (5)$$

where c_1 and c_2 are learning factors, r_1 and r_2 uniform random numbers within the range of [0, 1]. The first part in formula (4) represents the velocity before the particle, which ensures the global convergence of the algorithm. The second and third parts make the algorithm have local convergence ability. It can be seen that inertia weight *w* represents inheritance degree of the original velocity. The global convergence ability increases with *w*. Therefore, in this paper, a compression factor combined with a dynamic inertia weight updating speed and weight is adopted to ensure that the algorithm has strong global search ability in an early stage. This can guarantee local fine search ability in the later stage, and achieve fast convergence. Its formula is shown as follows:

$$\begin{aligned} v_{ij}(\mathrm{t}+1) = &\left(w_{\max} - \frac{(w_{\max} - w_{\min}) * t}{T_{\max}}\right) * v_{ij}(t) \\ &+ c_1 r_1(t)[p_{ij}(t) - x_{ij}(t)] + c_2 r_2(t)[p_{gi}(t) - x_{ij}(t)], \\ &t < \frac{2T_{\max}}{3} \end{aligned} \quad (6\text{-}1)$$

$$
\begin{aligned}
v_{ij}(t) = \lambda * v_{ij}(t) + c_1 r_1(t)[p_{ij}(t) - x_{ij}(t)] \\
+ \ c_2 r_2(t)[p_{gi}(t) - x_{ij}(t)], \\
\frac{2T_{\max}}{3} \le t \le T_{\max}
\end{aligned}
\tag{6-2}
$$

where λ is the compression factor

$$
\lambda = \frac{2}{\left|2 - \beta - \sqrt{\beta(\beta - 4)}\right|} \ (\beta = c_1 + c_2)
$$

$T_{\max}$ denotes the maximum iteration number. t denotes the current iteration number. And $w_{\max}/w_{\min}$ denotes the maximum/minimum inertia weight.

2.2.2 IPSO-ELM Learning Process

The optimized flow chart of the improved particle swarm optimization algorithm for the extreme learning machine is shown in Fig. 2.

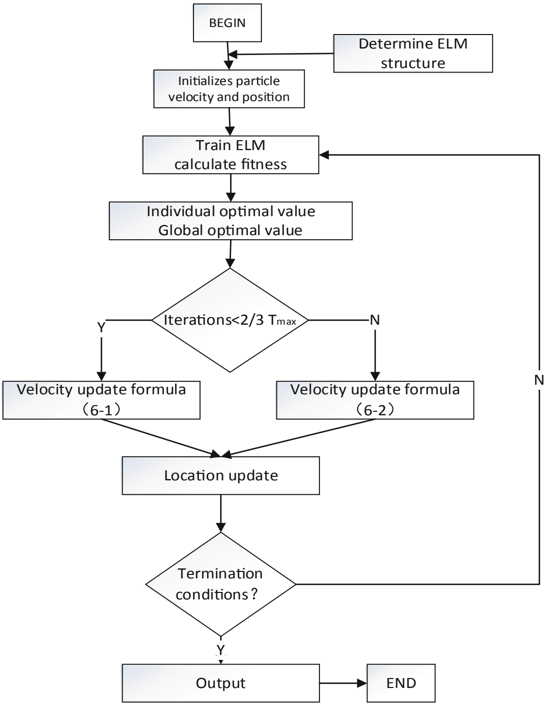

Fig. 2. IPSO-ELM optimization process

The pseudocode is as follows:

```
Step 1 : BEGIN
Step 2 : Initializing particle x= rand()、v =rand(); Tmax; times_iter;
Step 3 : Train ELM, fun_value = train_error; Record individual optimum p_best、global optimal g_best;
Step 4 : While fun_value> = 0.001
            if times_iter < 2/3*Tmax
                v = formula（5-1）;
            elseif  times_iter <= Tmax
                v= formula（5-2）;
                Update particle state;
                times_iter += 1;
                return to Step 3;
            else
                turn to Step 5;
Step 5 :Print g_best;fun_value;
Step 6 : END
```

3 Demand Prediction Model of Bike-Sharing

3.1 Prediction Period

In the bike-sharing system, its self-regulating ability has often met the demand for renting during peak time. In the scheduling problem, the user behavior during a peak period is one of main factors affecting the scheduling scheme. Thus, we discuss the demand prediction during a peak time as our scenario.

3.2 The Prediction Model

The demand prediction model of bike-sharing based on pseudo-double hidden layer feedforward neural network is given in Fig. 3.

3.3 Evaluation Criterion

In order to test the effectiveness of the pseudo-double hidden layer neural network combined with the improved PSO-ELM prediction model, the following error analysis method is selected in this paper, mean squared error (*MSE*). The formula is shown below.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Z_i - A_i)^2 \tag{7}$$

where A_i represents the final predicted value. Z_i is the measured value.

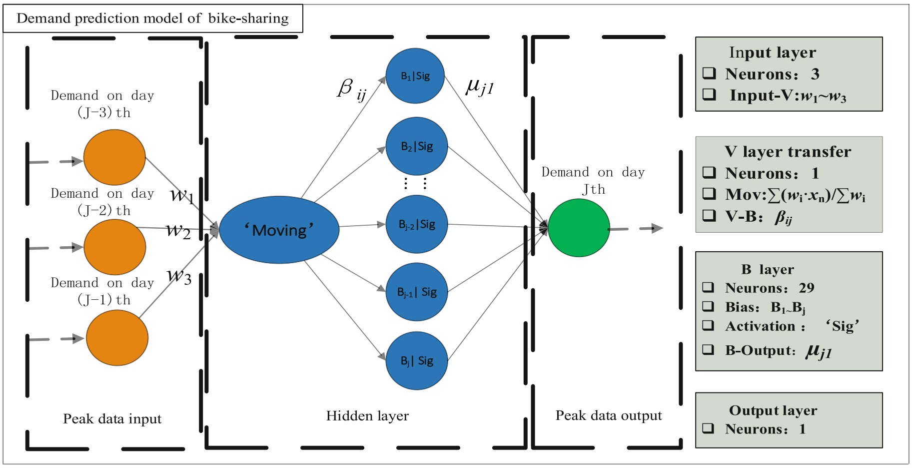

Fig. 3. Demand prediction model of bike-sharing based on PDLFN

4 Empirical Analysis

4.1 Data Collection

The data used in this section is from the official website of Divvy Bike in Chicago, USA. The original data amount is up to 2 million pieces (including 7 data features such as vehicle travel distance, user age, station number, etc.), we select Shedd Aquarium station (station_id = "3") as the study case column. In this paper, a total of 121 samples from April 1, 2018 to June 30, 2018 are selected as the training set. 30 samples from July 2018 are used as prediction sets.

In order to determine the distribution of the peak period in bike-stations, the distribution of bike rent in April is first drawn in Fig. 4.

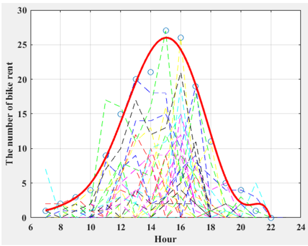

Fig. 4. Rent in April

We can note that the peak period of borrowing bikes is generally distributed from 13 pm to 17 pm, laying a foundation for this paper to select the peak time.

4.2 Prediction Result

In order to verify the performance of the improved particle swarm optimization algorithm, the fitness evolution curves before and after the improvement is shown in Fig. 5 and Fig. 6.

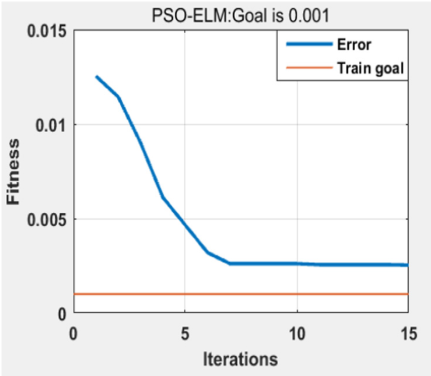

Fig. 5. Fitness evolution curve of PSO

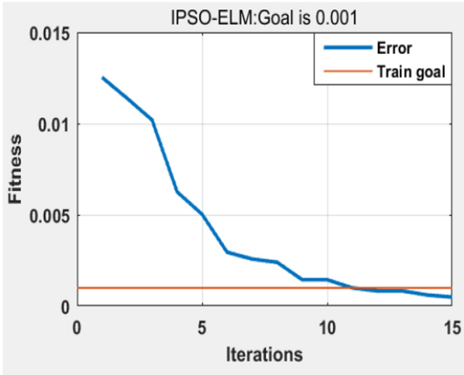

Fig. 6. Fitness evolution curve of IPSO

It can be seen that the global search ability of particle swarm is effectively improved by changing particle velocity through compression factor and dynamically adjusting inertia weight. Our particle swarm algorithm is more capable of jumping out of the local optimal solution and finding a better solution in the later stage.

Then, combined with the proposed PDLFN and the improved particle swarm algorithm, a prediction model based on PSO-ELM neural network was established. In this model, the activation function of neurons in B layer is ‘Sig’, and the TYPE value is 0 (representing fitting). Finally, 29 neurons in B layer are determined by the trial algorithm. The predicted results are shown in Fig. 7.

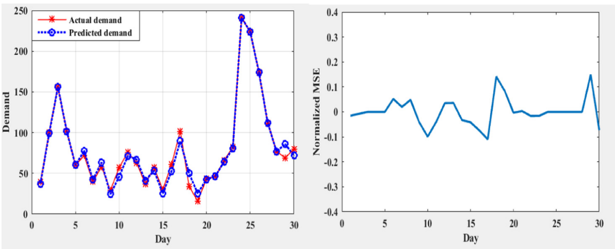

Fig. 7. Pseudo-double hidden layer-IPSO-ELM

In addition, in summarizing the experimental consideration on how to improve the prediction accuracy, it is found that the model proposed in this paper has the following two advantages:

(1) To solve the problem of demand prediction for bike-sharing, the method proposed in this paper to improve the prediction accuracy based on network structure is more effective than that to improve the prediction accuracy based on ELM improved by the optimization algorithm, and the implementation path is simpler and more efficient.
(2) In classification, fitting with time sequence features of large sample data, combined with neural network machine learning method is a frontier way, but by adding hidden layers of the network constantly to achieve the predetermined accuracy, overfitting phenomenon will inevitably occur, reducing the generalization performance of the model.

5 Conclusion

Aiming at the problem of demand prediction of bike-sharing, by referring to the learning method of multiple hidden layers in complex neural networks, the paper constructs a PDLFN model of "input layer - V layer - B layer - output layer", and improves the prediction accuracy of the model by optimizing ELM and improved particle swarm optimization algorithm. Finally, take Chicago, USA ("Shedd Aquarium" station) as an example for analysis, apply Newton interpolation method for processing of individual default values of data. It can be seen from the experimental results that the PDLFN model proposed has high prediction accuracy and good generalization ability. Generally speaking, compared with the existing methods to improve the prediction accuracy of the demand prediction model of bike-sharing, the method proposed in this paper is more concise and efficient, which has certain reference value in the research of demand prediction of bike-sharing.

References

Xu, H., Duan, F., Pu, P.: Dynamic bicycle scheduling problem based on short-term demand prediction. Appl. Intell. **49**(5), 1968–1981 (2019). https://doi.org/10.1007/s10489-018-1360-6
El-Assi, W., Salah Mahmoud, M., Nurul Habib, K.: Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto. Transportation **44**(3), 589–613 (2017). https://doi.org/10.1007/s11116-015-9669-z
Ermagun, A., Lindsey, G., Loh, T.H.: Bicycle, pedestrian, and mixed-mode trail traffic: a performance assessment of demand models. Landscape Urban Plann. **17**, 92–102 (2018)
Yang, F., Ding, F., Qu, X., Ran, B.: Estimating urban shared-bike trips with location-based social networking data. Sustainability **11**, 3220 (2019)
Negahban, A.: Simulation-based estimation of the real demand in bike-sharing systems in the presence of censoring. Eur. J. Oper. Res. **277**(1), 317–332 (2019)
Cheng, P., Hu, J., Yang, Z.D., Shu, Y.C., Chen, J.M.: Utilization-aware trip advisor in bike-sharing systems based on user behavior analysis. IEEE Trans. Knowl. Data Eng. **31**(9), 1822–1835 (2019)
Yang, H., Xie, K., Ozbay, K., Ma, Y., Wang, Z.Y.: Use of deep learing to predict daily usage of bike sharing. Transp. Res. Rec. **2672**(36), 92–102 (2018)
Ai, Y., Li, Z.P., et al.: A deep learning approach on short-term spatiotemporal distribution forecasting of dockless bike-sharing system. Neural Comput. Appl. **31**(5), 1665–1677 (2018). https://doi.org/10.1007/s00521-018-3470-9

Lin, L., He, Z.B., Peeta, S.: Predicting station-level hourly demand in a large-scale bike sharing network: a graph convolutional neural network approach. Transp. Res. Part C Emerg. Technol. **97**, 258–276 (2018)

Xu, C.C., Ji, J.Y., Liu, P.: The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets. Transp. Res. Part C Emerg. Technol. **95**, 47–60 (2018)

Chang, P.C., Wu, J.L., Xu, Y., et al.: Bike sharing demand prediction using artificial immune system and artificial neural network. Soft. Comput. **23**(2), 613–626 (2017). https://doi.org/10.1007/s00500-017-2909-8

Feng, C., Hillston, J., Reijsbergen, D.: Moment-based availability prediction for bike-sharing systems. Perform. Eval. **117**, 58–74 (2017)

Kim, K.: Investigation on the effects of weather and calendar events on bike sharing according to the trip patterns of bike rentals of station. J. Transport Geogr. **66**, 309–320 (2018)

Benkedjouh, T., Medjaher, K., Zerhouni, N., Rechak, S.: Health assessment and life prediction of cutting tools based on support vector regression. J. Intell. Manuf. **26**(2), 213–223 (2015). https://doi.org/10.1007/s10845-013-0774-6

Wu, Q., Ding, K., Huang, B.: Approach for fault prognosis using recurrent neural network. J. Intell. Manuf. **31**(7), 1621–1633 (2018). https://doi.org/10.1007/s10845-018-1428-5

Hu, H., Liu, Z., An, J.: Mining mobile intelligence for wireless systems: a deep neural network approach. IEEE Comput. Intell. Mag. **15**(1), 24–31 (2020)

Tamura, S., Tateishi, M.: Capabilities of a four-layered feedforward neural network: four layers versus three. IEEE Trans. Neural Netw. **8**(2), 251–255 (1997)

Cao, Y.-J., et al.: Recent advances of generative adversarial networks in computer vision. IEEE Access **7**, 14985–15006 (2018)

Jiawei, H., Kamber, M.: Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers Inc, San Francisco (2001)

Huang, G., Huang, G.B., Song, S.J., You, K.Y.: Trends in extreme learning machines: a review. Neural Netw. **61**, 32–48 (2015)