Machine Learning Based Key Performance Index Prediction Methods in Internet of Industry

Haowei Li[1], Liming Zheng[1], Yue Wu[2(✉)], and Gang Wang[1]

[1] Communication Research Center, Harbin Institute of Technology, Harbin 150080, China
lihwl0@qq.com, {zheng,gwang51}@hit.edu.cn
[2] Management School, Harbin Institute of Technology, Harbin 150080, China
wuy@hit.edu.cn

Abstract. The Internet of industry (IoI) has been well advanced in modem factory accompanying with increasing application of sensor network. The industry intelligent is just basing on techniques of IoI and industry data analysis and predication. In this paper, both the training process, in which the modeling between the input of environment and technological parameters and the output of key performance index (KPI) is built, and the functional process, where the model built in training process and current is used for KPI predication. Both multivariable linear regression and nonlinear BP neural network model are employed and verified with authentic data set.

Keywords: Internet of industry · Industry intelligent · Machine learning Neural network · Predication

1 Introduction

The Internet of industry driven by the wireless sensor network and industry data analysis are promoting the progress of the modern industry [1]. The modern industry with Internet of things (IOT) supported, which is the use of IOT, machines and other production facilities can be accessed to the internet by industrial enterprises to build a cyber-physical systems (CPS). Therefore, the real-time data information can be perceived, transmitted and processed intelligently, so as to achieve the real-time monitoring, early warning of the production process, optimal allocation of production resources and intelligent management.

The IOT has been greatly developed along with the integration of China's industrialization and information. IOT has various forms. According to the information acquisition, transmission, processing and dimension of utilization, the IOT can be divided into four layers: perceptual recognition layer, network transport layer, data intermediate layer and integrated application layer [2]. Among which, the data intermediate layer is responsible for data aggregation, management and intelligent analysis. The IOT has the characteristics of high real-time, large amount of data, heterogeneous

X. Gu et al. (Eds.): MLICOM 2017, Part II, LNICST 227, pp. 490–497, 2018.
https://doi.org/10.1007/978-3-319-73447-7_53

data sources etc. Internet data is one of the most important sources of big data and IOI data processing is the core of industrial intelligence systems.

In different industries, especially the Internet, the data is getting larger and larger. The traditional way is almost impossible to effectively find patterns to improve productivity today, only computers can be used to complete many complex missions. Machine learning can automatically learn programs from data. As a result, this rising discipline has become more and more important. It has been widely used in web search, spam filtering, recommendation systems, advertising, credit evaluation, fraud detection, stock trading and drug design and data mining etc. According to a recent report by the McKinsey Global Institute, machine learning (also known as data mining or predictive analysis) will drive the next round of innovation [3, 4]. This paper explores the application of machine learning methods in the prediction of IOI.

2 Predictive Modeling Method

2.1 Multivariable Linear Regression Model

Regression analysis was put forward by Galton in the late nineteenth century. The simplest method is linear regression, which is a regression analysis of the relationship between one or more independent variables and dependent variables, using least squares function of linear regression equation. Linear regression with only one independent variable is called simple linear regression. However, in practice a variable is usually affected by many factors, simple linear regression is not enough to achieve the purpose of solving problems. Multivariable linear regression modeling (MLRM) is a model with more than one independent variable. The general expression of MLRM is as follows:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon,\ \varepsilon \sim N(0, \sigma^2). \tag{1}$$

Where, $x_1, x_2, \ldots, x_p (p \geq 2)$ are the value of independent variables. Y is dependent variable, $\beta_1, \beta_2, \ldots, \beta_p$ are model coefficients and ε is random error.

If conduct n observations, the data collected by the n groups of samples is $x_{i1}, x_{i2}, \ldots x_{ip}, Y_i\ \ (i = 0, 1, \ldots, n, n > p)$ can be expressed in the form of matrix as follows:

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \ldots & x_{1p} \\ 1 & x_{21} & \ldots & x_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_p \end{bmatrix}. \tag{2}$$

Where, $Y = X\beta + \varepsilon$ is called linear regression data model, X is the design matrix and ε is the error vector which meets the random conditions.

The propose of linear regression is to find the mean square error of linear function on $x_1, x_2, \ldots, x_p$:

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p. \quad (3)$$

Where, $Q(\beta) = \|Y - X\beta\|^2$ is minimum unbiased estimator, that is, the regression equation:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_p x_p. \quad (4)$$

The least squares estimation of β is defined as $\hat{\beta} = (X^T X)^{-1} X^T Y$, through which the unbiased estimator of Y can be obtained.

2.2 Nonlinear BP Neural Network Model

In late 80s, the error back-propagation algorithm (BP algorithm) [6] and its application in artificial neural network learning process had greatly promoted the development of machine learning, and led the development of statistical machine learning model trend until now. Researchers found that with back-propagation algorithm, the artificial neural network model can automatically correct its parameters in the training process, making neural network model fit the training data to a greater degree. Using a large number of training samples, the neural network can be trained to learn statistical rules to predict unknown events. This kind of machine learning model based on statistical rules shows great superiority in many aspects, compared with the previous methods based on artificial rules. Although during this period, artificial neural network can also be referred to as multilayer perceptron, but in fact it is a shallow model, which contains only a layer of hidden nodes.

BP neural network has three layers: input layer, hidden layer and output layer. One topology of an ordinary BP neural network is shown in the following figure:

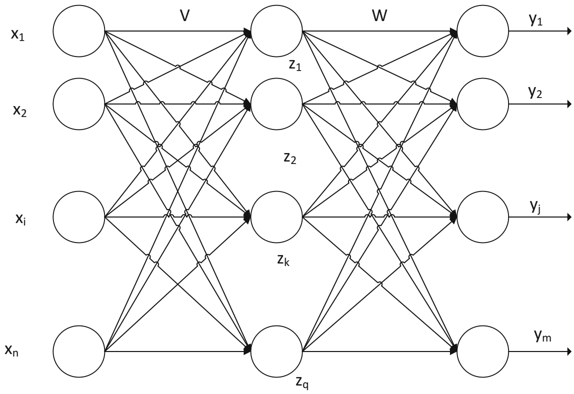

Fig. 1. Topological structure of BP neural network.

The learning process of a BP neural network can be represented by Fig. 1. The process consists of forward propagation of input data and back propagation of error.

The direction of forward propagation is from the input layer through the hidden layer to the output layer. Forward propagation processes data layer by layer, the state of each layer of neurons only affects the next layer of neurons. Then at the output layer the neurons compare the actual output with the desired output, if the result does not match, the deviation between the actual output and the desired output is calculated, which will be transmitted back. When the deviation is returned to the input layer, the weights and thresholds of each neuron in the hidden layer are modified so as to minimize the deviation.

v_{ki} is the weight between the input layer and the hidden layer, w_{jk} is the weight between the hidden layer and the output layer, f_1 is the transfer function of the hidden layer and f_2 is the transfer function of the output layer. Then the output of the hidden layer and layer node can be defined as:

$$z_k = f_1(\sum_{i=0}^{n} v_{ki}x_i),\ \mathrm{k} = 1, 2, \ldots, \mathrm{q}. \tag{5}$$

$$\mathrm{y}_j = f_2(\sum_{k=0}^{q} w_{jk}z_k),\ \mathrm{j} = 1, 2, \ldots, \mathrm{m}. \tag{6}$$

The number of learning samples $x_1, x_2, \ldots, x_p$ is p. Assuming the actual output of the l sample is y_j^l, expected output is t_j^l. Then the deviation between the actual output and the expected output of the l sample is:

$$E_l = \frac{1}{2}\sum_{j=1}^{m} (t_j^l - y_j^l)^2. \tag{7}$$

The global deviation produced by p samples is:

$$E = \sum_{l=1}^{p} E_l = \frac{1}{2}\sum_{l=1}^{p}\sum_{j=1}^{m} (t_j^l - y_j^l)^2. \tag{8}$$

The change of weight:

$$\Delta w_{jk} = -\eta\frac{\partial E}{\partial w_{jk}} = -\eta\frac{\partial}{\partial w_{jk}}(\sum_{l=1}^{p} E_l) = \sum_{l=1}^{p}(-\eta\frac{\partial E_l}{\partial w_{jk}}), \eta \in (0, 1). \tag{9}$$

Where, η is the learning rate, which needs to be chosen properly. The system will be unstable if η is too large. η being too small will lead to long training time and slow convergence. η is generally chosen between 0.01 and 0.8.

Deviation signal is defined as:

$$\delta_{lj} = -\frac{\partial E_l}{\partial S_j} = -\frac{\partial E_l}{\partial y_j}\frac{\partial y_l}{\partial S_j}. \tag{10}$$

Where, S_j is the net output of summation unit of output layer of neurons.

$$\frac{\partial E_l}{\partial y_j} = \frac{\partial}{\partial y_j}[\frac{1}{2}\sum_{j=1}^{m}(t_j^l - y_j^l)^2] = \sum_{j=1}^{m}(y_j^l - t_j^l). \tag{11}$$

$$\frac{\partial y_l}{\partial S_j} = f_2'(S_j). \tag{12}$$

From to Eqs. (10)–(12):

$$\delta_{lj} = \sum_{j=1}^{m}(t_j^l - y_j^l)\, f_2'(S_j). \tag{13}$$

Also:

$$\frac{\partial E}{\partial w_{jk}} = \frac{\partial E}{\partial S_j}\frac{\partial S_j}{\partial w_{jk}}. \tag{14}$$

Then:

$$\frac{\partial E}{\partial w_{jk}} = -\delta_{lj} z_k = -\sum_{j=1}^{m}(t_j^l - y_j^l)\, f_2'(S_j) z_k. \tag{15}$$

From Eq. (9) and Eq. (15), it can be seen that the weight of the output layer is changed:

$$\Delta w_{jk} = \sum_{l=1}^{p}(-\eta\frac{\partial E}{\partial w_{jk}}) = \eta\sum_{l=1}^{p}\sum_{j=1}^{m}(t_j^l - y_j^l)\, f_2'(S_j) z_k. \tag{16}$$

Similarly, the change of hidden layer weights:

$$\Delta v_{ki} = \eta\sum_{l=1}^{p}\sum_{j=1}^{m}(t_j^l - y_j^l)\, f_2'(S_j) w_{jk} f_1'(S_k) x_i. \tag{17}$$

The purpose of building the whole network is achieved by changing the weights of each neuron in the hidden layer.

3 Performance Simulation

In this paper, we used a set of 7 input 2 output food IOI data. Input is 7 performance indicators collected by the IOI sensor, which is: temperature, humidity, drying time, the original ratio, processing time of raw materials and other industrial and environmental parameters. The output value y_1 is the water content of the product, y_2 is nutritional content and data set size is 5000. Multiple linear regression model and BP neural network model were used to analyze the data. Fitting model is used to the prediction test of data and the mean square error (MSE) is used to evaluate the performance of the two different models. 3725 data is used for model learning and 1225 data is for verification.

Figure 2 is the result of the model training process using multiple linear regression method. The ordinate is the key index value and the abscissa is data record index for the validation of data set, where the blue line is the true value, the red line is the fitting output value. It can be seen that the fitting values of y1 and y2 converge to the true values except for very few peak values.

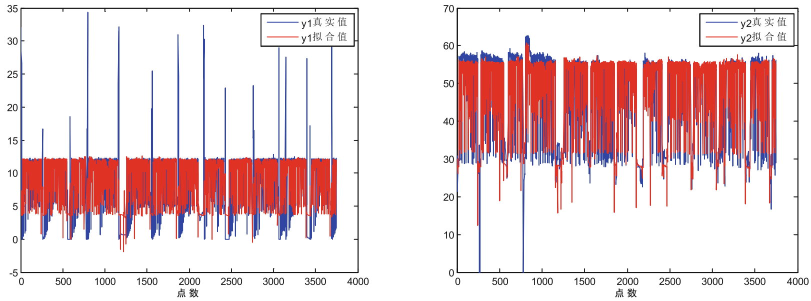

Fig. 2. Fitting value and real value of y1, y2 in Multivariable linear regression model (Color figure online)

Figure 3 shows the results of the validation using a multiple linear regression model in the workflow. The ordinate is the residuals of the predicted and true values of the key indicators and the abscissa is data record index for the validation of data set. It can be seen that the residual is basically 0, but the residual value is larger on more discrete points, which affects the overall prediction performance.

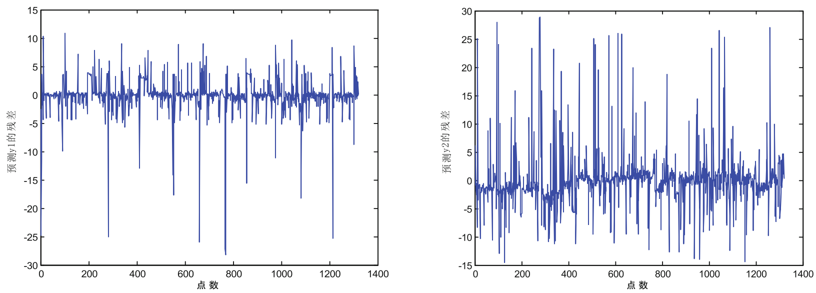

Fig. 3. Residual error between predicted and real values in multivariate linear regression model

Figure 4 shows the output result of the model training process using BP neural network. It can be seen that under the conditions of nonlinear model, the fitting values of y1 and y2 also converge to the true values in addition to very few peak values.

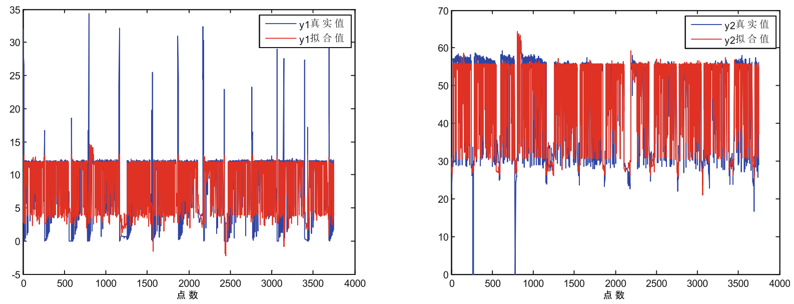

Fig. 4. Fitting value and real value of y1, y2 in BP neural network model

Figure 5 shows the results of the application of the model training process using multiple linear regression. It can be seen that the residual is basically 0. The peak value whose residual is larger is significantly reduced, compared with the linear multiple regression model. The prediction results are much better than the linear method.

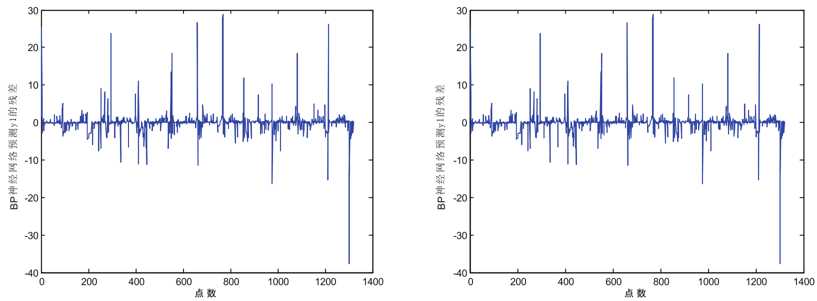

Fig. 5. Residual error between predicted and real values in BP neural network model

Table 1 shows the quantitative comparison of the mean square error between multiple linear regression model and BP neural network model during the training and working stage. It can be seen that the nonlinear BP neural network model can achieve better mean square error performance both in the training and working stage. The neural network model is improved by 43% and 37% respectively compared with linear model, especially the output value y2.

Table 1. The comparison of multiple linear regression model and BP neural network model.

MSE (y1/y2)	Multiple linear regression mode	BP neural network model
Training stage	10.0/20.5	7.5/11.1
Working stage	9.9/23.1	9.3/14.5

4 Conclusion

This paper discusses the application of machine learning methods in IOI. The multiple linear regression model and the nonlinear BP neural network model are introduced. The application of the two models above in the prediction of key performance indicators in IOI is given. At last, the model is validated by the real production environment data. The quantitative comparison of the mean square error of the linear method and the nonlinear method is given. Through the analysis and simulation, it is proved that the nonlinear BP neural network model can greatly improve the performance compared with the linear multiple regression.

Acknowledgement. This work was supported by the National Natural Science Foundation of China under Grant No. 61401120.

References

1. Wang, X.: Made in China 2015: from an industrial country to an industrial power. Internet Things Technol. **5**(5), 3–4 (2015)
2. Liu, Y.: Introduction to Internet of Things. Science Press, Beijing (2013)
3. Conway, D., White, J.: Machine Learning for Hackers. O'Reilly Media Inc., Newton (2012)
4. Domingos, P.: A few useful things to know about machine learning. Commun. ACM **55**(10), 78–87 (2012)
5. Bao, Y.: Data Analysis Tutorial. Tsinghua University Press, Beijing (2011)
6. Hecht-Nielsen, R.: Theory of the back propagation neural network. In: International Joint Conference on Neural Networks, IJCNN, pp. 593–605. IEEE (1989)