

Prediction of the Number of Online Doctors' Followers based on Machine Learning

Wei Wanhua

Department of Electronic Information
Engineering
Wuhan Donghu University
Wuhan, China
546386997@qq.com

Lv Renzhi

College of Science
China University of Petroleum
Beijing, China
865491236@qq.com

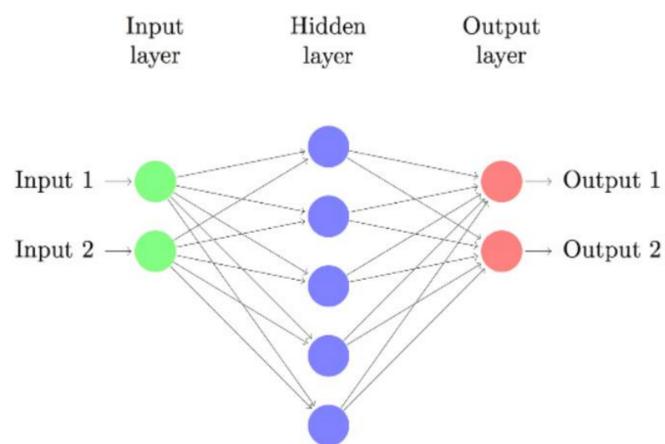
Abstract

The number of followers of online doctor can reflect the popularity and reputation of a doctor, and the number of a doctor's followers is affected by graphic consultation fee, video consultation fee, online consultation evaluation, hospital consultation evaluation, score, consultation volume and reservation volume. To explore the relationship between them, this paper establishes multiple linear regression model, Bayesian ridge regression model, elastic network regression model, support vector machine model, random forest, gradient boosting regression model and back propagation(BP) neural network model by python. The result shows the BP neural network model has the best performance and can be applied to the prediction of doctors' followers number.

MODELS PRINCIPLE

The model of the system magnetic coupler is shown in Fig 1, in which the transmitter in the magnetic coupler is a plane circular coil (namely, coil-P), and the receiver is composed of three coils. Coil-1 and coil-2 are crossed dipole coils wound on the cross core, and coil-3 is a circular coil. In Fig 1, D_1 , D_2 , and D_3 are the outer diameter of coil-P, the diameter of coil-2, and coil-3, respectively. L_1 is the width of coil-1 and coil-2, and L_2 is the length of coil-1 and coil-3. N_1 , N_2 , N_3 and N_4 are the turns of coil-P, coil-1, coil-2 and coil-3 respectively. In the MC-WPT system proposed in this paper, when the receiver rotates arbitrarily, at least one receiving coil will be coupled with the transmitting coil, so that the receiving coil can pick up power under any angle offset and improve the freedom of wireless power transfer. When a dependent variable is affected by multiple independent variables, and there is a linear relationship between them, the regression analysis we do is multiple linear regression. The model form is:

$$\hat{Y}_i = b_0 + b_1x_{1i} + \dots + b_kx_{ki}$$



MODEL ESTABLISHMENT

This paper grabs online doctor diagnosis data from a website. The original data contains 575 pieces of data. The independent variables are: graphic consultation fee, video consultation fee, the number of online consultation evaluation, the number of hospital consultation evaluation, score, the number of consultations, the number of appointments and those variables are represented by respectively. The dependent variable is the number of followers which is represented by .

Considering that the amount of data is not very large, the data is divided into a training dataset and test dataset in a ratio of 95:5. The amount of training dataset is 546 and the amount of training dataset is 29. Multivariate linear regression model, Bayesian Ridge regression model, Elastic network regression model, support vector machine model, gradient boosting regression model, random forest model and BP network are established respectively on the training dataset by python. The partial code is as follows:

```
br_model = BayesianRidge()
lr_model = LinearRegression()
etc_model = ElasticNet()
svr_model = SVR()
gbr_model = GradientBoostingRegressor()
forest=RandomForestRegressor(n_estimator=80,criterion='mse',random_state=42)
model_names=['BayesianRidge','LinearRegression', 'ElasticNet', 'SVR', 'GBR','RF']
model_dic = [br_model, lr_model, etc_model, svr_model, gbr_model,forest]
pre_y_list = []
for model in model_dic:
    pre_y_list.append(model.fit(X_train, y_train).predict(X_test))
```

The 6 models' fitting effects on the test dataset are shown as figure 2-figure 7.

TABLE I PERFORMANCE COMPARISON TABLE OF 7 MODELS

model	EV	MAE	MAPE	MSE	R ²
Linear Regression	0.86	165.66	1.548	53653.90	0.86
Bayesian Ridge	0.88	149.23	2.06	52341.18	0.87
Elastic Net	0.89	143.15	1.804	48019.52	0.88
SVR	0.06	379.47	3.393	382577.59	0.03
GBR	0.93	119.44	0.784	29700.56	0.93
RF	0.91	123.71	0.804	34986.09	0.91
BP Network	0.97	81.162	0.604	16164.48	0.96

Conclusion

In summary, the model that should be selected for the prediction of online doctors concerned is the BP neural network model.

The reason why the performance of BP neural network is so well is that it has nonlinear mapping capabilities, enabling a mapping function from input to output. It proves that the three-layer neural network is capable of approximating any nonlinear continuous functions, which makes it special suitable for solving complex internal mechanisms. However, the selection of the BP neural network hidden layer lacks theory guidance, and the network is very sensitive to the initial network weight. So the initialization of the network with different weights will lead to converges to different local minimal, which is also the reason why each training gets different results.