

## Research on Freshman Registration Prediction based on GRNN and PNN

Lei Yang<sup>a</sup>, Longqing Zhang<sup>b</sup>, Liang Yu<sup>c</sup>, Yungui Chen<sup>d</sup> and Jun Nie<sup>e</sup>

Guangdong University of Science and Technology, Dongguan, China

<sup>a</sup>gdstyl@qq.com, <sup>b</sup>bjzqlq@qq.com, <sup>c</sup>840493518@qq.com, <sup>d</sup>370064596@qq.com,

<sup>e</sup>13739149@qq.com

### Abstract

The prediction of Freshmen's registration in Colleges and universities is a very difficult problem. Machine learning, as a discipline derived from artificial intelligence and statistics, makes it possible to solve this problem. Based on the enrollment and registration data of a university in previous years, this paper analyzes it by using generalized regression neural network (GRNN) and probabilistic neural network (PNN). The results show that whether freshmen registration or not can be predicted.

### Keywords

Freshmen; Forecast; GRNN; PNN.

### 1. Introduction

All colleges and universities will encounter such a situation: it is impossible to accurately predict the number of new students. Before the arrival of freshmen, colleges and universities need to do a lot of preparatory work, such as how many dormitories to prepare, how many new teachers to recruit, and whether the school canteen needs to be expanded. Once the preparation is insufficient, the school management will fall into a very passive situation after the beginning of school, and may even affect the reputation of the school.

The usual practice of colleges and universities is to make a rough estimation based on the registration rate in previous years, which is a very vague practice. We try to use machine learning to solve this problem. Among many machine learning algorithms, we choose generalized regression neural network (GRNN) and probabilistic neural network (PNN) to machine learn the data set.

We divide the data set into two parts: training set and test set. Firstly, we generate the neural network models of GRNN and PNN according to the training set, and then use these two models to predict the test set. The results are exciting. The accuracy of GRNN can reach more than 80% and that of PNN can reach more than 65%, which shows that our work is effective.

### 2. The Dataset

The admission data comes from the candidate information database of the admissions office. Basic information includes student's number, name, gender, ID card number, professional code, etc. The check-in data comes from the school's educational administration system. Basic information includes student's name, gender, ID number, date of birth, profession, etc.

The whole data set contains 200 pieces of data. We cleaned and transformed the data. Finally, the data set contains four features. We randomly selected 160 samples from the data set as the training set and the remaining data as the test set, that is, the total number of samples in the test set is 40.

### 3. Algorithms and Models

Generalized regression neural network is a radial basis function network based on mathematical statistics. Its theoretical basis is nonlinear regression analysis. GRNN has strong nonlinear mapping ability and learning speed. Finally, the network generally converges to the optimal regression with more sample size. When there are few sample data, the prediction effect is very good, and the network can also deal with unstable data. GRNN consists of four layers in structure, namely input layer, mode layer, summation layer and output layer. The number of neurons in the input layer is equal to the dimension of the input vector in the learning sample. Each neuron is a simple distribution unit, which directly transmits the input variables to the mode layer.

Probabilistic neural network PNN uses Gaussian function as basis function to form the estimation method of joint probability density distribution and Bayesian optimization rules, and constructs a neural network for probability density classification estimation and parallel processing. Similar to GRNN, PNN is composed of input layer, hidden layer and output layer. Different from GRNN, the output layer of PNN uses competitive output instead of linear output. Each neuron only sums and estimates various probabilities according to Parzen method, so as to compete for the response opportunity of input mode. Finally, only one neuron wins the competition, In this way, the winning neuron represents the classification of input patterns. PNN not only has the characteristics of general neural network, but also has good generalization ability and fast learning ability.

### 4. Experiment and Analysis

In this section, we use MATLAB tools to learn the data. The training set is trained with GRNN and PNN algorithms respectively to generate the corresponding neural network model. The two models are used to predict the data of the test set. Finally, the performance of the two algorithms is compared and analyzed.

#### 4.1. GRNN Algorithm Model

The data of the training set is predicted by GRNN, and its model diagram is shown in Figure 1:

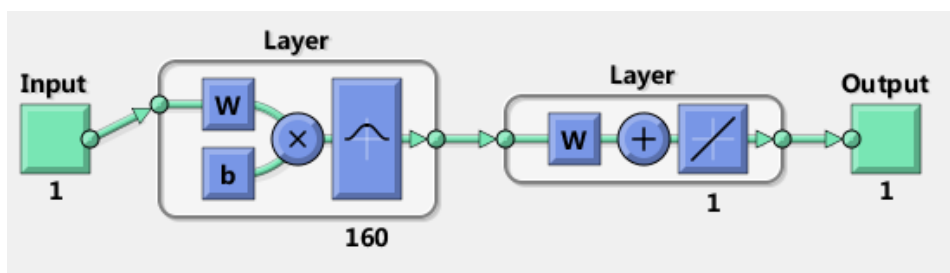


Figure 1. GRNN model

The input layer is a vector, the dimension is 4, the number of samples is 160, the linear function is a transfer function, the hidden layer is fully connected with the input layer, there is no connection in the layer, the number of neurons in the hidden layer is equal to the number of samples, that is, 160, the transfer function is a radial basis function, and there are two nodes in the summation layer. The first node is the output sum of each hidden layer node, The second node is the weighted sum of the expected result and each hidden layer node, and the output layer output is the second node divided by the first node.

### 4.2. PNN Algorithm Model

The PNN is used to predict the data of the training set, and its model diagram is shown in Figure 2:

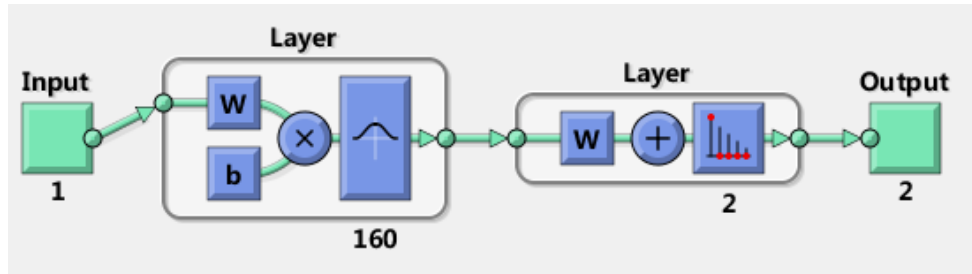


Figure 2. PNN model

The input layer receives the value of the sample, and the number of neurons is equal to the length of the input vector. The hidden layer is a radial base layer, and each neuron corresponds to a center (corresponding to a sample data). The input data is divided into two categories, because PNN is used for classification, that is, first train the network with samples, then input data, and use this network to identify which type of data it belongs to. The number of neurons in the summation layer is equal to the number of data classification. The summation layer obtains the average value of each type of data, and then compares the average value of each type to classify the data into the category with the largest value.

### 4.3. Algorithm Comparison

After the test set is predicted according to the two algorithm models, the obtained data are shown in Figure 3:

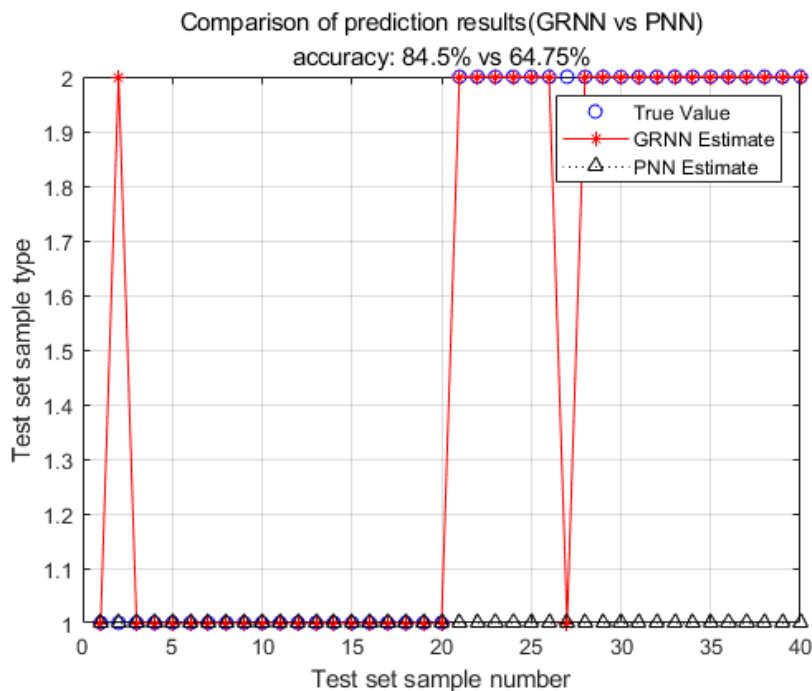
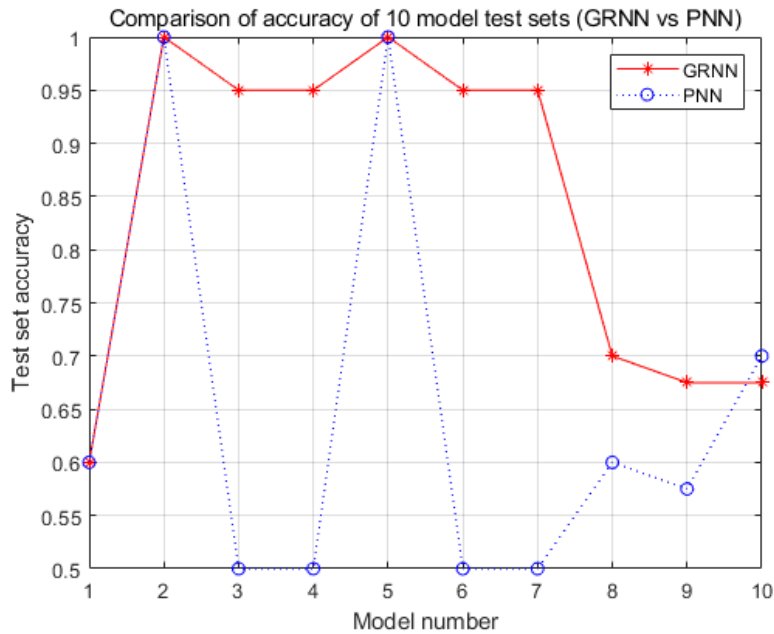


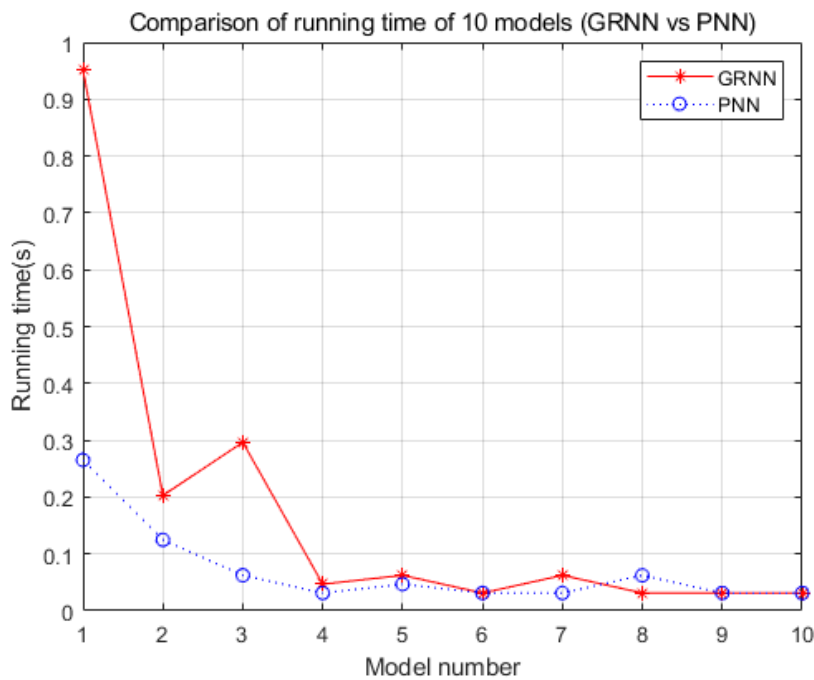
Figure 3. Comparison of prediction results

As can be seen from the above figure, the accuracy of GRNN can reach 84.5%, and the accuracy of PNN can also reach 64.75%, indicating that both algorithms can predict the enrollment data of Freshmen in Colleges and universities.



**Figure 4.** Comparison of accuracy of 10 model test sets

As can be seen from the above figure, GRNN performs well in models 2, 3, 4, 5, 6 and 7, while PNN performs well only in models 2 and 5.



**Figure 5.** Comparison of running time of 10 models

As can be seen from the above figure, PNN is slightly better than GRNN in running time.

## 5. Conclusion

This paper uses machine learning to predict the enrollment of freshmen. GRNN and PNN algorithms in machine learning are used. After the model training of the training set, the prediction accuracy of the test set is 84.5% and 64.75% respectively. There is still room for further improvement in the prediction accuracy, which we will further study in the future work.

## Acknowledgments

This paper was financially supported by general projects in Natural Science held by Guangdong University of Science and Technology NO.GKY-2019KYYB-31, 36, 42 and NO.GKY-2020KYYBK-22, 24, 27, General projects of social science and technology development in Dongguan in 2020NO.2020507154643, project of Guangdong Provincial Department of Education: 2020KTSCX166, project of Guangdong Institute of science and technology: GKY-2020KYZDK-10.

## References

- [1] Bing Wang, Chen Hu. Simulation study of device physics and design of GeOI TFET with PNN structure and buried layer for high performance[J]. Chinese Physics B, 2020, 29(10): 547-552.
- [2] Ting Lan, Chudong Tong, Haizhen Yu, Xuhua Shi, Lijia Luo. Nonlinear process monitoring based on decentralized generalized regression neural networks[J]. Expert Systems With Applications, 2020, 150.
- [3] Jun Guo, Chen-xu Gu, Jun-jie Yang, Yin Zhang, Heng Yang. Data mining and application of ship impact spectrum acceleration based on PNN neural network[J]. Ocean Engineering, 2020, 203.
- [4] Rui Hou, Yang Hu, Yunhao Zhao, Huan Liu. Hyperspectral image quality evaluation using generalized regression neural network[J]. Signal Processing: Image Communication, 2020, 83.
- [5] Yuhao He, Bin Yang, Hui Lin, Junqiang Zhang. Modeling Polarized Reflectance of Natural Land Surfaces Using Generalized Regression Neural Networks[J]. Remote Sensing, 2020, 12(2).
- [6] Yan, Mou, Xu, Du, Wang, Liu. A comparative assessment of predicting CH<sub>4</sub> adsorption on different activated carbons using generalized regression neural network (GRNN), and adaptive network-based fuzzy inference system (ANFIS)[J]. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 2019, 41(16).
- [7] Yunyun Xie, Chaojie Li, Youjie Lv, Chen Yu. Predicting lightning outages of transmission lines using generalized regression neural network[J]. Applied Soft Computing Journal, 2019, 78.
- [8] Guo Jun, Gu Chen xu, Yang Jun jie, Zhang Yin, Yang Heng. Data mining and application of ship impact spectrum acceleration based on PNN neural network[J]. Ocean Engineering, 2019, 203.
- [9] Xingliu Zhang, Zili Wang, Chen Lu. Fault diagnosis of gearbox based on the nonlinear output frequency response functions and PNN[J]. Vibroengineering PROCEDIA, 2018, 19.
- [10] Silicon Research; Findings from Department of Mechanical Engineering Has Provided New Data on Silicon Research [Predictive Analysis on Responses in WEDM of Titanium Grade 6 Using General Regression Neural Network (GRNN) and Multiple Regression Analysis (MRA)][J]. Computers Networks & Communications, 2018.
- [11] Energy; Research Conducted at Payame Noor University Has Updated Our Knowledge about Energy [A Comparative Assessment of Predicting Daily Solar Radiation Using Bat Neural Network (BNN), Generalized Regression Neural Network (GRNN), and Neuro-Fuzzy (NF)][J]. Computers Networks & Communications, 2018.