Stellar Spectral Classification with 2D Spectrum and Fully Connected Neural Network

Yakun Lu, Bo Qiu*, Guanjie Xiang, Mengci Li and Zhendong He
Hebei University of Technology, Tianjin 300401

*Corresponding author email: qiubo@hebut.edu.cn

Abstract. It is one of the basic tasks to realize the automatic classification of celestial spectrum. This paper presents a method of star classification based on two-dimensional spectral data. The data used in the experiment are two-dimensional spectral data of LAMOST DR6 from the National Astronomical Observatory. There are about 2000 two-dimensional spectral data in total to use to classify F, G and K stars spectra. By observing a large number of two-dimensional data of stars, different region data are selected as input data according to the stars. Then the data are normalized and put into the full connection neural network for training. The whole neural network consists of one input layer, nine hidden layers and one output layer. Each hidden layer is activated by rule function, and the output layer is activated by softmax function. Experiments show that the accuracy of classification is 80% by classifying the two-dimensional spectra of F, G and G, K. It shows that the two-dimensional characteristics of the spectrum can effectively classify the spectrum.

Keywords: Two-dimensional spectral; Spectrum; LAMOST; Fully connected neural network.

1. Introduction
The development of modern science and technology has greatly improved the astronomical observation ability, and the astronomical observation data has increased rapidly. With the development of data processing technology such as machine learning and data mining, people can extract more information from the observed spectral data and find more celestial phenomena. LAMOST is a powerful and competitive fiber spectroscopic telescope developed by China. LAMOST can get tens of thousands of two-dimensional Celestial Spectra every observation night, which can effectively help us to mine the information of celestial bodies. How to deal with these massive data accurately and quickly is a very important work of current astronomical work. Spectral classification of celestial bodies is one of the basic tasks of astronomy. To distinguish different celestial bodies from massive spectral data accurately and quickly can not only help to clean the early data, but also help the follow-up astronomers to study different stars.

In recent years, with the development of machine learning theory, researchers at home and abroad have proposed a lot of celestial spectral classification algorithms based on machine learning. In reference 1, a three-layer feedforward artificial neural network (ANN) model is proposed to classify 26 kinds of stars[1]. However, Ann has few hidden layers, simple structure, and can not effectively extract spectral features. In reference 2, wavelet transform is used to deal the data, and then support vector machine (SVM) is used to classify the spectral data of stars, and a high accuracy is obtained[2]. In reference 3, random forest (RF) is proposed to classify QSO, star and galaxy, and the accuracy of stars and Galaxy classification is 99%, which proves the feasibility of RF in celestial spectral classification[3]. In reference 4, Tingting Xu proposed spectral classification based on deep learning, and applied deep learning technology to the field of spectral classification[4]. In reference 5, a star spectrum classifier based on
convolution neural network is proposed, which can effectively extract star spectrum features\cite{5}. Traditional classification methods are based on one-dimensional spectral data, but the initial data observed by LAMOST, SDSS and other optical telescopes are two-dimensional spectral data. Then, one-dimensional spectrum is extracted from two-dimensional spectral data, and the two-dimensional celestial spectrum is transformed into one-dimensional celestial spectrum. Therefore, the above classification algorithm has high requirements for data, and the spectrum extraction technology directly affects the quality of one-dimensional spectrum classification. If the high-quality one-dimensional spectrum is not obtained, the results of one-dimensional spectrum classification will be greatly reduced. Although the two-dimensional spectrum extraction method is more and more perfect, but the process is more complicated, and from the data point of view, each processing operation will introduce data errors, destroy the integrity of the data, and the data will be lost. In the two-dimensional space, the spectral data will have more characteristics. The conversion of the two-dimensional data to the one-dimensional data ignores the two-dimensional characteristics of the target spectrum, such as space, direction and so on. Therefore, one-dimensional spectral data can not fully reflect the original characteristics of the object, and there is a large error in the classification algorithm based on one-dimensional spectrum.

In order to overcome the above shortcomings, this paper studies from another direction, skipping the process of two-dimensional to one-dimensional, and directly studies the initial data to achieve the classification of two-dimensional spectral data. In this paper, based on the two-dimensional data of LAMOST, the deep neural network is used to train and predict the two-dimensional spectral data, and a two-dimensional spectral classification processing method based on the deep neural network is proposed. In this paper, the deep neural network is an eleven layer all connected neural network, and 2000 two-dimensional observation data of LAMOST DR6 are selected. In order to verify the effectiveness of the two-dimensional spectrum classification method, FG and GK stars are classified. This paper explains the significance of two-dimensional spectral data for spectral classification from a new point of view, and proves the practical value of the data.

2. Fully Connected Neural Network Model

Fully connected neural network is an important algorithm in deep learning. Compared with other deep learning algorithms, it has the advantages of simple structure. The training complexity of the network model is related to the input data dimension and the depth of the network. The fully connected neural network mainly includes input layer, hidden layer and output layer, in which the hidden layer can contain multiple layers. The structure of the fully connected neural network is shown in Figure 1.

![Figure 1. Fully connected neural network model.](image)

In this paper, we use the full connection neural network model, which includes one input layer, nine hidden layers and one output layer. There are eleven layers of neural network. Each hidden layer uses Relu as the activation function, and the output layer uses softmax function to activate. The function
representation of the Relu activation function is
\[ f(x_i^j) = \max(0, x_i^j) \]

The mathematical expression of softmax function is
\[ \sigma(z)_j = e^{z_j} \left( \sum_{k=1}^{K} e^{z_k} \right)^{-1} \]

In order to prevent over fitting of the model, dropout is added between each hidden layer. Dropout can effectively prevent the over fitting problem, it will randomly lose some neuron nodes in the training process of the model, so that it does not play a role.

3. Data
The experimental data used in this paper are two-dimensional spectral data observed by LAMOST DR6. LAMOST has 4000 optical fibers, and 32 CCD cameras record red end data and blue end data respectively. The pixel size of LAMOST image is 4136 * 4160. By cutting the whole image, 250 images of 15 * 4136 pixels can be obtained. Each spectral image contains the spectral information of a star. Through the observation of two-dimensional spectrogram, we select different regions for different stars to extract features, reducing data redundancy. The two-dimensional spectra of different stars are shown in Figure 2.

For F, G stars spectra, in the range of 5-11 rows and 500-800 columns of pixels, we find that the gray image difference is quite obvious. Most F stars spectra will have two columns of pixels with smaller gray values in this range, while most G stars spectra are in this range, and this feature is not obvious. So we choose this range to classify F, G stars spectra.

For G, K stars spectra, in the range of 5-11 rows and 2700-3100 columns of pixels, we propose that most...
of the G stars spectra will have two columns of pixels with smaller gray values in this range, while most of the K stars spectra will have three columns of pixels with smaller gray values in this range, so we choose to classify G, K stars spectra in this range.

4. Experimental Results and Analysis

The experiment uses the keras framework and the software programming environment is python3.7. The data of LAMOST DR6 is used, the data of training set and test set are shown in Table 1.

In order to speed up the convergence of data, it is necessary to normalize the data. The normalization formula is

\[ x^* = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \]

Table 1. Spectral data.

| Spectral type | data | training set | Test set |
|---------------|------|--------------|---------|
| F, G          | F    | 647          | 527     | 120     |
|               | G    | 653          | 500     | 153     |
| G, K          | G    | 1510         | 1250    | 260     |
|               | K    | 1510         | 1250    | 260     |

The normalized data is put into the classification model of full connection neural network and iterated 600 times. Table 2 shows the results of F, G and G, K stars spectra classification. P is the accuracy, representing the ratio of the correctly classified spectrum to the spectrum classified in the model classification. R is the recall rate, which represents the ratio of the correctly classified spectrum to the spectrum in the model. A is the accuracy, representing the ratio of the correct classification spectrum to the number of test spectra.

Table 2. Classification experiment and result analysis.

| Spectral type | correctly classified | P/% | R/% | F-score/% | A/% |
|---------------|---------------------|-----|-----|-----------|-----|
| F             | 105                 | 88  | 61  | 72        |     |
| G             | 125                 | 76  | 93  | 83        |     |
| F+G           |                     |     |     |           | 80  |
| G             | 192                 | 80  | 78  | 79        |     |
| K             | 190                 | 79  | 79  | 80        |     |
| G+K           |                     |     |     |           | 80  |

5. Conclusion

The experimental results show that it is meaningful to realize spectral classification based on LAMOST two-dimensional spectral data. In this paper, it is proved that the spectral classification of celestial bodies can be realized without data processing from two-dimensional spectrum to one-dimensional spectrum. Not only the result of classification is good, but also the speed of classification is improved. From the perspective of two-dimensional data to study the spectrum, the proposed method is a new attempt in the field of astronomy to process the spectrum, which has a certain reference significance for the study of two-dimensional spectral data and practical value for the processing of massive spectral data. The experimental results of this paper can be further improved. In the future, more two-dimensional spectral data can be added to improve the performance of the system. The preprocessing operation can also be used to improve the quality of the two-dimensional spectral image.

Acknowledgments

This work was supported by the National Natural Science Foundation of China the Joint Research Fund
References
[1] Schierscher, F. & Paunzen, E. (2011). An artificial neural network approach to classify sdss stellar spectra. Astronomische Nachrichten, 332(6), p.597-601.
[2] Zhong-bao, L. Stellar Spectral Classification with Locality Preserving Projections and Support Vector Machine. J Astrophys Astron 37, 12 (2016).
[3] Bai, Y., Liu, J. F., Wang, S., & Yang, F.. (2018). Machine learning applied to star-galaxy-qso classification and stellar effective temperature regression. Astronomical Journal, 157(1).
[4] Xu Tingting, Ma chenye, Zhang Jingmin, et al. Study on LAMOST spectral classification based on deep learning [J]. Acta astronoma Sinica, 60 (02): 15-24
[5] Shi Chaojun, Qiu Bo, Zhou Yatong, et al. Automatic classification of stellar spectra based on convolutional neural network. Spectroscopy and spectral analysis, 39 (04), 322-326
[6] Xiao-Qing Wen, Jin-Meng Yang. Classification of star/galaxy/QSO and star spectral types from LAMOST Data Release 5 with machine learning approaches. 2020.
[7] Li Chao, Zhang Wen-hui, Lin Ji-ming. Study of Star/Galaxy Classification Based on the XGBoost Algorithm. 2019, 43(4):539-548.
[8] Xiang-Ru Li,Yang-Tao Lin,Kai-Bin Qiu.Stellar spectral classification and feature evaluation based on a random forest[J].Research in Astronomy and Astrophysics,2019,19(08):56-62.
[9] WANG Jing, YI Zhen-ping, YUE Li-li, et al. Spectral Classification of M-Type Stars Based on Ensemble Tree Models. 2018, 39(07):2288-2292