www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume – 4 Issue - 12 December, 2015 Page No. 15226-15228

Model For Spectral Classification

S.R.Gedam[1], Dr R.A. Ingolikar[2]

[1]Inter Institutional Computer Centre, RTM Nagpur University,
Nagpur(MS), India
Email: shmanke@rediffmail.com
[2]Department of Computer of Science, S.F.S College, Seminary Hills,
Nagpur(MS), India
Email: r_ingolikar@rediffmail.com

Abstract: Classification of astronomical data has always been a challenging problem faced by astronomers. Classification if performed by ensemble method gives more accurate results than single classifier. In the present study, model for spectral classification is developed using Salford predictive modeler in context of astronomical objects catalogue SDSS. Random forest ensemble method is used for model development and classification results are analyzed. Experimental results show that the model is beneficial for spectral classification.
Keywords: Data mining, Ensembles, random forest.

1. Introduction

Classification is the categorization of data for its efficient use. Classifying the available spectra of star is an important work, because by indentifying the star type one can find the physical properties of star. Over years various classification techniques have been explored for spectra data classification. These techniques include artificial neural network, support vector machine, nearest neighbor method and principle component analysis. LaSala et al[1] used the nearest neighbor method and got the average error of 1.14 spectral subtypes. Gulati et al[2] and others applied multi-layer feedback neural network 10 158 stellar spectra, the classification accuracy is 2 spectral subtypes. Qin [3] applied nearest neighbor method and principle component analysis and obtained spectral classification rate of 91%. As above methods can accurately separate the spectral type but still needs further improvement.

Recently more attention has been paid to combination of classifiers, which gives better results. One of the effective combination classifier is random forest developed by Breiman[4] and composed of many classification trees .In this paper we apply random forest ensemble method to spectral classification. Our aim is to evaluate the effectiveness of random forest on spectral classification. We used astronomical objects catalogue SDSS[5] to test random forest. This paper is organized as follows. Section Method introduces the random forest ensemble method. Section Data describes the data used in the experiments. Section Experiments and Results shows the experimental results. Section Conclusion presents the conclusion of this work.

2. Method

Random forest is an ensemble classifier consisting of a number of decision trees. Since the performance of the ensemble classifier is highly related to the correlation among each model in it, the trees are often constructed with some randomization. Randomization comes from two points: (1) Sub sampling the training data and each tree are grown with different data. (2) for each internal node, selecting some attribute for the split. Besides, each internal node contains a best split of the training data. While testing, the test sample is applied on each tree, and the final result is an aggregation of the results from all trees. This method was first introduced in [6], and further developed in [4].

3. Data

DOI: 10.18535/Ijecs/v4i12.24

The Slogan Digital Sky Survey (SDSS) is the largest optical survey of the astronomical bodies(objects) including stars,
galaxies, asteroids etc., and contains data of ~10^9 objects(data release 9) covering 1/3 of sky[5]. The images are taken in five
photometric bands u, g, r, i and z in the optical wavelength range 0.3-1.0µm. These bands provide enough information to
broadly classify these objects as stars. From the available spectra of the individual object's redshift, velocity, intensity of light and temperature are calculated. Figure 1 shows the image of star and Figure 2 shows its spectra. 1005 similar spectra were considered for the study. As the wavelength in the available spectra ranges from 3800 to 9200 Å [5].We choose only the visible part of wavelength from 3800 to 9200Å .

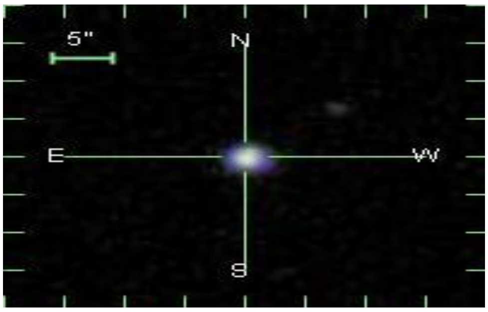

Figure 1 : Image of star

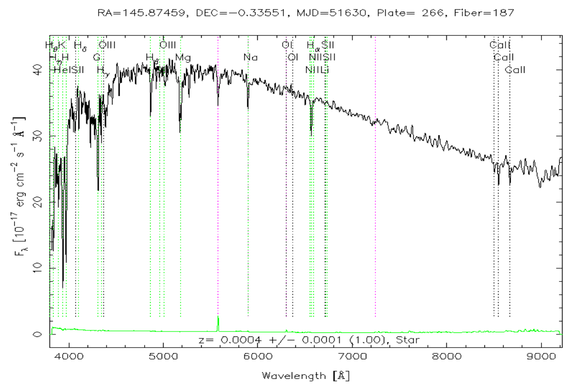

Figure 2 : Spectra of star

4. Experiments and Results

We designed random forest classifier using Salford Predictive Modeler (R) Software Suite: Randomforest (R) [7]. Morgan-Keenan (MK) system is widely used in astronomy classification. According to MK classification, stellar spectra were divided into totally 10 types: O, B, A, F, G, K, M, R, S, N. We generate class type as the target output of the random forest. As the wavelength in the available spectra ranges from 3800 to 9200Å , our model can only classify A, F, G, K, M class types. This section shows the prediction results of the developed model.

Sample size of 1005 records is taken. 3 predictors are tested at each node. Table 1 shows the Gini variable importance for all the attributes. 1 to 500 trees are generated and each time error rates are calculated.

Figure 3 shows the relation between the number of trees in random forest and error rates.

Figure 3 shows that as the number of trees in the random forest increases the error rate tends to decrease and becomes zero.
As error rate is zero the classification rate increases and the developed model can therefore be used for spectral classification.

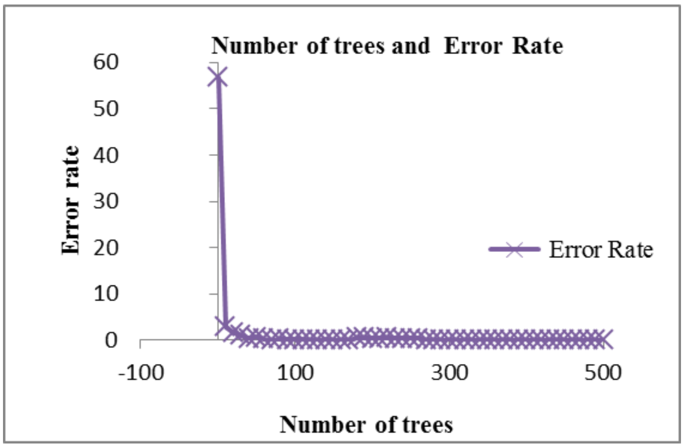

Figure 3: Number of trees in forest and error rate

Table 1: Variable Importance

Attribute	*Importance*
CONT_PEAK	100.00000
TEMPERATURE	80.79070
COLOUR__OF___STAR$	65.27440
U	24.42912
G	12.68100
R	7.73426
Z	7.32712
I	4.37654
INTENSITY	2.58200
DEC	1.28773
REDSHIFT_Z_	1.14793
RA	0.93758
VELOCITY	0.84313
ERROR_REDSHIFT$	0.00000

5. Conclusion

In this paper , we applied random forest ensemble method to spectral classification. The error rates are calculated by varying the number of trees in the random forest. As a result, at some point of time the error rate becomes zero and the model gives best classification results.

As a future work, we are going to consider more physical properties and different astronomical objects.

References

[1] Lasala J. in Chris corbally, Gray R O, Garrison RFE ditor. The MK process at 50 years. A Powerful tools for Astrophysical Insight. Astronomical Society of the pacific Conference Series, Proceedings of a Workshop of the Vatical Observatory, Tucson Arizona, USA, September, 1993.312

[2] Gulati R K, Gupta R, Ran NK. Astronomy and Astrophysica, 1997,322:933

[3] Qin DM, Hu Z, Zhao Y. In: Shen J, Pankranti S, eds. Object Detection, Classification and Tracking Technologies. Proceedings of SPIE. Bellingham: SPIE,2001, 4554:268.

[4] L. Breiman,"Random forests," Mach. Learn, vol. 45, pp5-32, 2001.

[5] D.G.York, et al.,and SDSS Collaboration. The Sloan Digital Sky Survey:Technical Summary. AJ, 120:1579-1587, September 2000.

[6] Y. Amit and D. Geman,"Shape quantization and recognition with randomized trees." Neural Comput., vol. 9, no.7,pp.1545-1588. Oct. 1997.

[7] http://www.salford-systems.com/