Hierarchical Classification of Variable Stars Using Deep Convolutional Neural Networks

M. Abdollahi [1], N. Torabi [2], S. Raeisi [2], S. Rahvar [2]

[1] School of Astronomy, Institute for Research in Fundamental Sciences (IPM), P. O. Box19395- 5531, Tehran, Iran

[2] Department of Physics, Sharif University of Technology, P. O. Box 11365-9161, Tehran, Iran

Institute for Research in Fundamental Sciences

Introduction

Variable Stars are important objects in astrophysics and cosmology. Most importantly, there are well-defined relations between period and absolute magnitude of Cepheid stars, which are relatively young, massive, radially pulsating stars. These stars are used as standard candles for measuring cosmological distance and calibrating type Ia supernovas [1], [2]. The Cepheids as the distance indicators also provide essential information on the size of our galaxy [3]. Type II Cepheids are also useful for studying stellar evolution as each subclass of this class is in a different stage of stellar evolution [4]. RR Lyrae stars as another class of variable stars can be used for studying the early history of galaxies. They are one of the oldest observable populations of stars which makes their chemical and dynamical evolution useful for understanding evolution of stars [5].

Data & Pre-processing

We used the OGLE-IV [6] variable stars database for training. Data contains 7 classes and 16 sub-classes in total. For each star, the data includes observation time in Julian days, magnitude and error bar of magnitude. The raw data from different surveys differ in the sampling rate and even in the number of observation points due to the science objectives and natural phenomena such as the weather and the moon. To make the models independent of sampling rate, the pre-processing method produces (folded and binned) light curves as input data.

To start pre-processing, first we use the period of variable stars, which is provided by the OGLE catalogue to fold the raw data. This step leads to obtaining the periodic behavior of the star in the phase space. Then, each folded light curve is divided into 50 equal bins to make the length of data points the same, and the value of each bin is set to the average of the values of the points in it.

The folded light curve has the disadvantage of losing period information which is a critical feature for classification of variable stars. To solve this problem, the value of the period is provided as an separate feature to the model.

Hierarchical Classification

We use a Hierarchical classification model. Namely, we use several classifiers to achieve better accuracy. The first classifier which we refer to as Classifier (a) separates ECL, RRLYR, LPV, DSCT, and a 5th class containing Cep, ACep, and T2Cep classes. we group the three classes that are less populated and similar together. When the first classifier predicts the 5th class, a second classifier is used to specify which subgroup, i.e. Cep, ACep, or T2Cep the star belongs to. Up to this point, the combination of the two classifiers makes one classifier that can identify the class of the variable star. Next, for each class that is composed of subclasses, a new classifier is trained to identify the subclasses. A schematic of the structure of our hierarchical classification model is depicted in Fig 1.

Having several classes with different subclasses increases the probability of misclassification. The Hierarchical Classification technique helps to reduce these errors when we deal with multi-class classification.

The classifiers used for the hierarchical model are identical in the network structure for both Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) [7],[8].

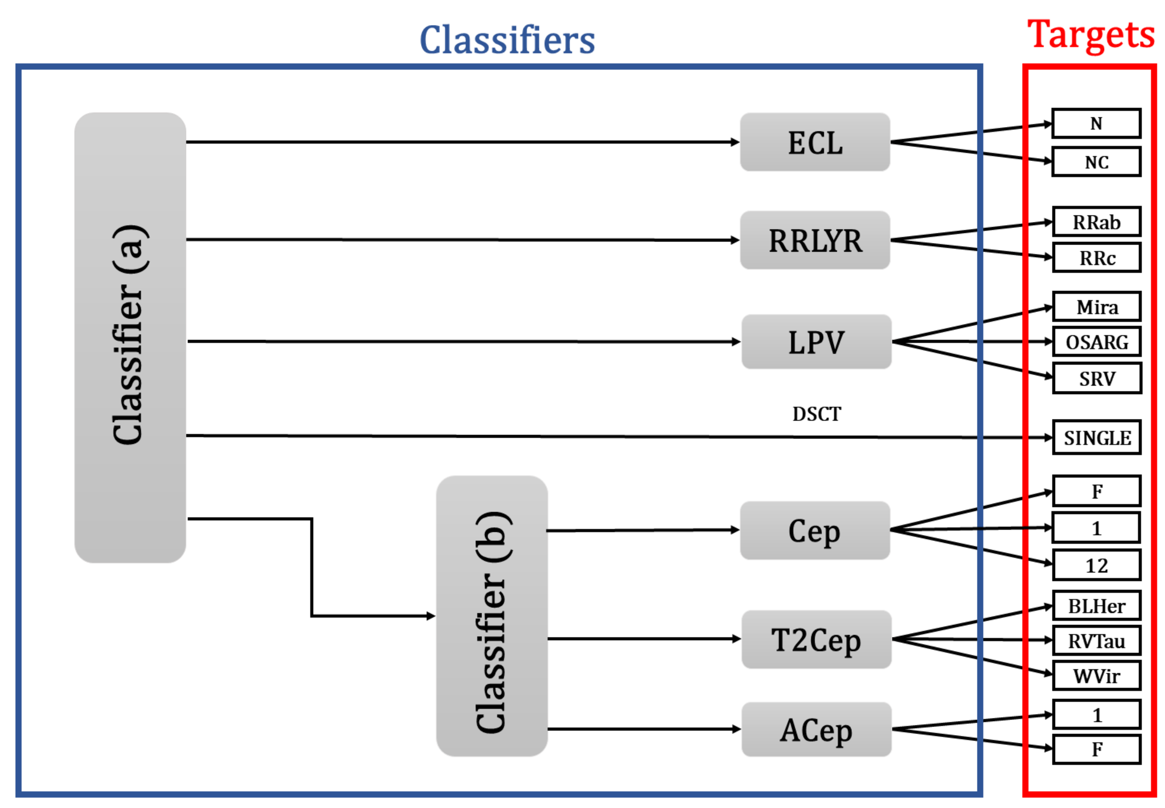

Figure 1. Schematic of Hierarchical Classification.

Classification Results

We have trained and tested different models (e.g. CNN, RNN). We consider two main metrics for evaluation: the accuracy of the classification and the time required for training. Table 1 shows the accuracy and time cost of our models. This indicates that the CNN model has better performance and is more efficient, i.e. faster.

Table 1. Comparing accuracy and runtime of each type of neural network used to classify variable stars.

Model	Accuracy	Training Time (min)
CNN	0.93	~ 24
RNN	0.90	~ 324

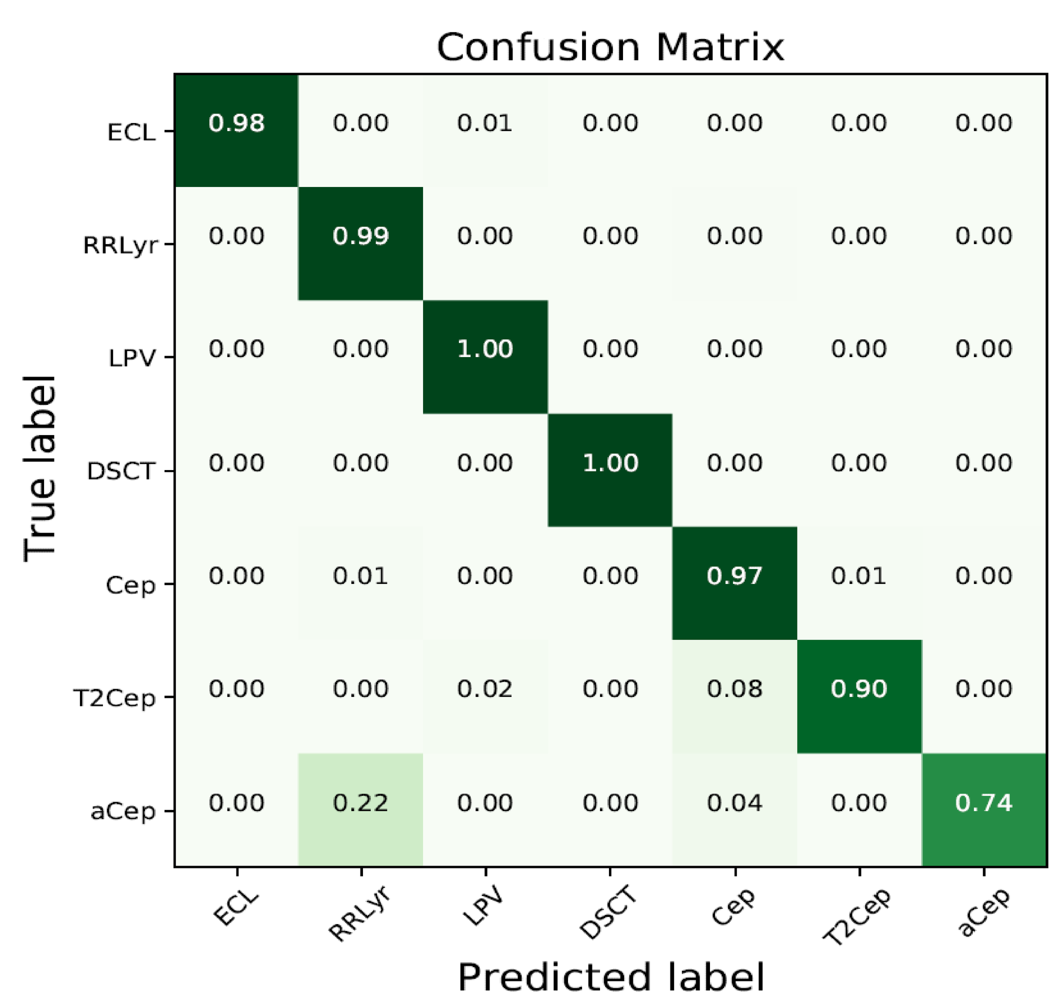

Figure 2. Confusion matrix of classification using CNN model.

The results of the best model in class and subclass classification are presented in two methods, confusion matrix and classification report.

The more the confusion matrix is diagonal, the more the classification is accurate. Checking the confusion matrix allows us to find misclassification made by the model. Therefore, the numbers out of the diagonal show misclassified classes and subclasses of.

Considering the confusion matrix, we find that misclassification in some classes are more than others. The main reason for these mistakes is the similarity of features of the two classes. Classes with a small population can be hard to predict, too, due to the data not being enough for training the network.

The Confusion matrix of classes is shown in Fig. 2. As an example, 22% of ACep stars have been predicted as RRlyr stars. The reason for this is the similarity between period and light curves of these two types of stars. To be able to enhance prediction of ACep stars, information of distance and absolute magnitude is needed [9].

As mentioned, we can evaluate the model by classification reports and find which subclasses have less accurate classification based on standard metrics, including recall and precision.

Table 2 shows that there are some subclasses with low recall and precision. For example, subclass 1 from ACep has recall 0.43 which means that our model classified 43% of all actual labels of this subclass. And, precision 0.25 means that only 25% of predicted stars were accurately classified as subclass 1 from ACep class.

This information implies that we have misclassification in this subclass which is a consequence of a small population compared to other subclasses, in addition to the similarity mentioned before.

Table 2. Results of subclasses classification using CNN model.

Class	Subclass	Precision	Recall
ECL	NC	0.91	0.92
ECL	C	0.68	0.57
RRLYR	RRab	0.98	1.00
RRLYR	RRc	0.96	0.98
LPV	MIRA	1.00	0.86
LPV	OSARG	0.96	0.97
LPV	SRV	0.79	0.84
DSCT	SINGLE	0.99	1.00
Cep	F	0.98	0.98
Cep	1	0.87	0.93
Cep	12	0.72	0.45
T2Cep	BLHer	0.94	0.88
T2Cep	RVTau	0.86	0.87
T2Cep	Wvir	0.95	0.89
ACep	F	0.92	0.98
ACep	1	0.43	0.25
weighted avg		0.93	0.93

Conclusion

We present a hierarchical model for classification of variable stars. The model has two neural networks for classification of the class of the star. Then each class is passed to a new neural network to identify the subclasses of the corresponding class. With this architecture, we manage to classify less populated classes/subclasses with high performance.

We see that Convolutional Neural Networks work faster than Recurrent Neural Networks and have more accurate predictions. We obtain an accuracy of 98% for class classification and 93% for subclasses classification. This work was done using OGLE-IV dataset. It would be interesting to extend this to other surveys like WISE and Gaia.

References

[1] Freedman W. L., et al., 2001, The Astrophysical Journal, 553, 47
[2] Riess A. G., et al., 2011, The Astrophysical Journal, 730, 119
[3] Catelan M., Smith H. A., 2015, Pulsating Stars
[4] Soszynski I., Udalski A., Szymanski M. K., Kubiak M., PietrzynskiG., Wyrzykowski L., Ulaczyk K., Poleski R., 2010, Acta As-tron., 60, 91
[5] Soszynski I., et al., 2009, Acta Astron., 59, 1
[6] Udalski A., Szymanski M. K., Szymanski G., 2015, Acta Astron.,65, 1
[7] Aloysius N., Geetha M., 2017, 2017 International Conference on Communication and Signal Processing (ICCSP), pp 0588{0592
[8] Lipton Z. C., 2015, ArXiv, abs/1506.00019
[9] Soszynski I., et al., 2008, Acta Astron., 58, 293

You can find all information about this project by scanning these two QR-code.

github

arXiv