Classification Between Galaxies and Stars with Two-Layer Perceptron

Wooseok Lee[*] Department of Energy Resources Engineering,
Seoul National University, Seoul, Korea
Email: *andylws@snu.ac.kr

***Abstract*—In this paper, I propose a quick and concise method to distinguish between galaxies and stars. Thanks to two-layer perceptron, classification between galaxies and stars could be done quickly in the early stages of astronomical research.**

I. Introduction

If you are at somewhere far from artificial lights, you can see a lot of bright spots in the night sky. You might think that all of those far tiny spots would be stars. That is correct, but to be more precisely, if you take a bright point in the sky, it could be a single star but also it could be a cluster of a lot of stars which is called galaxy.

Distinguishing between galaxies and stars is a very important task in starting astronomical research, especially in observational astronomy.

However, those unknown celestial bodies and structures are literally astronomical distances away. Therefore in so many cases, we cannot figure out the physical features of each target with only one photograph. That leads us to take multiple pictures of a target which are taken at a time interval and there, time delay follows. This also applies to the distinction between galaxies and stars.

What if a two-layer perceptron can distinguish them only with a single picture?

II. Problem Definition

The goal of this research is to construct a two-layer perceptron which can successfully recognize whether the object of the given image is a galaxy or a star. Also, it aims to minimize an error rate of the perceptron. I checked if the pixel size or the number of training epoch affects significantly in error rate.

III. Solving Approach

First of all, I converted 10,000 galaxies and stars picture data which are given as PNG files of 512×512 pixel size into PNG files with 28×28 pixels(Figure 1) and 56×56 pixels(Figure 2) each. This is to reduce time and memory required in calculating. 4,000 pictures for each galaxy and star are used for training and the other 1,000 pictures for each are used for testing.

Then, converting PNG files into CSV files follows. In front of each data, I added target numbers which indicate the real classification of the object; 0s for galaxies and 1s for stars.

Construction of two-layer perceptron with sigmoid function and a hidden layer with 100 and 200 dimensions for each image sets makes four perceptron structures in total. These perceptrons are compared with each other in terms of error rate to find out presence or absence of influence of the pixel size and the number of training epoch on error rate.

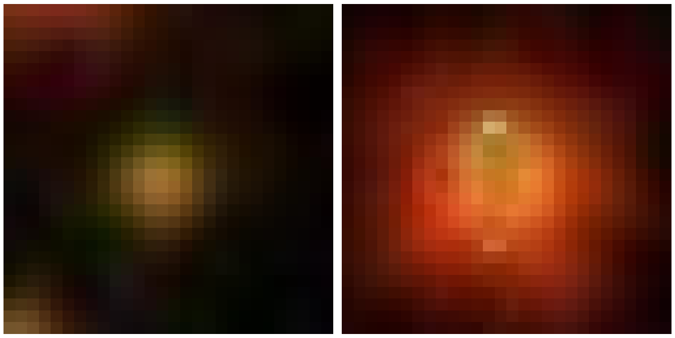

Fig. 1. The sample images of galaxy(left) and star(right) of 28×28 pixels.

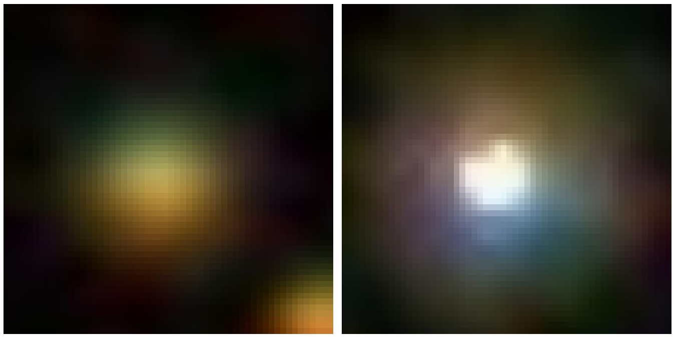

Fig. 2. The sample images of galaxy(left) and star(right) of 56×56 pixels.

IV. Evaluation

All four perceptrons are constructed successfully and four error rates are calculated. See the Table I.

Four error rates are almost same and are about 8.2% $\sim$ 8.5%. Error rates is changed slightly as the number of training epoch increased, but it is hard to say this amount of change

Training Epoch	28 × 28 pixels	56 × 56 pixels
5	0.0845	0.0815
10	0.082	0.082

TABLE I
Error rates from each perceptrons and image data.

is noticeable and significant. Rather in terms of image data of 56×56 pixels, the error rate increased as the number of training epoch increased.

V. Conclusion

I constructed concise galaxy-star classification algorithm using two-layer perceptron. It is encouraging that the error rate is lower than 10%. This error rate is low enough to be used in the early stages of astronomical research for quick and concise classification of galaxy and star. I hope that this perceptron structure and method will be used by astronomy researchers and save their precious time.

However, it is not recommended to use this in research requiring precision. Because the error rate is still large, the drastic reduction in error rate is certainly needed. Other kinds of machine learning methods might be the way for it.