Galaxy Zoo Challenge

Project Proposal

Irfan Hudda [11319]
Rohit Kumar Jha [11615]

March 1, 2014

1 Introduction

We are planning to work on an ongoing challenge on Kaggle which requires participants to classify the morphologies of distant galaxies in our Universe.
With each passing day telescopes around and above the Earth capture more and more images of distant galaxies. As better and bigger telescopes continue to collect these images, the datasets begin to explode in size. In order to better understand how the different shapes (or morphologies) of galaxies relate to the physics that create them, such images need to be sorted and classified [1].
Galaxies in some sets have already been classified once through the help of hundreds of thousands of volunteers, who collectively classified the shapes of these images by eye in a successful citizen science crowd-sourcing project. However, this approach becomes less feasible as data sets grow to contain of hundreds of millions (or even billions) of galaxies. [1]
This competition asks participants to analyze the JPG images of galaxies to find automated metrics that reproduce the probability distributions derived from human classifications. For each galaxy, participants are required to determine the probability that it belongs in a particular class. [2]

2 Motivation

Understanding how and why we are here is one of the fundamental questions for the human race. Part of the answer to this question lies in the origins of galaxies, such as our own Milky Way. Yet questions remain about how the Milky Way (or any of the other 100 billion galaxies in our Universe) was formed and has evolved. Galaxies come in all shapes, sizes and colors: from beautiful spirals to huge ellipticals. Understanding the distribution, location and types of galaxies as a function of shape, size, and color are critical pieces for solving this puzzle. [2]

3 Formal Problem Statement

The problem statement asks participants to answer a sequence of question thus forming a tree with these questions. [2]
Q1. Is the object a smooth galaxy, a galaxy with features/disk or a star? 3 responses
Q2. Is it edge-on? 2 responses
Q3. Is there a bar? 2 responses
Q4. Is there a spiral pattern? 2 responses
Q5. How prominent is the central bulge? 4 responses
Q6. Is there anything "odd" about the galaxy? 2 responses
Q7. How round is the smooth galaxy? 3 responses
Q8. What is the odd feature? 7 responses
Q9. What shape is the bulge in the edge-on galaxy? 3 responses
Q10. How tightly wound are the spiral arms? 3 responses
Q11. How many spiral arms are there? 6 responses

4 Approach

One obvious approach is to use Random Forest Classifier and an ensemble of such classifiers with varying parameters. It actually works well in practice when one is not sure on what to use. One another approach can be to use deep learning a it can be used to learn the underlying intricate features in images. We also plan to look into the physics of universes to use the related rules to improve classification.

5 Data

Data is provided on the Galaxy Zoo Challenge Page at http://www.kaggle.com/c/galaxy-zoo-the-galaxy-challenge/data/

References

[1] http://www.galaxyzoo.org/.

[2] http://www.kaggle.com/c/galaxy-zoo-the-galaxy-challenge.

[3] C.Cardamone. Galaxy zoo green peas: Discovery of a class of compact extremely star-forming galaxies.

[4] D.Darg. Galaxy zoo: The fraction of merging galaxies in the sdss and their morphologies.

[5] D.Darg. Galaxy zoo: The properties of merging galaxies in the nearby universe - local environments, colours, masses, star-formation rates and agn activity.