Classification of Morphologies for Galaxy

Authors: Alvin Dey, Pranay Raj Anand, Swagatam Chakraborti

Motivation

Improving the accuracy of manual morphological galaxy classification in order to aid the efforts of astronomers through automated metrics that reproduce the probability distributions derived from human classifications. The proposed system can extend the spectrum of classification problems across astronomical entities of different varieties.

Problem Statement

Understanding how and why we are here is one of the radical mysteries for the human race. Fragment of the answer to this problem lies in the origins of galaxies, such as our own Milky Way. As per reports from The University of Chicago ,the Sloan Digital survey conducted by Apache Point Observatory will produce more than 50 million images of galaxies in the near future. Stratification of these images is usually done by visual examination of photographic plates. Interpreting the distribution, location and types of galaxies as a function of shape, size, and colour are critical pieces for evaluating our place in the universe. We plan to build an algorithm to classify properties of galaxies with a degree of probability close to the annotation done by volunteers as part of classification project.

Objective

We will build a model that will categorize on the following features:

- Smoothness of galaxy
- Presence of a spiral pattern
- Prominence of central bulge
- Number of spiral arms
- Presence of odd features
- Roundness of a smooth galaxy

We will train the model using four approaches least square regression, ridge regression, lasso regression and random forest.

Project Tasks

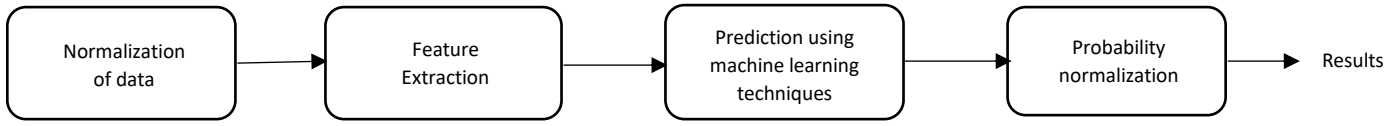

Figure 1 Project tasks pipeline

- Acquire image dataset from Kaggle challenge.
- Apply appropriate affine transformations for normalization of images.
- Feature extraction using PCA, SIFT and HOG.
- Classification using least square regression, ridge regression, lasso regression and random forest.
- Probability normalization to obtain results.
- Comparative analysis between above mentioned approaches.

Dataset

Figure 2 NGC 4414, a typical spiral galaxy

The training dataset contains 61,578 images of galaxies. For each particular image 37 different categories are identified as morphological representation of the galaxies by volunteers as part of Galaxy Zoo 2 project. A higher number close to 1 indicates that many volunteers voted for this morphology category with a high level of confidence while low number indicates the feature is likely not present.

Literature Review

Neural network[1], locally weighted regression using principal component analysis and Naïve-Bayes[2] classification techniques have been applied for this problem with various success rates.

The locally weighted regression technique[2] garnered 90% accuracy for binary classification(spiral and elliptical). Goderya and Lolling [1] achieved 97% success on 171 training examples using neural networks.

Evaluation-Metrics

The proposed methods will be evaluated on:

- Root Mean Square Error
- Mean Absolute Error

Time-Line

Mid- Evaluation
1.Image normalization
2.Feature extraction
3.Classification using LS regression.

Final Evaluation
1. Classification using ridge, lasso and RF.
2.Comparitive analysis

Reference

[1] Goderya, S "Morphological Classification of Galaxies Using Computer Vision and Artificial Neural Networks. "
[2] Calleja, J., Fuentes, O., "Machine Learning and Image Analysis for Morphological classification of galaxies."