Identifying Interesting Stellar Objects

Group Members: Ishaan Bhojwani, Alexander Tyan, Kevin Sun, Tyler Amos

Objectives

The core objectives of this project were to identify, from multispectral imagery, interesting stellar objects that warrant further investigation. The metric for success is whether or not our algorithms efficiently identify objects which could plausibly be:

1. Young Stellar Objects
2. Active Galactic Nucleii
3. Colliding Galaxies

This metric will be evaluated by reviewing images of objects which are identified by Algorithm 1, and comparing the output of Algorithm 2 with selected imagery.

Hypotheses

1. Notable objects can be identified as extrema in terms of infrared light
2. Some notable objects are grouped into interesting structures, which can be identified based on pairwise distances between objects

Dataset

The dataset used is drawn from the NASA/IPAC Infrared Science Archive, specifically the Wide-field Infrared Survey Explorer (WISE). This mapping of the entire sky contains records on each object and readings on the energies they emit. The entire dataset is approximately 800 million objects and 815 GB.

[add description of fields]

Overall Approach

This project was implemented using the MapReduce framework in Python. Algorithms were either developed from scratch or customized for use in a distributed environment. Details of both algorithms are included in the following sections.

- Big Data approaches used (e.g. "MapReduce")

- any Big Data related tools or techniques you learned that we did not cover in lecture

Algorithm 1: Pairwise Edges and Nodes

### Description of Algorithm

### Estimated Complexity

Algorithm 2: Random Walks on Sub-grids

### Description of Algorithm

### Estimated Complexity

Runtime Comparison

| Algorithm | Runtime | Master Spec. | Worker Spec | Master: Worker Ratio | Comment |
| --- | --- | --- | --- | --- | --- |
| Algorithm 1 |  |  |  |  |  |
| Algorithm 2 |  |  |  |  |  |

[ here we should put one graph which compares both algorithms runtimes]

Lessons Learned (Challenges/Solutions)

* challenges you faced, and which you successfully overcame or worked around, and which you were unable to

Summary of Results

| Algorithm | Number of Results | Success Metric | Degree of Success |
| --- | --- | --- | --- |
| Algorithm 1 |  | Comparison of sample of results to WISE images |  |
| Algorithm 2 |  | Comparison of results to larger WISE imagery |  |

Conclusion

It appears to be possible to identify interesting stellar objects from multispectral data using distributed computing environments. Adapting existing algorithms proves useful in flagging outliers and identifying unusual objects which appear to be clustered with other outliers. Applying standard graph algorithms such as random walks, and image thresholding methods from computer vision we are also able to generate a less granular but more comprehensive picture of where the most promising objects lie.

We believe these outputs provide a rough guideline for where follow-up astronomical investigation would be most fruitful. Future investigations of this topic could leverage labelled training sets and forms of machine learning which are amenable to distributed implementations.

Appendix