MAIS 202 Deliverable 2

Stephen Fay

October 2019

1 Problem Statement

I have some simulated experimental data. There are 200 x 200 x 200 images that represent different objects in a 300 Mega parsec cubed region of the univers; I am interested in two types of these, the mass-density ones found in the "updated_smoothed_deltax" files, and the ionization maps found in the "delta_T" files. My first goal will be attempt to match the files.

2 Data Pre-processing

I cut up the blocks of data into smaller blocks. Then in each sub-block I will cut 2D slices of the sub-block, and use these as the images and make a first attempt at classifying them. For now I will only be looking at ionization maps. The data is divided into folders based on two variables that were input into the simulation. There is the ESC_FRAC (the fraction of something [i'm not quite sure, maybe light] that escapes the system and ionizese the universe); and M_TURNOVER (the something Joelle explained it to me a while ago but I can't quite remember...). For now I will just work with ESC_FRAC.

So the pre-processing involved cutting up the boxes into smaller boxes, slicing up these smaller boxes, normalizing the slices and storing them as numpy-arrays.

3 Method - PCA on ESC FRAC

The most important thing at the moment is to obtain some form of successful result. I will look at data from a single ESC_FRAC value - 0.070 - but at different red-shift values (different distances). I used the euclidean distance and k-nearest neighbours method to try to find the red-shift value of a bunch of a sub-blocks.

The idea is to guess how far away it is for a certain ESC FRAC, then to guess the ESC FRAC based off of the picture alone. This part was done with principal component analysis using k-nearest neighbours method. In the end I was able to get 33 percent accuracy on each images, which is not bad, if we assume the odds of predicting it wrongly out of the other 16 classes are homogeneous, this gives us a not bad odds of predicting the correct class.

4 Future Work

Next I'm going to try to implement a CNN to do this better. Also I think CNNs will have better predictions, and will allow me to recognise the ESC FRAC and also hopefully the Turnover Mass.

5 References

1. https://academic.oup.com/mnras/article-abstract/483/2/2524/5228756?redirectedFrom=fulltext - a free pdf of this article can be found here https://arxiv.org/pdf/1807.03317.pdf

2. Had a look at Samuel Gagnon's research project as a resource (uses same / similar? data that I use) - https://github.com/samgagnon/remove-wedge