Weekly Journal

Leila Uy

July 19, 2021

1 Work Update

This week I spent a lot of my time working on a bash script process that would initalize a set number of EC2 instances on AWS to compute multiple k values at the same time. From our last meeting, you saw the format of my code and it is all in a GitHub repo. The problem I am encountering is that there is a CPU limitation on our tier in AWS that does not allow me to test more than 64 cores at a time. This means that I would have to total all the instances initalized to at most 64 cores and so if I want to test multiple, I cannot get more than a 16/24 core machine to run two other instances from the primary script running instance.

This makes testing difficult because if the size of the instance cannot handle the data/computation I ask for it to run, it will kill the process. Since Jishnu ran the world clustering on a 40 core machine with the 50 clusters taking 1 hour itself, I cannot test my process because my instances do not have the necessary memory or computational power.

The pro of this process is that for large lists of k we want to compute, it will shorten time by a significant amount because the computational time upperbound ($O(n)$) is limited to the highest cluster size + initalization process. I want to put it on the poster for future research, but because we are not able to fully test it, we cannot talk about it in-depth, possibly.

2 Literature Review

Since Jishnu and I are analyzing the methodology of ecoregion delineation, I tried to find articles on ecoregion delineation. One of the articles I read talked about the more philosophical side of vague and granular delineations like ecoregions, ecosystems, and biomes. It says that the goal of creating precise delineations contradicts building a sophisticated classification systems using the Aristotelian method. The article concluded that the basis of the ontology is the "notion of homogeneity with respect to geographic qualities" [1] which has lead to many misunderstandings of the resulting trade-offs. Bittner found that it is possible that many geographic regions exist and can be identified using scientific methods.

I had a really hard time understanding this article because it used a lot of definitions and axioms to prove their point. A lot of which I had to re-read but it was interesting to read about because we have so many classification systems and of course, there is no correct or standardized way of evaluating them. Ecoregion delineation is supported by the purpose behind the delineation and we can see that with a lot of the articles we read previously. In our case, the bioclim variables are extremely important, but it makes you wonder what variables should be added or removed.

The other article I read was "A freshwater ecoregion delineation based on freshwater macroinvertebrate community features and spatial environmental data in Taizi River Basin, northeastern China. [2]" They used ISODATA which from what I remember is a popular variant of k-means clustering algorithm which allows for a different number of clusters. It is really interesting because they had three important features to their approach:

1. delineation variables were filtered to avoid redundancy

2. quantitative and repeatable

3. boundaries were checked for consistency with spatial attributes

These features are reflective of the process I want to create especially an approach that is quantitative and repeatable. Now their conclusions are not really relevant to us because they are based mostly on freshwater delineation rather than our analysis on land, but the interesting part of their research is the way they filtered through the variables. I do not think we have the time for this but it was interesting to see that they used correlation and ordination analysis to filter through their results.

References

[1] Thomas Bittner. Vagueness and the trade-off between the classification and delineation of geographic regions - an ontological analysis. *International journal of geographical information science*, 25(5):825–850, 2011.

[2] Weijing Kong, Wei Meng, Yuan Zhang, Christopher Gippel, and Xiaodong Qu. A freshwater ecoregion delineation approach based on freshwater macroinvertebrate community features and spatial environmental data in taizi river basin, northeastern china. *Ecological research*, 28(4):581–592, 2013.