Weekly Journal

Leila Uy

May 17, 2021

1 Work Update

This past week I read the assigned article but spent most of my time focused on learning the concepts and architectures of parallel programming and the integration of parallel programming in R and GRASS.

1.1 Parallel programming

I watched many videos and readings about parallel programming because I was finding it difficult to understand academic writings about it without knowing the basics. So, I started from the beginning. There are two major types of parallel architectures: distributed memory and shared memory. Distributed memory uses a message passing system, while shared memory access the main memory via uniform memory access (UMA) or non-uniform memory access.

When parallel programming, there are different computation bounds aside from the CPU. Therefore although in theory adding processors would linearly increase the thoroughout of computation, in reality, we are bounded by memory, I/O, and network.

1.1.1 R

Several parallel programming packages in R include parallel, caret, multidplyr, and doParallel [2, 4]. Different packages are useful for different situations like the parallel::mclapply function. It is useful for repetitive independent tasks that can be done in parallel.

1.1.2 GRASS

Parallel programming in GRASS is interesting because GRASS is not a monolithic application but consists of over 300 autonomous modules with their own memory management and error handling. An article I read used a shared memory parallelization tool (OpenMP) in GRASS [1]. In the Literature Review, I will discuss this article more but from what I understand there is more freedom for parallel programming in GRASS than R. R feels very restrictive in the packages and functions but it is easier to implement.

1.2 Difficulties/Questions

I am trying to learn the concepts and architectures because I believe we need to find the best fit for the situation. So, I am trying to learn about the best possible solution for a parallel k-means clustering for large datasets.

- Should we use a distributed or shared memory?
- How can we separate a k-means clustering into independent tasks?
- What are the bounds for our algorithm?
- How should we analyze the optimal amount of threads?

I am hoping to start practicing and coding parallel functions to deepen my learning next week because it is difficult for me to fully grasp the concepts without trying it myself.

2 Literature Review

The article I focused my time on throughly reading was "Parallelization of interpolation, solar radiation and water flow simulation modules in GRASS GIS using OpenMP". The study explored the implementation of shared memory parallelism for three computationally intesive modules: v.surf.rst, r.sun, and r.sim.water. The article throughly went through the thought process and steps that came into the parallelization of their code and I found this extremely useful because I am hoping to implement the same steps when I arrive at that step in my process. The article was also very through on describing the terminology easily for those learning parallel programming and providing sources of other sources. [1]

The assigned article for this week had me thinking of the location of my study area. In the article, they focused on California because the current and future trends in climate will impact the agriculture and economy of California. It is a very diverse landscape and permanent crops will be hit hard because they commonly grow for more than 25 years which make them vulnerable to impacts in climate. I am currently not strict on the location I want to focus on but, California is high on my list. [3]

References

[1] Jaroslav Hofierka, Michal Lacko, and Stanislav Zubal. Parallelization of interpolation, solar radiation and water flow simulation modules in grass gis using openmp. *Computers & Geosciences*, 107:20–27, 2017.

[2] Matt Jones. Quick intro to parallel computing in r. https://nceas.github.io/oss-lessons/parallel-computing-in-r/parallel-computing-in-r.html, 2017.

[3] Tapan B Pathak, Mahesh L Maskey, Jeffery A Dahlberg, Faith Kearns, Khaled M Bali, and Daniele Zaccaria. Climate change trends and impacts on california agriculture: a detailed review. *Agronomy*, 8(3):25, 2018.

[4] Eduardo A Rubia. Multicore data science with r and python. https://blog.dominodatalab.com/multicore-data-science-r-python/, 2017.