
A Distributed Architecture Proposal for Regressional Generative Adversarial Networks for Biostatistical Analysis Modeled after the Decentralized-Internet SDK and BOINC

The Lonero Foundation

Andrew M.K. Nassief

andrew@etherstone.org

Abstract

Distributed computing and parallel processing are often used for offloading large amounts of data in instances such as BOINC. Projects, such as the Decentralized-Internet SDK also allow for people to build instances of cluster computing projects for the offloading of data or decentralized architecture. Generative Adversarial Networks are currently used by AI experts in order to generate data that would have otherwise been non-existent. Given that certain biomedical datasets only have a small amount of donors or case studies available, means that more data would allow for a higher degree of accuracy. Since, certain diseases may not have enough donors or resources to collect that data, one method may be mathematically creating viable artificial data. This however, requires large amounts of processing. The utilization of a regressional model that would allow for a generative adversarial network (GAN) to recursively build medical data sets based off of pre-existing data in order to increase the statistical pool of accuracy should be feasible with distributed computing. This approach should also be worth trying in the case of absolute unknowns and false positives.

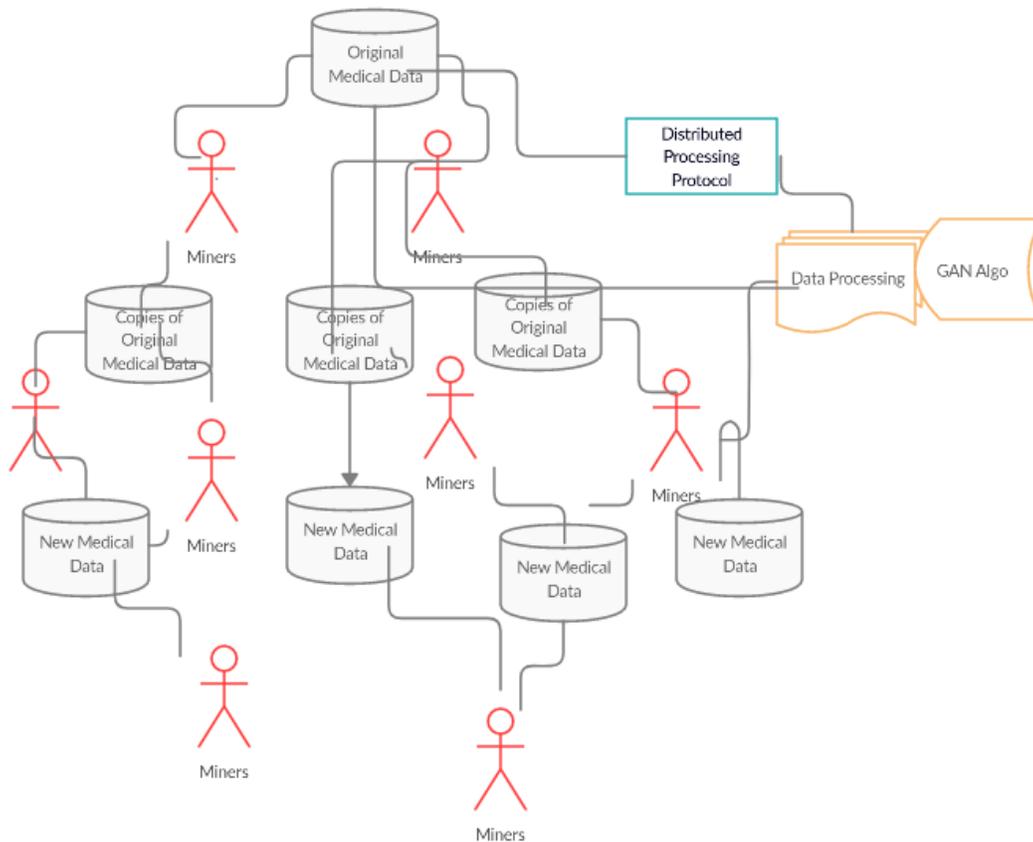
1.0 Problem Statement

For findings in statistics, one wants to have a high degree of significance.[1] The more data you have, the higher the degree of accuracy. For example, the cancer known as Diffuse Intrinsic Pontine Glioma (DIPG)[2] with a high degree of rarity relies many statistical unknowns. Given this, and the rarity of survival, likely not much data is readily available to get a full sense of knowledge on DIPG. Other abnormalities could include medical diseases that have genetic variants, cardiac diseases, and such that could utilize a better sense of higher degree of accuracy in data. Grid Computing architecture such as the introduction of BOINC[3] i.e “Berkley’s Open Infrastructure for Network Computing” allow for

offloading of large amounts of data through parallel processing. Other projects such as the Decentralized Internet SDK[4] allow for people to build distributed computing clusters and instances in support of decentralization. A proposal is to create a distributed processing program for the allowance of a regressional GAN that would increase the amount of biomedical data being analyzed in order to receive a higher degree of accuracy for the researcher to viably conclude results upon.

2.0 Proposal

In order for the distributed processing architecture to be configured correctly for a biostatistical GAN, the following UML diagram is proposed:



Starting with the original medical data, the data is distributed into copies instead of centralized and those copies are slowly being utilized into a distributed GAN network through parallel processing. Over time, the network should receive better latency given the higher processing capacity.

3.0 Technological Contributions

The Lonero Foundation have created the decentralized-internet SDK and contributed to the development of [Cancer@home](#). [5] [Cancer@Home](#) utilizes a regressional AI for analyzing cancer genomic case studies, and is built untop of a decentralized operating system. Currently, bioinformatics

research is being done in order to actually create a protocol to allow miners/workers to run a generative adversarial network in order to solve challenges computational genomics research companies are facing. Besides [Cancer@home](#) and the distributed-internet SDK, other contributions include: working towards decentralized variations of PyTorch[6], regressionally analyzing ecological datasets following the Darwin core format[7], and Data Mining Cancer Genomics Case Studies[8]. Work is being done towards creating a series of open standards, software and research towards grid computing and biostatistic-related GANs.

References

- [1] Ali, Z. & Bhaskar, S. B. Basic statistical tools in research and data analysis. *Indian journal of anaesthesia* (2016). Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5037948/>. (Accessed: 23rd December 2019)
- [2] Diffuse Pontine Glioma: Boston Children's Hospital. *Boston Childrens Hospital* Available at: <https://www.childrenshospital.org/conditions-and-treatments/conditions/d/diffuse-pontine-glioma>. (Accessed: 23rd December 2019)
- [3] High-Throughput Computing with BOINC. *BOINC* Available at: <https://boinc.berkeley.edu/>. (Accessed: 23rd December 2019)
- [4] Kamal, A. M. decentralized-internet. *npm* (2019). Available at: <https://www.npmjs.com/package/decentralized-internet>. (Accessed: 23rd December 2019)
- [5] Cancer@Home. *Stark Drones - OS* (2019). Available at: https://www.starkdrones.org/home/os#h.p_-bdQF7aLVCfp. (Accessed: 23rd December 2019)
- [6] Kamal, A. DecentTorch. *Devpost* (2019). Available at: <https://devpost.com/software/decenttorch>. (Accessed: 23rd December 2019)
- [7] Kamal, A. Regression Analysis for Ecological Datasets. *Devpost* (2017). Available at: <https://devpost.com/software/regressional-analysis-for-ecological-datasets>. (Accessed: 23rd December 2019)
- [8] Kamal, A. Data Mining Cancer Genomics Case Studies. *Devpost* (2017). Available at: <https://devpost.com/software/data-mining-cancer-genomics-case-studies>. (Accessed: 23rd December 2019)