# Big Data Approach to fMRI data analysis with Intel DAAL and Full Correlation Matrix Analysis

Haoran Shu (CUHK) , Yin Lok Wong (HKU)
Mentors: Pragnesh Patel, Kwai Wong, Junqi Yin

## Abstract

In this project we make use of the open source Spark Nifti reader library – biananes, dedicated for scalable fMRI data analysis, together with Intel Data Analytics Acceleration Library(DAAL), to realize analytic operations of fMRI data on the Big Data framework Apache Spark and replacing the use of Apache Spark MLlib with DAAL.

At current stage, tested results on computations with DAAL on Apache Spark do excel the performance of that with Apache Spark MLlib.

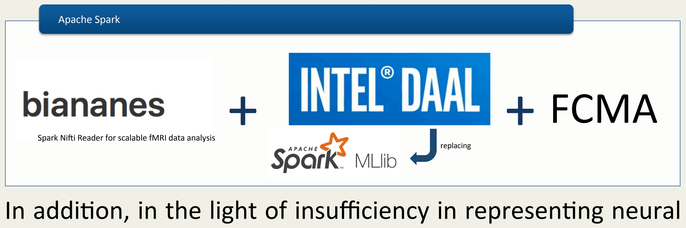
In the end of the project, we want to explore the possibility of multivariate pattern analysis, or more specifically, the implementation of Full Correlation Matrix Analysis under Apache Spark and evaluate the performance.

## Introduction

fMRI data analysis deals with data represented in large scale matrices and operations numbered easily in Giga magnitude, which breeds an ideal scenario for the use of Big Data framework such as Apache Spark. However, the use of Big Data framework in fMRI data analysis is limited due to the difficulty of representing fMRI data in distributed data sets that can then be utilized by such frameworks.

The use of open source Spark Nifti Reader library - biananes enables nifti image files of fMRI data to be read into distributed data sets for Apache Spark, and thus available to MLlib function calls.

However, some MLlib functions are limited in performance and restricted to certain input size. To enhance the library, we try to incorporate the use of Intel Data Analytics Acceleration Library with Apache Spark, in the hope to boost performance and expand usability.

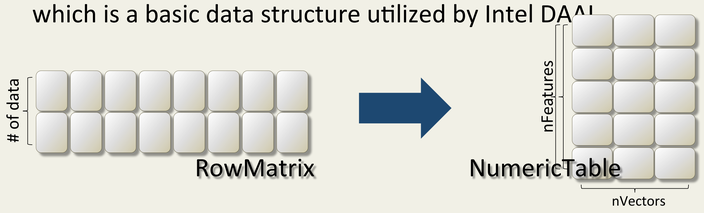



In addition, in the light of insufficiency in representing neural activities of the current univariate analysis approach, we want to experiment the use of Full Correlation Matrix Analysis under the Big Data framework with the above improvements.

## Spark Nifti Reader library - biananes

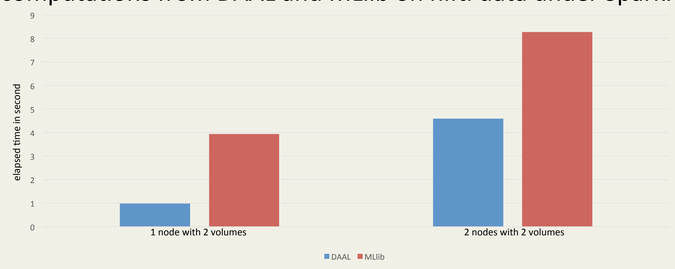
biananes is a library that reads fMRI data in Nifti(.nii) format and returns results as RowMatrix RDD for the use of libraries and operations under Apache Spark, such as MLlib.

However, in the attempt to experiment with Intel DAAL, the return type of RowMatrix is no longer useful. Instead, the main task in this part of the project is to transform the row oriented RowMatrix into Homogeneous Numeric Table, which is a basic data structure utilized by Intel DAAL.



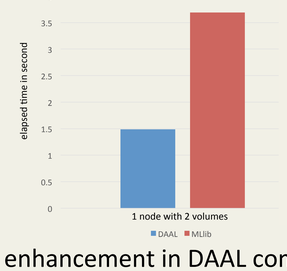

## Performance Comparison

The following shows performance comparison of different computations from DAAL and MLlib on nifti data under Spark.




The above diagram shows the time taken for SVD computation with DAAL and MLlib respectively on the same data set.

While both results are generated from computations under Apache Spark, the ones on the left specify results from single node computation while the ones on the right from computation in cluster mode with 2 compute nodes.




The diagram on the left shows performance comparison on QR decomposition, while computation in MLlib was launched via the interactive shell in Spark.

While results show performance enhancement in DAAL comparing to MLlib, further testings with multiple volumes on multiple nodes will be needed.

An additional note is that computation attempts on Principal Component Analysis and Covariance were taken but failed to draw comparison as with input matrix size larger than 65535, these computations in MLlib are not supported.

## Full Correlation Matrix Analysis

FCMA is a method that performs unbiased multivariate analyses of whole-brain functional connectivity, which makes use of the temporal correlation in fMRI activity of each voxel in the brain with every other voxel. It extends MVPA to not only concern the activities of voxels but also the correlations (interactions) between them.

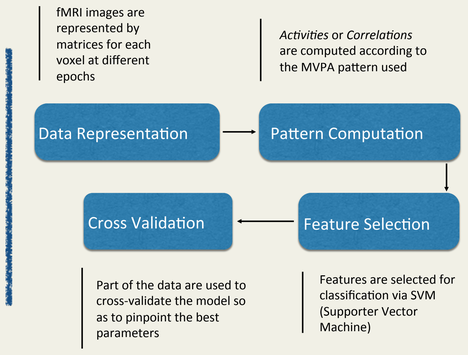
Parallelism

Correlation Computation
- Optimized to a matrix multiplication by normalizing the data
- Paralleled with MPI

Classifier Analysis
- LibSVM used
- GotoBlas to further speed up

MVPA
- Activity Patterns
- Correlation Patterns

Classification Procedure




## Computing Platform


All computation and testing mentioned are done on the HPC cluster "Beacon" at NICS, on single node or in cluster mode as specified, under the Big Data framework - Apache Spark, version 1.5.2.

## References

1. Roland N Boubela et al., Big Data approaches for the analysis of large-scale fMRI data using Apache Spark and GPU processin: A demonstration on resting-state fMRI data from the Human Connectome Project, Frontiers in neuroscience 9 (2015)

2. Yida Wang et al., Full correlation matrix analysis of fmri data, technical report.