ATHENA Tutorial

Installation:

- Download the ATHENA source file from http://ritchielab.psu.edu/ritchielab/software
- Unzip the tar ball `athena-1.1.tar.gz`
  `tar -xvzf athena-1.1.tar.gz`
- `./configure`
- `make`
- `make install`

Note: The configure script will compile a parallelized version of ATHENA if MPI is installed on the system. A serial version can be compiled on a system with MPI using "--with-mpi=no" when calling configure. A full list of options for the configure script can be seen by running "./configure –help"

Files included in the tutorial:

All files used in this tutorial are contained in the /samples directory of the package except for the series of 100 datasets in the /sample_sets directory.

DATASETS:

casecon_data.txt
DESCRIPTION: Dataset was generated using genomeSIMLA.
-25 SNPs
-Heritibility = 0.1
-Minor allele frequency of 2 functional SNPs = 0.4
-Two functional SNPs are G1 and G10
-1000 cases and 1000 controls
-100 datasets for each model

continout_data.txt
DESCRIPTION: Dataset contains continuous outcome and matches with the contінfile.txt.
4000 individuals

continfile.txt
DESCRIPTION: Continuous variable input dataset for use with continout_data.txt
50 variables
4000 individuals

/sample_sets/

DESCRIPTION: 100 datasets were generated using genomeSIMLA to demonstrate how to run a simulation power study with ATHENA.
-25 SNPs
-Heritibility = 0.1
-Minor allele frequency of 2 functional SNPs = 0.4
-Two functional SNPs are G1 and G10
-1000 cases and 1000 controls
-100 datasets for each model

MAPFILES:

casecon_map.txt
DESCRIPTION: Matches the casecon_data.txt file with information on 25 variables in that set.

contin_map.txt
DESCRIPTION: Matches the continfile.txt file with information on 50 variables in that set.

CONFIGURATION FILE:

casecon.GENN
DESCRIPTION: Runs the analysis on the case control dataset (casecon_data.txt) provided.

continout.GENN
DESCRIPTION: Runs the analysis on the continuous outcome dataset (continout_data.txt) provided.

HELPER SCRIPT:

makeathenaconfig.pl

Script will make 100 configuration files that correspond to the 100 datasets for each model. While running the script, please specify the following five options:

- `-p` Prefix for datasets
- `-s` Suffix for datasets
- `-d` Directory containing datasets
- `-a` Algorithm to use (GENN or GESR, must match grammar!)
- `-g` Full path to grammar file (must match algorithm!)

Specifying the output directory is optional:

- `-o` Output directory (optional -- set to dataset directory when omitted)

Command running the sample sets (you can also type `./makeathenaconfig.pl` for USAGE):

```
./makeathenaconfig.pl -p Sample_Dataset-  -s .txt -g
/path/to/grammar/file -d /path/to/datasets –a GENN
```

GRAMMAR FILES:

Grammar files describe the grammars for the algorithm. One might choose one of them for the models. The grammar files are in the /samples directory.

genn_all.gram – Grammar File for GENN (all nodes)

genn_add.gram - Grammar File for GENN (Add-only nodes)

symbolic_regress.gram - Grammar File for GESR (+, -, /, * functions).

Running ATHENA

If you want to run ATHENA for single dataset just execute the command:

```
./bin/athena #.config
```

For example, if you are in the samples directory and wish to run the case control analysis:

../bin/athena casecon.GENN

If you have compiled ATHENA with MPI, you will need to call mpirun in addition as below:

mpirun –np 2 ../bin/athena casecon.GENN

The –np argument controls the number of processes that will be started.

If you are on a system that uses PBS queuing (as we have at PSU), you will need to submit a PBS script to have your job run. A sample PBS script called SampleAthena.pbs is included in this tutorial. The script **run_athenap.pl** will be used to run athenap on all 100 configuration files.

In the **pbs script** SampleAthena.pbs specifies the number of processors you want and make sure it matches the number after -np in the run_athenap.pl script so as to run it more efficiently.

Examples of the two paths in pbs:

```
/path/to/run_athenap.pl
/path/to/configs/*.config
```

Commands for Running ATHENA simulations:

Create Configuration Files

Execute makeathenaconfig.pl:

```
./makeathenaconfig.pl -p Sample_Dataset- -s .txt -d
/path/to/sample/dataset -a (GENN/GESR) -g
/path/to/grammar/file/genn_add.gram  -o
/path/to/output/directory
```

Update run_athena.pl and run_athenap.pl

You need to update these to PERL scripts by providing the path to ATHENA executables.

Update PBS Script

Update the PBS script by specify the paths of **run_athenap.pl** and (***.config**)configuration files directory

```
vi SampleAthena.pbs
```

Run ATHENA

Execute PBS script **SampleAthena.pbs:**

```
qsub SampleAthena.pbs
```

RESULTS:

By default configuration, ATHENA generates following result files:

- *.athena.sum (Summary File)
- *.best (Best model for each cross validation)
- *.dot (Dot compatible file for visualization)

athena_power.pl summarizes results from *.sum files after ATHENA is run.
Execute **athena_power.pl** to see the usage:

```
./athena_power.pl
```

This script summarizes the number of times either or both of the functional SNPs for example G1,G10 were present in your best model. Run in the directory where your summary files (*.athena.sum) are located.

Try on your own:

It is recommend playing with parameters in configuration file like popSize and number of generations to compare the power of ATHENA to find either one or both of the functional SNPs in the best model.

It should be noted that due to the nature of GENN programming, running multiple times may yield different results.