Genetic Approach With Guided Mutation How To's

Data Format

To construct the network with a specific data set you need to give the data input in the following manner. Each row in the input file will be a comma separated line. First attribute will be the index number. The next comma separated values are input attributes and the last one is the output for the input vector.

e.g. Suppose a input vector is (1 1 0.9 5.9) and the output is 5. The row entry for this data may be contain 3,1,1,0.9,5.9,5 . Which means its the Third input vector, the input vector is (1 1 0.9 5.9) and its output is 5.

How To Give Input

To use the code for a specific dataset, several global variables and data file name should be specified in the code. They are mentioned below :

File Name : *main.cpp*
line 516 $->$ set training data file name.
line 517 $->$ set validatioin data file name.
line 014 $->$ set $TRAIN\_SIZE$, it is number of training data in the given data set.
line 015 $->$ set value of $ATTR\_NUM$, it should be the number of input attributes.
line 027 $->$ set value of $VALIDATION\_SIZE$, it should be the number of validation data in the given data set.

File Name : *marchand.h*
line 008 $->$ set $TRAIN\_SIZE$, it is number of training data in the given data set.
line 009 $->$ set value of $ATTR\_NUM$, it should be the number of input attributes.

File Name : *dnc.h*
line 018 $->$ set $TRAIN\_SIZE$, it is number of training data in the given data set.
line 019 $->$ set value of $ATTR\_NUM$, it should be the number of input attributes.

File Name : *test_network.cpp*
line 012 $->$ set $TRAIN\_SIZE$, it is number of test data in the given test data set.
line 013 $->$ set value of $ATTR\_NUM$, it should be the number of input attributes.
line 088 $->$ set validation data file name.

How To Run

Follow these steps to run training and testing process :

- Locate the *Makefile* in the source folder.
- Issue *make* command from the directory of the *Makefile*. It will produce a binary file named network.
- Now run *./network* from the console, this will complete the training process.
- From the console issue this command *g++ -o test test_network.cpp f2n2.cpp*
- Run *./test* to see the error on test data for our approach, dnc and marchand.

Tuning The Parameters

In file *main.cpp* there are some adjustable parameters. By changing the value of those different output can be achieved. Some of them have some constraints over their values. The adjustable parameters are -

- $ALPHA$: weight on number of hidden neurons of the selection function($0.0 \leq ALPHA \leq 1.0$). The greater value of $ALPHA$ will increase the probability of reducing the number of hidden neurons.
- $BETA$: weight on the validation error ($0.0 \leq BETA \leq 1.0$). The greater value of $ALPHA$ will increase the probability of reducing the validation error. $ALPHA + BETA$ must be 1.0 for all values of $ALPHA$ and $BETA$.
- $INIT\_NET$: The number of networks formed with DNC and Marchand in each generation. Therefore the total networks formed in each generation is $2 \times INIT\_NET$.
- NUN_GEN : Number of generations.
- $EPOCH\_DNC$: Number of maximum epochs the DNC may run.
- $TRAIN\_TIME$: Training time in seconds for each instance of DNC and Marchand in each generation.