

Stack Overflow is a question and answer site for professional and enthusiast programmers. It's 100% free, no registration required.

Take the 2-minute tour ×

Batch training of a very large data set using matlab neural network toolbox

HANDS-ON CUDA[®] TRAINING IN THE CLOUD

OPENACC · CUDA C/C++ · CUDA FORTRAN · CUDA PYTHON · GPU-ACCELERATED LIBRARIES [LEARN MORE >>](#)

I want to use MATLAB neural network toolbox for speech recognition in KTH data set. Now my training data is so large that I cannot load them into a large matrix for batch training at once. One solution I found is use incremental training using adapt with chunks of data at one time. But as far as I know, this may reduce the accuracy.

I was using NICO toolkit earlier. In it we can give the input as names of file containing training data. It will read the files and do batch training. But I couldn't find such an option in MATLAB.

Is there a way to do batch training for such large data sets in MATLAB ?

[matlab](#) [machine-learning](#) [speech-recognition](#) [bigdata](#)

edited Mar 2 at 20:10



[Parag S. Chandakkar](#)

2,910 1 9 25

[add comment](#)

asked Mar 2 at 17:45



[Hari](#)

6 1

1 Answer

I would not recommend adaption for very large datasets. Adaptive learning is best for datasets that represent a changing relationship (dataset) over time.

If you have access to the Parallel Computing Toolbox and MATLAB Distributed Computing Server you can use the Neural Network Toolbox to spread calculations and data across multiple machines.

To increase the size of dataset that can be trained within a specific amount of RAM, use TRAINSCG (Scaled conjugate gradient, which uses less memory than Jacobian methods, or alternatively use TRAINBR (Bayesian Regularization Jacobian) with memory reduction (trades off time for memory space).

Assuming you still don't have enough RAM, one possible solution is to train multiple networks on different random subsets of the data using the TRAINBR with memory reduction. TRAINBR attempts to find the simplest possible solution, yet each training session will likely find a quite different solution. After training several dozen (or more) neural networks on different sets of data the outputs of each network can be averaged. Essentially a lot of simple solutions are combined to find a complex relationship.

Be sure to maintain some data which is not used to train any of the TRAINBR neural networks so that the generality of their combined (averaged) outputs can be independently measured.

answered Mar 21 at 0:12



[Mark Hudson Beale](#)

111 1

[add comment](#)

Not the answer you're looking for? Browse other questions tagged [matlab](#)

[machine-learning](#) [speech-recognition](#) [bigdata](#) or [ask your own question](#).