

Natural Language Processing First Stage
By Ryan Meuth
Introduction to Neural Networks Semester Project
5/14/04

Objective:

To produce a neural network based word classification tool and determine suitability for use as the first stage of a Natural Language processing scheme. Performance of various learning algorithms, and architectures are analyzed and compared.

Background:

Accurate Natural Language processing has long been a largely unfulfilled desire of computer science. The ability to speak to our computers in the way that we would speak to another human being, without any special training or language, would promote meaningful interaction with our machines, and allow a level of productivity previously unfathomed in the computer science world. Despite the usefulness of such a technology, the ambiguities of language have proven to be too drastic for conventional rule-based programming methods, and non-deterministic systems, such as neural networks, have only been applied to the problem in recent years. The application of neural networks to language processing has allowed systems to learn the patterns of language implicitly, by example, rather than the explicit definition of rules of language that are by no means absolute.

Intention:

This report deals primarily with the development of a neural network based word classification stage of a natural language processing system. This early stage classifies a word as it's part of speech, such as a noun, verb, adjective, etc. This information, and its relative location in a sentence, could then be used by a later stage to determine the subject and predicate of a sentence, which could then be used to determine meaning.

Approach:

A list of the top 1000 most common English words and corresponding parts of speech were compiled to be used as training exemplars for the classifier network. A 5-bit binary encoding scheme was devised to reduce the necessary size of the network, and a parsing routine was written to construct appropriate input patterns from typed words for the MatLab-based neural network. The words were classified into 9 categories:

- “dart” – Definite Article
- “iart” – Indefinite Article
- “prep” – Preposition
- “conj” – Conjugate
- “adverb”
- “noun”
- “verb”
- “adjective”

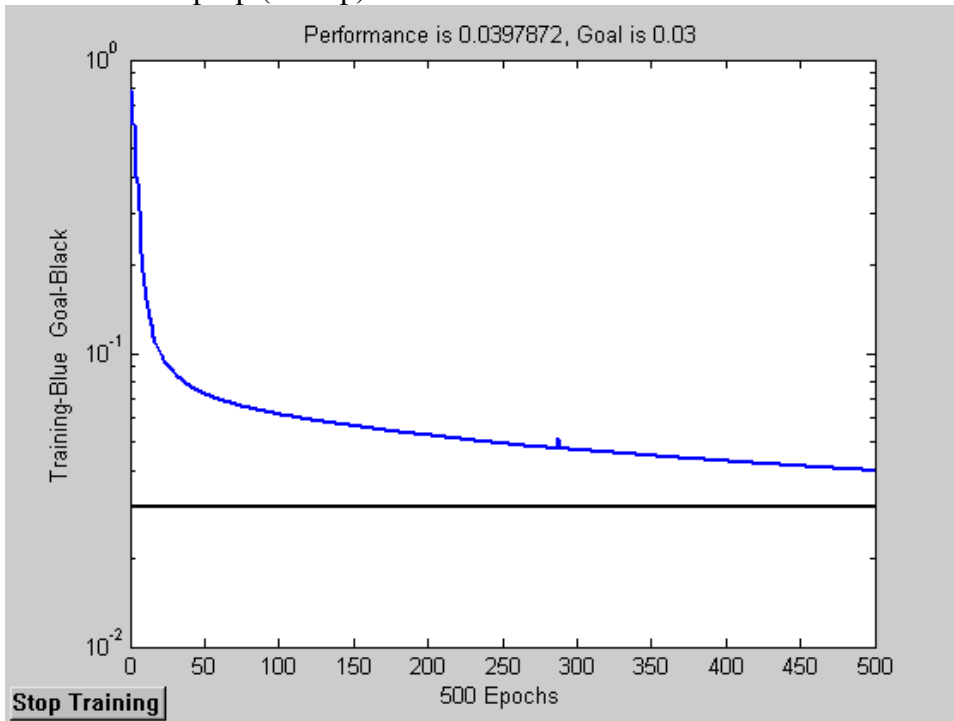
“pronoun”

A two-layer back propagation network with 55 inputs and 9 outputs was constructed, and the training set was presented for 500 epochs with a training goal of 0.03 RMS. Testing functions were implemented, and the network evaluated.

Results:

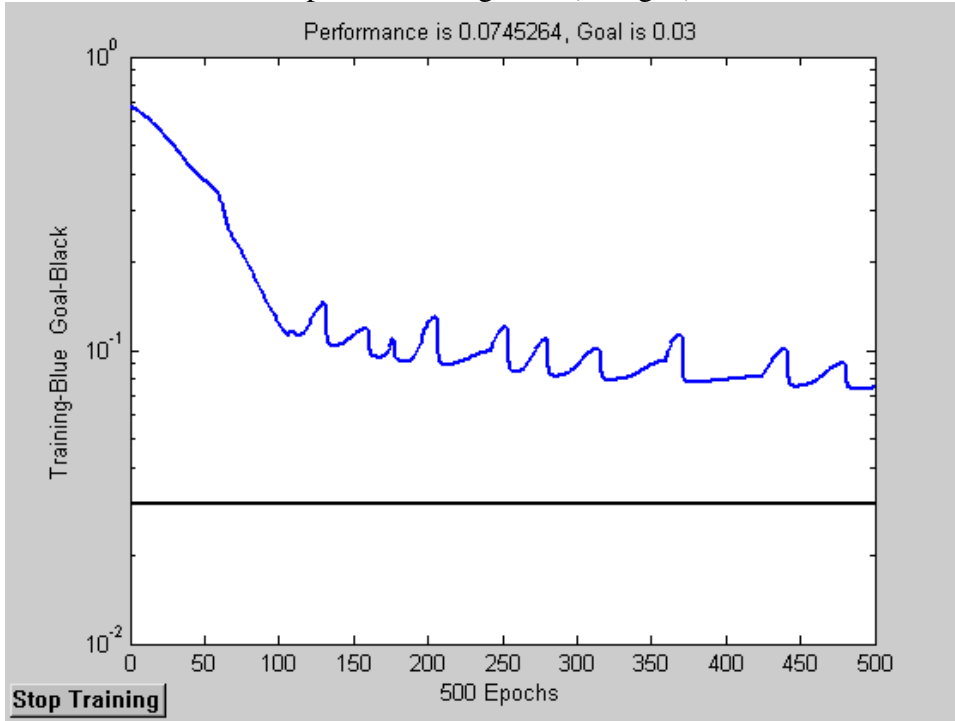
Networks were trained using a variety of learning algorithms, and their performance was evaluated in terms of elapsed training time, and lowest achieved RMS.

Resilient Backprop (trainrp):



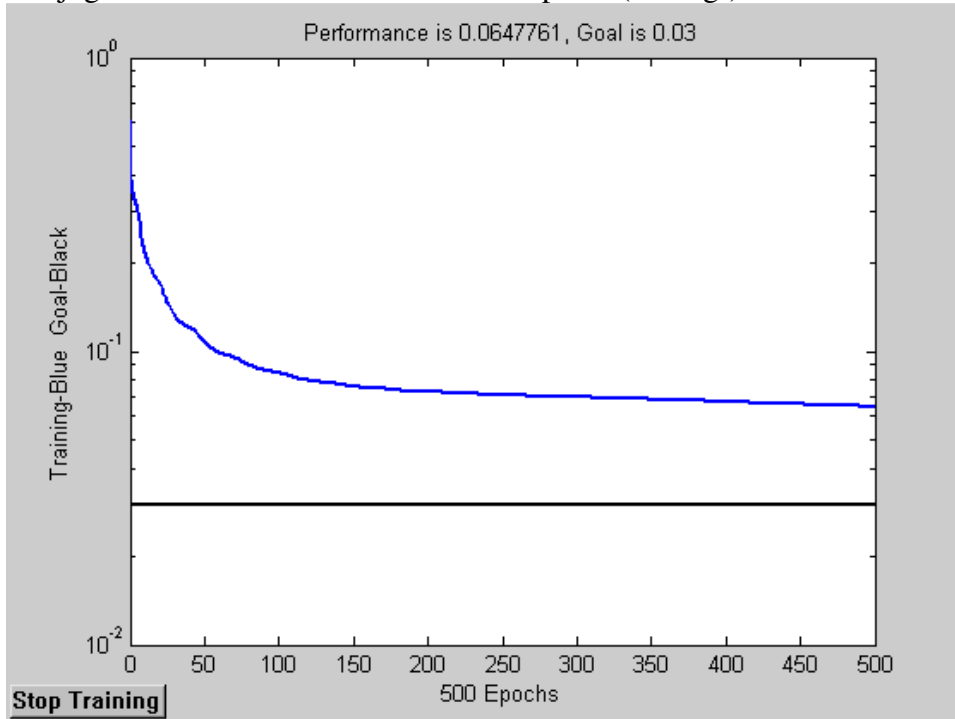
Elapsed Time: 88.141 seconds.

Gradient Descent / Adaptive Learning Rate (trainгда):



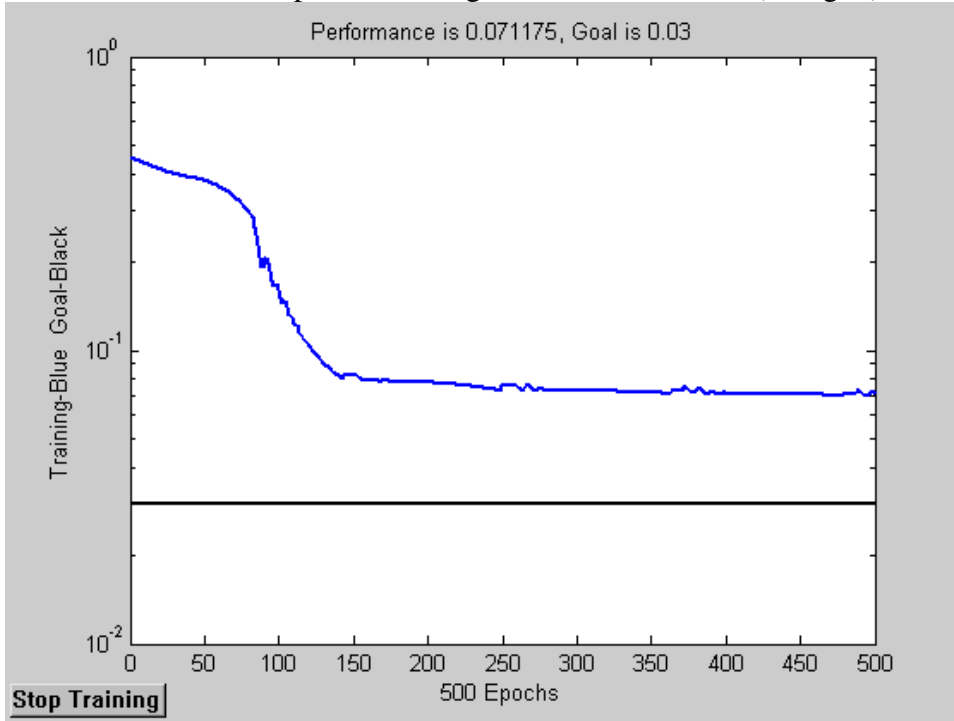
Elapsed Time: 96.844 seconds

Conjugate Gradient / Fletcher – Reeves Update (traincgr):



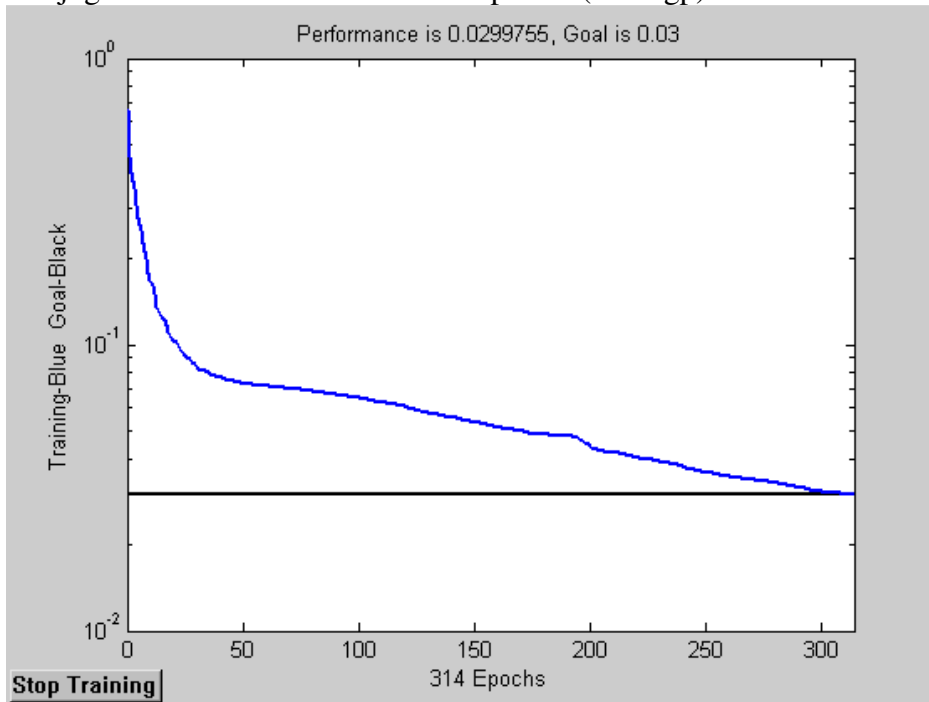
Elapsed Time: 128.453 seconds

Gradient Decent / Adaptive Learning Rate w/ Momentum (traingdx) :



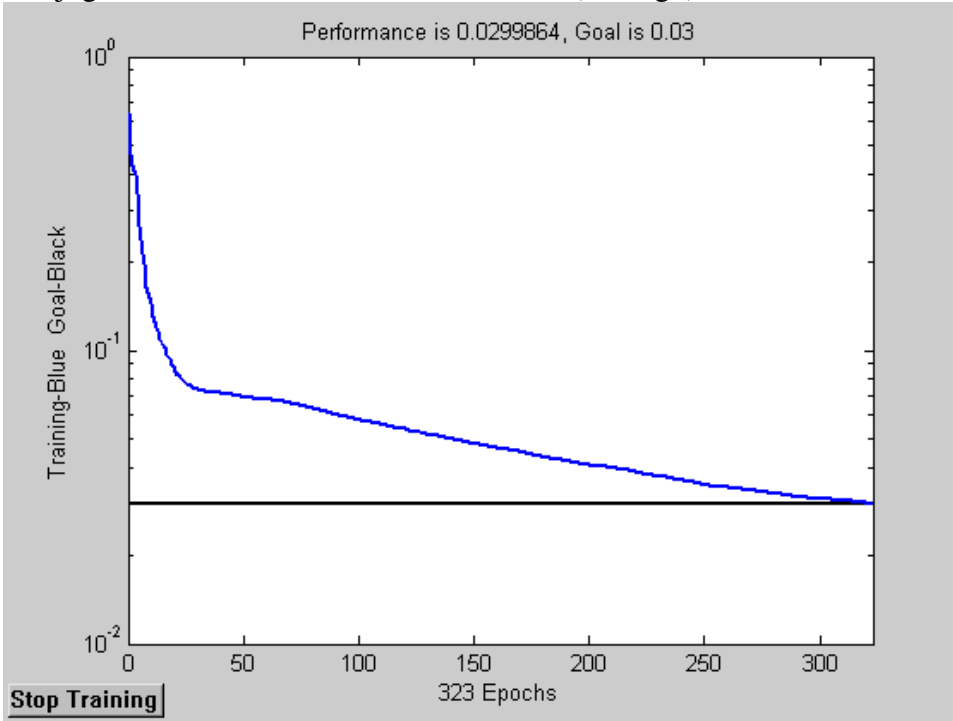
Elapsed Time: 99.047 seconds

Conjugate Gradient / Polak-Ribiere Updates (traincgp) :



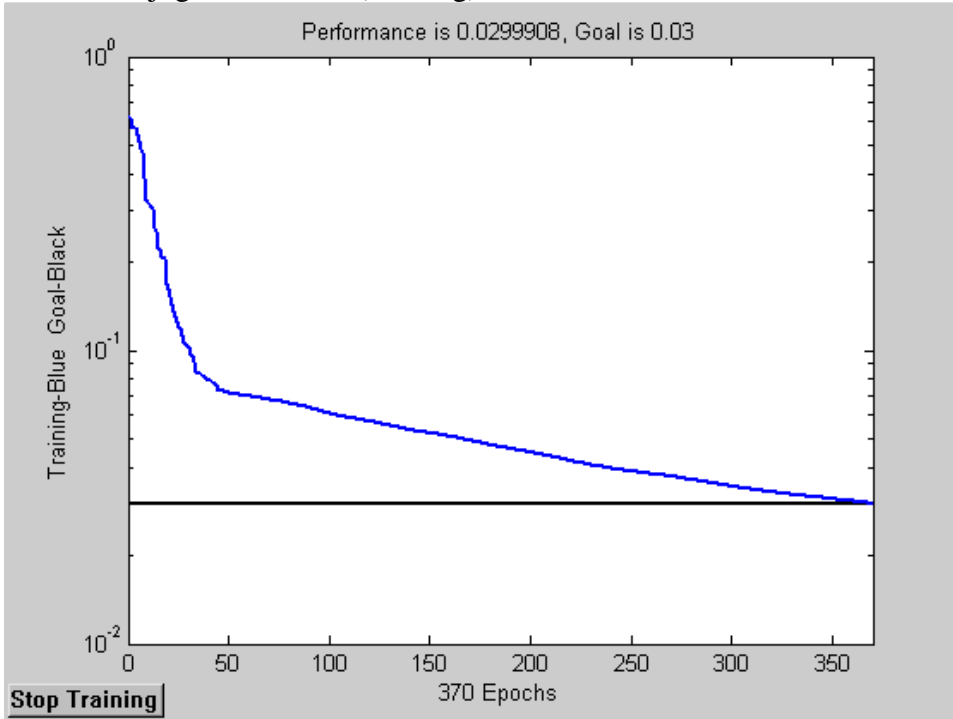
Elapsed Time: 115.984 seconds

Conjugate Gradient / Powell-Beale Restarts (traincgb) :



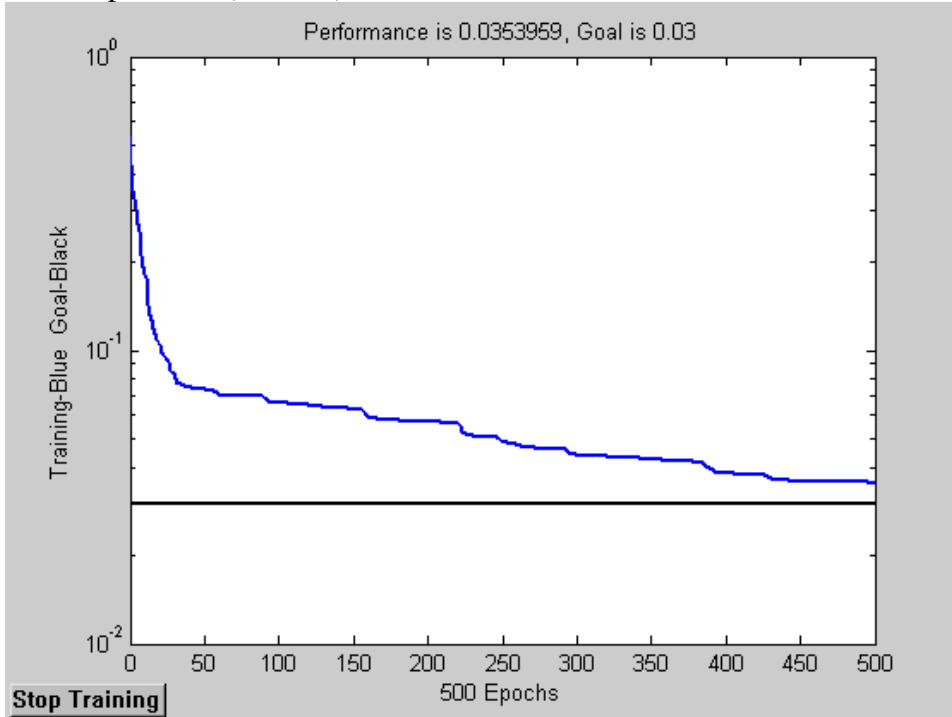
Elapsed Time: 134.015 seconds

Scaled Conjugate Gradient (trainscg) :



Elapsed Time: 121.844 seconds

One Step Secant (trainoss) :



Elapsed Time: 155.469 seconds

**** Due to shortage of processing capability, the Levenberg-Marquardt training algorithm was not tested.**

Upon review of the above results, the Conjugate Gradient method with Polak-Ribiere updates (traincgp) was selected for further experimentation because it achieved the lowest RMS in the least amount of time.

Further training was conducted with 1000 epochs, and the RMS dropping to 0.015. The network was then evaluated by presenting its inputs with examples from its learning set, and new examples not in its learning set.

The network has difficulty correctly identifying the less frequent cases in the example set, such as both types of articles and conjugates. However, the network is very good at classifying simple plural forms of nouns. Otherwise, examples outside of the training set stand a random chance of being correctly classified.

Conclusions:

For the purpose of simple classification such as above, the limitations of the network such as incorrect classification, and the need for all possible words to be presented and learned, and the large amount of resources consumed makes the use of a neural network undesirable to the alternative of using a simple lookup table. However, the next stage – identification of subject and predicate, would more likely be a good

application for neural classification, as the rules for subjects and predicates are much less firm than that of word classifications.