Classification of Thiamine-Repressible Promoters Using a Feed-forward Backpropagation Neural Network

Jacqueline Lam

Supervisors: Professor Parvin Mousavi and Professor Paul G. Young

School of Computing, Queen’s University, Kingston, ON, K7L 3N6

1. Introduction and Objectives

- nmt1, nmt2, bsa1, pho4, thi4 and thi9 genes in Schizosaccharomyces pombe are essential components of the thiamine biosynthesis pathway (Figure 1)
- Promoters of the genes are thiamine-repressible/thiamine-regulated
- Thi1 is a Cys6 zinc finger transcription factor that positively regulates the expression of all known thiamine-repressible genes
- Thi5, another Cys6 zinc finger protein, also regulates the nmt1 promoter and may work to regulate the expression of thi1 as well
- The sequences for the promoter regions of the thiamine-repressible genes in S. pombe, Schizosaccharomyces japonicus and Schizosaccharomyces octosporus are presumed to be recognized by Thi1 and to thus have some sequence similarities
- The objective of this study is to use a feed-forward backpropagation (FFBP) neural network that may learn to differentiate thiamine-repressible promoter sequences from random, non-thiamine-repressible sequences.

2. Methodology

2.1 Data

- Used sequences from S. pombe, S. japonicus and S. octosporus

2.2 Feed-forward backpropagation neural network

- Each nucleotide is represented as a set of four binary digits (e.g., A=0100; G=0010; C=0001; and T=1000)
- The feed-forward network consists of two layers - hidden and output layers (Figure 2)
- Weighted inputs are calculated by applying the dot product weight function to inputs
- The net input function calculates the layer’s net input by combining its weighted inputs and biases
- Transfer functions transform the output of the layer
- The first and second layers each have a weight as input from the previous layer
- Each layer's weights and biases are initialized by random

2.3 Training of the feed-forward backpropagation neural network using the conjugant gradient method. The neural network was trained 20 times.

3. Results

- Motifs that were consistently classified incorrectly were removed from the dataset
- The best results were achieved when the conjugate gradient method was used for training the neural network
- The neural network was trained 20 times (Figure 3), for 100 cycles
- The input dataset was divided into two-thirds training and one-third validation
- The size of the hidden layer was one neuron
- The best performance of the neural network was 83% accuracy on the validation dataset
- Figure 4 shows which positions in the motifs are significant in classification

4. Conclusion

- The highly conserved DNA binding motifs in thiamine-regulated promoters for the zinc finger transcription factors Thi1 and Thi5 have not yet been identified
- The FFBP neural network can differentiate between thiamine-repressible and non-thiamine-repressible promoter sequences
- The regulatory sequences of the thiamine-repressible promoters may be one of the 30 motifs that were classified correctly by the neural network
- One of the limitations in this study was the small number of known thiamine-regulated genes in the dataset
- A further exploration of this study would be to use longer motifs and to train the neural network with known non-thiamine-repressible promoter sequences