Distributed Mentor Project Final Report

On June 3rd 2002, I traveled from my suburban home in Staten Island, New York to a rural town in North Dartmouth, Massachusetts. I was one of thirty-five young ladies chosen to participate in a 10-week Mentor Project sponsored by the Computing Research Association to increase the number of females in the computer science field. I had the honor of working with Professor Iren Valova at the University of Massachusetts Dartmouth. The project focused on neural networking and the different types of algorithms used in specifically pattern recognition.

I arrived at the school and was shown to room. I lived in the Cedar Dell South Apartments, which I shared with five other girls. This was the first time I was on my own. I immediately hooked up my laptop so that I could begin working as soon as possible. I met with Professor Valova and we discussed the plan that we would follow for the duration of the project. The first two weeks of the project focused on Adaptive Resonance Theory and its applications. The next three weeks of the project focused on Self Organizing Map Algorithms. The last five weeks were pertinent to the whole idea of how well self-organizing maps are in recognizing and visualizing patterns specifically protein sequences.

From June 3rd-June 10th I continuously researched Adaptive Resonance Theory and its applications. I started out with researching broad areas of applying the ART-1 algorithm. However, I found a vast amount of information on using ART-1 for diagnostic purposes. Through my research I found that ART-1 is successful in diagnosing early symptoms of schizophrenia through the brain waves of the patient. ART-1 is also used in recognizing certain patterns of behavior in schizophrenics. ART-1 has been very useful in showing damaged brain circuits in patients with Parkinson's disease and other diseases related to neuronal degeneration. I wrote a 6-page paper on Adaptive Resonance Theory and its application to medical diagnoses. I used Internet databases such as Lexis-Nexis and Ebsco to research articles, journals and other reference guides. I also used three neural network textbooks to fully capture the architecture and algorithm of ART. I used a total of 13 references to write this paper.

On June 11th I started to design a program that would use ART to recognize letters. I went to a freeware site and studied different codes that are written in MATLAB to accomplish this task. On June 17th I finished my program and the results were successful. I designed the program so that any matrix size can be used for the program to recognize the letter. I also designed the program so that it not only prints out the final weight matrix but also each letter in a graphical pattern.

On June 18th Professor Valova and I discussed the second part of the project. I researched all of the possible algorithms used for Self-Organizing Maps and decided which was best. I studied the classic algorithm carefully and then studied how the classic algorithm could be improved. I found that the Growing Hierarchical Self Organizing Map provided the best clustering results however the Bayesian Algorithm converged the fastest. On June 25th I found more algorithms that use kernel based functions during the initialization process. Self Organizing Map Algorithms grew from one to the other. The classic algorithm is still preserved in architecture while the mathematical functions are calculated using new techniques. Through my research I concluded that an algorithm

using both Bayesian learning and layering would prove very efficient for clustering large amounts of data. On June 28th I finished my paper on Self Organizing Map Algorithms. The 10- page paper consisted of 24 references all strictly dealing with the different algorithms used for Self-Organization.

On July 1st Professor Valova and I discussed the next project. For the next five weeks I researched proteins, amino acid sequences and Self Organizing Feature Maps. For the first week I researched if soms could be used to predict the shape of the proteins. I found that it would be too difficult to accomplish in five weeks. On July 8th I found a few papers that showed how the shape of a protein could be predicted from the amino acid sequence. On July 9th I downloaded the SOM Toolbox for MATLAB to see if I could simply differentiate between two different proteins based on their amino acid sequence. On July 15th I gathered all of my references to write a 5- page paper on protein shape prediction using Self Organizing Maps.

On July 16th I analyzed protein databases and obtained the amino acid sequence for myoglobin. On July 18th I wrote a C++ file that converted the amino acid letters into ASCII characters. These ASCII characters were used as the training pattern for the program. The amino acid sequence for myoglobin is 154 amino acid residues long. I used three different encoding techniques for training the SOM. The first encoding was the ASCII characters. The second was taking the ASCII characters as two separate numbers and using them as the x-coordinate and y-coordinate. On July 24th I improved the Self-Organizing Map in order to capture the entire data set. On July 26th I began writing my paper on SOM and protein analysis. I discovered that how the amino acids are encoded is significant to how successful the SOM will be. From July 27th- July 29th I worked meticulously on my paper and code. I increased the map sizes to 15 x 15 and realized that if the map size is too large the graphical representation of the protein is inaccurate. Therefore 15 x 15 turned out to be an appropriate size. I also increased the number of training epochs in order to allow the map to self organize properly. The program I designed was successful in recognizing proteins.

From July 30th –August 6th I worked on a presentation for the Freshman Summer Institute College of Engineering on “Human Vision Modeling”. I researched facial recognition and self -organizing maps. The faces are presented to the network by way of eigenvalues and the map recognizes the grayscale pattern set by the eigenvalues of the facial space. I used the Internet as my research tool.

From August 7th-August 9th I continued working on my SOM and protein analysis program. The program was a 2-Dimensional map; I captured the data more accurately using a 1-Dimensional map.

In addition to working on the project I updated my webpage designed specifically for the project regularly and included Power Point Presentations and three 6-10 page reports on Adaptive Resonance Theory and Self Organizing Maps. My webpage also contains my results from my programs and daily log entries. Though the project has ended I will continue to work on 1-Dimensional Self Organizing Maps and protein recognition.