

The proposal for the Senior Design Project

Automated Annotation of Amino Acid Sequences Using Hierarchical Motif Vectors

Student Name, Number: Mert Ekren, 200206014

Objective : This project aims to accomplish identifying clusters of hierarchical motif vectors from the DNA-binding human proteins and labeling the responsible cluster with regards to the obtained motif vectors.

Motivation: The project roots from developing a statistical learning algorithm that can manage to predict functional regions of all amino acid sequences based on experimentally verified sites. Hierarchical motif vectors achieve higher predictive performance compare to other algorithms in the literature. Here, the motivation of clustering the motif vectors and labeling the corresponding cluster is to provide analytic approach for hierarchical motif vectors.

Specific Aims:

Specific Aim 1: Construction of motif vectors and researching cluster analysis algorithms.

Specific Aim 2: Testing the clustering algorithms and determining a specific algorithm.

Specific Aim 3: Finalizing the project by clustering the motifs, and labeling the amino acid sequences in terms of the obtained motif vectors.

Table 1: TimeTable of the Project

Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13	Week 14
			Initiation										
			Specific Aim 1										
						Specific Aim 2							
										Specific Aim 3			
			Progress Report			Progress Report				Final Report, Poster			