ABSTRACT

The role of Probability from observation is important for analyzing biological based problems. Using HMM we estimate different factors in molecular biology. They are protein sequence, RNA, CpG Islands, Protein and DNA.

To estimate probability distribution over discrete alphabets from observation, we use a mixture model of common ancestors. Also to estimate the amino acids in common ancestors, we have used Viterbi and Baum Welch Algorithm.

Insertion, deletion and substitution for RNA sequence are described by a phylogenetic tree. In this, we use a multiplicative, Quasi stationary, non reversible, continuous time Markov process and Felsenstein's Peeling Algorithm. This method is applied to any probabilistic model of sequence alignment.

CpG is the region of DNA. To detect the average length of a CpG islands and average length between two CpG islands, we use Veterbi method, Forward method, Backward method and Baum Welch Algorithm. Among the various methods, we presume Baum Welch method is the best one.

Probability of protein sequence can be found by multiplying the Probabilities of amino acids in the sequence. HMM is used as a tool in the study of protein molecule. Using this tool, we obtain a score detail about
protein molecules. If it is a high score, the sequence of the interest will be a member of the class, and if it is a low score, it will not be a member of the class.

HMM is used to analyze the protein sequence. Since Hidden Markov Models are more sensitive to perturbations in the emission probabilities than to perturbations in the transition probability matrix hence the initial distribution of underlying Markov Chain.

Pair Hidden Markov models (PHMMs) can be applied to alignment problems. Generalized Hidden Markov models (GHMMs) can be applied to gene finding problems. We combine both the models known as Generalized Pair Hidden Markov models (GPHMMs) that can be applied to alignment and gene finding problems by using Baum Welch algorithm.