Difference between Edit distance, Hidden Markov model and Probabilistic Suffix Tree

All of the 3 methods can be used as a basis to cluster sequential data.

Edit Distance: It is a similarity measure which takes place of only optimal global alignment whereas it discards the local alignments (which can be an important feature as well). The computation of the edit distance with block operations is a N-P hard problem. It is an inefficient method.

Hidden Markov model: It is a statistical model, in which the states are not visible but the output of the states is visible. Each state has a probability distribution assigned over the output tokens.

Probabilistic Suffix tree: It is an improvement over suffix tree, in which all the nodes have a probability vector to store the probability distribution of the next symbol when the preceding segment is provided as a label.

Advantage

PST is very efficient in identifying the cluster of sequential data. Edit distance and Hidden markov model, both are computationally very expensive models. In case of limited memory scenario, the tree can be pruned according to the available memory which is not possible in other cases. The accuracy of PST was also better than the other 2 methods.