

Introduction

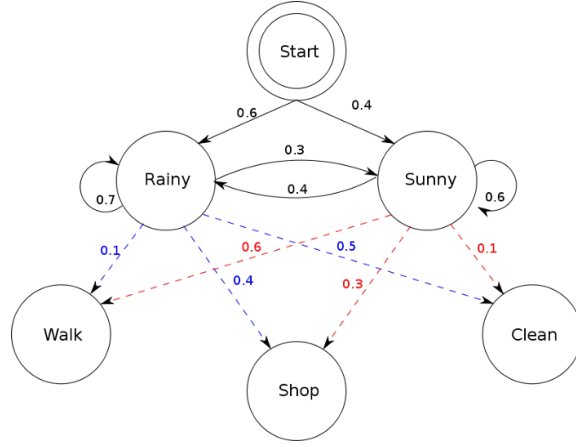
Hidden Markov Model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (i.e. hidden) states.

Hidden Markov models are especially known for their application in reinforcement learning and temporal pattern recognition such as speech, handwriting, gesture recognition, part-of-speech tagging, musical score following, partial discharges and bioinformatics.

Biological sequence analysis is frequently as simple as affixing the appropriate label to each residue. We want to label nucleotides as exons, introns, or intergenic sequences in gene identification. We want to associate residues in a query sequence with homologous residues in a target database sequence in sequence alignment.

Terminology in HMM

For example, the term "hidden" refers to the first-order Markov process that lies behind the observation in question. An observation is any data that we are aware of or that we can observe. These HIDDEN STATES can be seen in the following diagram, where "Rainy" interacts with "Sunny".



As shown in the diagram, OBSERVATIONS refers to the words "Walk", "Shop", and "Clean". Our observations are our training data in machine learning, and the number of hidden states is our hyper parameter for our model. The model's evaluation will be discussed in a future post.

T = length of the observation sequence

N = number of states in the model

M = number of observation symbols

Q = {

$$q_0, q_1, \dots, q_{N-1}$$

} = distinct states of the Markov process

V = { 0, 1,, M - 1 } = set of possible observations

A = state transition probabilities

B = observation probability matrix

Π = initial state distribution

O = (

$$O_0, O_1, \dots, O_{T-1}$$

) = observation sequence.