Predicting User Behavior using Clickstream Data with Hidden Markov Models

Surya Anish Nadupalli
Jawaharlal Nehru Technological University, Hyderabad, Telangana, India

Shyama Vasa
Jawaharlal Nehru Technological University, Hyderabad, Telangana, India

Abstract

A Hidden Markov Model (HMM) is a type of machine learning model that helps understand and predict data that occurs in a sequence, such as speech, text or biological data. It works by assuming that there are hidden states that are not directly visible, but which influence the observable data we see. The model uses probabilities to determine how states change over time, following the Markov process, and how each state leads to observable outcomes. Key algorithms such as the Forward–Backward method help in learning the model, while the Viterbi algorithm is used to find the most likely sequence of hidden states. HMMs are widely used in areas like speech recognition, natural language processing, and bioinformatics. Although deep learning methods are currently very popular, HMMs remain important because they are easier to understand, require less data, and serve as a foundation for many modern sequence models.

Keywords

Hidden Markov Model, Forward-Backward method, Viterbi algorithm, deep learning, Baum-Welch/EM algorithm.