Abstract

Information on the World Wide Web is increasing tremendously. To get the relevant information from very large data sets is essential. In traditional retrieval systems, the query is given to large corpus to retrieve the relevant documents. The traditional models for information retrieval are just one subclass of retrieval techniques that have been studied in many years. Although many techniques share common characteristics in the information retrieval hierarchy, they all share a core set of similarities that justify their own class and these algorithms are design for isolated datasets. But in most of cases, relationships among different datasets are always existed. A new probabilistic Hidden Markov model is proposed and based on this model new information retrieval (IR) technique is presented. Hidden Markov models (HMMs) are widely used in science, engineering and many other areas. In a HMM, there are two types of states like hidden states and observable states. HMM is powerful modeling of context as well as
the current observations. Hidden Markov model is finite state machine which offer a good balance between simplicity and expressiveness of context. IR is performed by determining the sequence of states that was most likely to have generated the entire document, and retrieving the information that were associated with certain designated target states. Determining this sequence is efficiently performed by dynamic programming with the Viterbi algorithm.

Reference


Index Terms

Computer Science  Information Retrieval

Key words

Information Retrieval  Statistical model  Hidden Markov Model  Viterbi algorithm