Available online at www.elixirpublishers.com (Elixir International Journal)

Computer Science and Engineering

Elixir Comp. Sci. & Engg. 56A (2013) 13530-13533

ISSN: 2229-712X

Hidden Markov Model as Classifier: A survey

Sanjay Shimpi and Vijay Patil

Department of Computer Engineering, Vidyalankar Polytechnic, Wadala, Mumbai-37.

ARTICLE INFO

Article history:

Received: 9 July 2013;

Received in revised form:

11 March 2013;

Accepted: 18 March 2013;

Keywords

HMM,

Markov Chain,

Probability.

ABSTRACT

This paper summarizes the introduction and importance of hidden markov model (HMM) as a classifier, learning and classification. A Markov process is a particular case of stochastic process, where the state at every time belongs to a finite set, the evolution occurs in a discrete time and the probability distribution of a state at a given time is explicitly dependent only on the last states and not on all the others. In this survey we present details of hmm, its mathematical foundations, advantages and applications in the field recognition.

1. Introduction

During the last decade, hidden Markov models (HMMs), which can be thought of as a generalization of dynamic programming techniques [1].

A Hidden Markov Models (HMM) is a finite state machine which has some fixed number of states. It provides a probabilistic framework for modeling a time series of multivariate observations.

Hidden Markov models were introduced in the beginning of the 1970's as a tool in speech recognition. This model based on statistical methods has become increasingly popular in the last several years due to its strong mathematical structure and theoretical basis for use in a wide range of applications [2].

A hidden Markov model is a doubly stochastic process, with an underlying stochastic process that is not observable (hence the word hidden), but can be observed through another stochastic process that produces the sequence of observations. The hidden process consists of a set of states connected to each other by transitions with probabilities, while the observed process consists of a set of outputs or observations, each of which may be emitted by each state according to some probability density function (pdf). Depending on the nature of this pdf, several HMM classes can be distinguished. If the observations are naturally discrete or quantized using vector quantization [1].

The Hidden Markov Model is a finite set of states, each of which is associated with a (generally multidimensional) probability distribution Transitions among the states are governed by a set of probabilities called transition probabilities. In a particular state an outcome or observation can be generated, according to the associated probability distribution. It is only the outcome, not the state visible to an external observer and therefore states are "hidden" to the outside; hence the name Hidden Markov Model [3].

A hidden Markov model can be considered a generalization of a mixture model where the hidden variables (or latent variables), which control the mixture component to be selected for each observation, are related through a Markov process rather than independent of each other.

Tele:

E-mail addresses: vijay.patil@vpt.edu.in

A hidden Markov model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (*hidden*) states. An HMM can be considered as the simplest dynamic Bayesian network. In a *hidden* Markov model, the state is not directly visible, but output, dependent on the state, is visible. Each state has a probability distribution over the possible output tokens. Therefore the sequence of tokens generated by an HMM gives some information about the sequence of states.

2. Brief introduction to HMM

An introduction to hidden markov model is as fallows. The diagram of a typical Markov models with 5 states in shown figure.1.

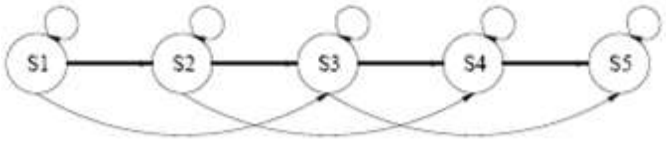

Figure 1: Markov Model with 5 states

Such a model is called a left-right model with a start state S1 and a legal end state S5. The like hood of transition between states is governed by a state transition matrix given in figure 2. as follow:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & 0 & 0 \\ 0 & a_{2,2} & a_{2,3} & a_{2,4} & 0 \\ 0 & 0 & a_{3,3} & a_{3,4} & a_{3,5} \\ 0 & 0 & 0 & a_{4,4} & a_{4,5} \\ 0 & 0 & 0 & 0 & a_{5,5} \end{pmatrix}$$

Figure 2: State Transition Matrixes

The variable $a_{i,\ j}$ represents the like hood of making a transition from state Si to state Sj. According to the particular structure chosen, some transitions are not permitted and therefore have a like hood of 0. For example, in the model of figure 2.3., the direct transition from S1 to S4 is not permitted and consequently a1, 4 is set to 0.

The word 'hidden' in "Hidden Markov Model" has much the same meaning as the word hidden when referring to hidden "hidden layers" in an artificial neural network. Just as the single layer neural network needs a second 'hidden' layer of weights in order to be able to model complex functions, the Markov chain needs a second 'hidden' random process in order to model complex observations. A Hidden Markov Model is therefore sometimes represented as shown in Figure.3. as follows:

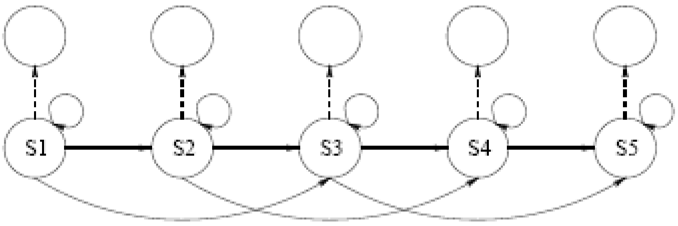

Figure: 3 Hidden Markov Model with 5 states

The second process is represented by an observation probability matrix given in figure 4 as Follow:

$$B = \begin{pmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,M-1} & b_{1,M} \\ & \cdots & & \cdots & \\ & & b_{i,m} & & \\ & \cdots & & \cdots & \\ b_{5,1} & b_{5,2} & \cdots & b_{5,M-1} & b_{5,M} \end{pmatrix}$$

Figure 4: Observation probability matrix

The term $b_{i,m}$ gives the probability of producing (observing) the discrete symbol (codeword) m when being in state s_{i}..

An HMM should be seen as a "Finite state machine" capable of generating observation string Starting with state S1. The model generates a symbol based on the probabilities $b_{i,\ m.}$ then a transition is made to another state based on the transition probabilities $a_{i,\ j.}$ such a mechanism is repeated until the final state (S5 in figure.2.5) is reached. Upon reaching S5, it is still possible to generate an infinitely long sequence of symbols as the transition from S5 to itself is allowed. It should be noted that even though we view the HMM as a generator, our goal is to use it as recognizer [4].

The Hidden Markov Model is characterized by the following

1) Number of states in the model
2) Number of observation symbols
3) State transition probabilities
4)Observation emission probability distribution that characterizes each state
5) Initial state distribution

3. Architecture of a hidden Markov model

The diagram below shows the general architecture of an instantiated HMM. Each oval shape represents a random variable that can adopt any of a number of values. The random variable *x*(*t*) is the hidden state at time *t* (with the model from The above diagram, x(f) Є {x1, x2, x3}). The random variable Y(f) is the observation at time t (with y(f) Є {y1, y2, y3, y4}). The arrows in the diagram (often called a trellis diagram) denote conditional dependencies as shown in figure 5.

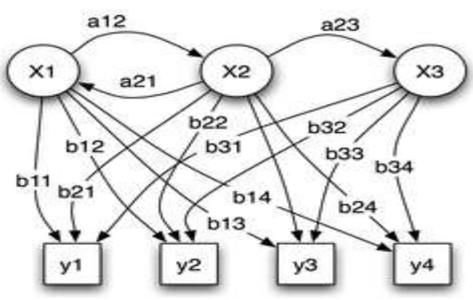

Figure 5: Probabilistic parameters of a hidden Markov model

x - states- possible observations, *a*- state transition probabilities, *b*- output probabilities

From the diagram, it is clear that the conditional probability distribution of the hidden variable *x*(*t*) at time *t*, given the values of the hidden variable *x* at all times, depends *only* on the value of the hidden variable $x(t-1)$: the values at time $t-2$ and before have no influence. This is called the Markov property. Similarly, the value of the observed variable *y*(*t*) only depends on the value of the hidden variable *x*(*t*) (both at time *t*). The parameters of a hidden Markov model are of two types, *transition probabilities* and *emission probabilities* (also known as *output probabilities*). The transition probabilities control the way the hidden state at time *t* is chosen given the hidden state at time *t* – 1[7].

Following notations will be used regarding HMM

N = number of states in the model

M = number of distinct observation symbols per state (Observation symbols correspond to the physical output of the system being modeled)

T = length of observation sequence

O = observation sequence, i.e., O1, O2, O3,………OT

Q = state sequence q1, q2, ……., qT in the Markov model

A = {aij} transition matrix, where aij represents the transition probability from state i to state j

B = {bj(Ot)} observation emission matrix, where bj(Ot) represent the probability of observing Ot at state j

π = {πi} the prior probability, where πi represent the probability of being in state i at the beginning of the experiment, i.e., at time t = 1

λ = (A, B, π) the overall HMM model [2].

4. Markov Chain

In mathematics, a Markov chain, named after Andrey Markov, is a stochastic process with the Markov property. Having the Markov property means that, given the present state, future states are independent of the past states. In other words, the description of the present state fully captures all the information that could influence the future evolution of the process. Being a stochastic process means that all state transitions are probabilistic [5].

At each step the system may change its state from the current state to another state (or remain in the same state) according to a probability distribution. The changes of state are called transitions, and the probabilities associated with various state-changes are called transition probabilities. An example of a Markov chain is a random walk on the number line which starts at zero and transitions +1 or -1 with equal probability at each step.

Formal definition of Markov chain: A Markov chain is a sequence of random variables X_1, X_2, X_3, ... with the Markov property, namely that, given the present state, the future and past states are independent.

Formally, set of the possible values of X_i form a countable set *S* called the state space of the chain. Markov chains are often described by a directed graph, where the edges are labeled by the probabilities of going from one state to the other states.

There are two variations in Markov chain, Continuous-time Markov processes have a continuous index and Time-homogeneous Markov chains (or, Markov chains with time-homogeneous transition probabilities) are processes where index varies according to time for all *n*.

Markov chains display some important properties, [5] those are as follows:

- **Reducibility:** A Markov chain is said to be irreducible if its state space is a single communicating class; in other words, if it is possible to get to any state from any state.
- **Periodicity:** A state *i* have period *k* if any return to state *i* must occur in multiples of *k* time steps. If $k = 1$, then the state is said to be aperiodic; otherwise (k>1), the state is said to be periodic with period *k*. It can be shown that every state in a communicating class must have the same period.
- **Recurrence:** A state *i* is said to be transient if, given that we start in state *i*, there is a non-zero probability that we will never return to *i*. If a state *i* is not transient (it has finite hitting time with probability 1), then it is said to be recurrent or persistent.
- **Ergodicity:** A state *i* is said to be ergodic if it is aperiodic and positive recurrent. If all states in a Markov chain are ergodic, then the chain is said to be ergodic. It can be shown that a finite state irreducible Markov chain is ergodic if its states are aperiodic.

5. Three basic problems with HMM

To work with HMM, the following three fundamental questions should be resolved

a) Given the model $\lambda= (A, B, \pi)$ how do we compute $P(O|\lambda)$, the probability of occurrence of the observation sequence O = O1,O2, ….. , OT.

b) Given the observation sequence O and a model λ, how do we choose a state sequence q1 , q2 , ….. , qT that best explains the observations.

c) Given the observation sequence O and a space of models found by varying the model parameters A, B and π, how do we find the model that best explains the observed data.

There are established algorithms to solve the above questions. In our task we have used the forward-backward algorithm to compute the P (O| λ), Viterbi algorithm to resolve problem b, and Baum-Welch algorithm to train the HMM.

6. Solution to three basic problems in HMM

Once a system can be described as a HMM, three problems can be solved. The first two are pattern recognition problems: Finding the probability of an observed sequence given a HMM (evaluation); and finding the sequence of hidden states that most probably generated an observed sequence (decoding). The third problem is generating a HMM given a sequence of observations (learning) [5].

Problem-1: Evaluation

Consider the problem where we have a number of HMM s (that is, a set of (∏, A, B) triples) describing different systems, and a sequence of observations. We may want to know which HMM most probably generated the given sequence. Use the forward algorithm to calculate the probability of an observation sequence given a particular HMM, and hence choose the most probable HMM.

This type of problem occurs in speech recognition where a large number of Markov models will be used, each one modeling a particular word. An observation sequence is formed from a spoken word, and this word is recognized by identifying the most probable HMM for the observations.

Problem-2: Decoding

Finding the most probable sequence of hidden states gives some observations. Another related problem, and the one usually of most interest, is to find the hidden states that generated the observed output. In many cases we are interested in the hidden states of the model since they represent something of value that is not directly observable. We use the Viterbi algorithm to determine the most probable sequence of hidden states given a sequence of observations and a HMM. [5, 6].

Problem-3: Learning

Generating a HMM from a sequence of observations. The third, and much the hardest, problem associated with HMMs is to take a sequence of observations (from a known set), known to represent a set of hidden states, and fit the most probable HMM; that is, determine the (π,A,B) triple that most probably describes what is seen.

7. The advantage of HMM

Hidden Markov models are used for find the states for which a given stochastic process went through. In order to use a markov chain, the process must depend only in its last state.

a) HMM has strong statistical foundation

b) It is able to handle new data robustly

c) Computationally efficient to develop and evaluate (due to the existence of established training algorithms).

d) It is able to predict similar patterns efficiently [2].

8. Applications of hidden Markov models

HMM-based applications are common in various areas such as speech recognition, bioinformatics, and genomics. In recent years, Joshi and Phoba [9] have investigated the capabilities of HMM in anomaly detection. They classify TCP network traffic as an attack or normal using HMM. Cho and Park [10] suggest an HMM-based intrusion detection system that improves the modeling time and performance by considering only the privilege transition flows based on the domain knowledge of attacks. Ourston et al. [11] have proposed the application of HMM in detecting multistage network attacks. Hoang et al. [12] present a new method to process sequences of system calls for anomaly detection using HMM. The key idea is to build a multilayer model of program behaviors based on both HMMs and enumerating methods for anomaly detection. Lane [13] has used HMM to model human behavior. Once human behavior is correctly modeled, any detected deviation is a cause for concern since an attacker is not expected to have a behavior similar to the genuine user. Hence, an alarm is raised in case of any deviation.

HMMs can be applied in many fields where the goal is to recover a data sequence that is not immediately observable (but other data that depends on the sequence is). Applications include:

- Cryptanalysis
- Speech recognition
- Speech synthesis
- Part-of-speech tagging
- Machine translation
- Partial discharge
- Gene prediction
- Alignment of bio-sequences
- Activity recognition
- Protein folding
- Pattern recognition

9. Conclusion

This paper has presented a survey on hidden Markov models. Herein, a list of properties was subjected to a new HMM definition, and it was found that HMMs are extremely powerful, given enough hidden states and sufficiently rich observation distributions. Moreover, even though HMMs encompass a rich class of variable length probability distributions, for the purposes of classification, they need not precisely represent the true conditional distribution. Even if a specific HMM only crudely reflects the nature of a speech signal, there might not be any detriment to their use in the recognition task, where a model need only internalize the distinct attributes of its class.

10. References:

[1] A. El-Yacoubi, M. Gilloux,R. Sabourin, Member, IEEE, and C.Y. Suen, Fellow, IEEE , An HMM-Based Approach for Off-Line Unconstrained Handwritten Word Modeling and Recognition, IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 21, No. 8, Auguest 1999.

[2] Md. Rafiul Hassan and Baikunth Nath, StockMarket Forecasting Using Hidden Markov Model: A New Approach, Proceedings of the 2005 5th International Conference on Intelligent Systems Design and Applications (ISDA'05)0-7695-2286-06/05 $20.00 © 2005 IEEE.

[3] Mohit Mehta, Member, IACSIT, Rupesh Sanchati and Ajay Marchya , Automatic Cheque Processing System , International Journal of Computer and Electrical Engineering, Vol. 2, No. 4, August, 2010 1793-8163

[4]Didier Guillevic, Ching Y .Suen, HMM Word Recognition Engine, CENPARMI, Suite GM-606 Concordia University.

[5] Lawrence R. Rabiner, A Tutorial on Hidden Markov Model and Selected Applications in Speech Recognition, Proceedings of the IEEE, vol. 77, no.2, February 1989.

[6] A. L. Koerich, R. Sabourin, and C. Y. Suen. Fast two–level viterbi search algorithm for unconstrained handwriting recognition. In *Proc. 27th International Conference on Acoustics, Speech, and Signal Processing*, Orlando, USA, 2002. To appear.

[7] Cheng-Lin Liu, Hiromichi Fujisawa, Classification and Learning for Character Recognition: Comparison of Methods and Remaining Problems, National Laboratory of Pattern Recognition (NLPR) Institute of Automation ,Chinese Academy of Sciences P.O. Box 2728, Beijing 100080, P.R. China.

[8] Abhinav Srivastava, Amlan Kundu, Shamik Sural, Senior Member, IEEE, and Arun K. Majumdar, Senior Member, IEEE, Credit Card Fraud Detection Using Hidden Markov Model, IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 5, NO. 1, JANUARY-MARCH 2008

[9] S.S. Joshi and V.V. Phoha, "Investigating Hidden Markov Models Capabilities in Anomaly Detection," Proc. 43rd ACM Ann. Southeast Regional Conf., vol. 1, pp. 98-103, 2005.

[10] S.B. Cho and H.J. Park, "Efficient Anomaly Detection by Modeling Privilege Flows Using Hidden Markov Model," Computer and Security, vol. 22, no. 1, pp. 45-55, 2003.

[11] D. Ourston, S. Matzner, W. Stump, and B. Hopkins, "Applications of Hidden Markov Models to Detecting Multi-Stage Network Attacks," Proc. 36th Ann. Hawaii Int'l Conf. System Sciences, vol. 9, pp. 334-344, 2003.

[12] X.D. Hoang, J. Hu, and P. Bertok, "A Multi-Layer Model for Anomaly Intrusion Detection Using Program Sequences of System Calls," Proc. 11th IEEE Int'l Conf. Networks, pp. 531-536, 2003.

[13] T. Lane, "Hidden Markov Models for Human/Computer Interface Modeling," Proc. Int'l Joint Conf. Artificial Intelligence, Workshop Learning about Users, pp. 35-44, 1999.

[14] Dong Xiang, Huahua Yan, Xianqiao Chen, Yanfen Cheng, "Offline Arabic Handwriting Recognition System based on HMM" 978-1-4244-5540-9/10/$26.00 ©2010 IEEE

[15] Li Z, Wu Z, He Y and Fulei , Hidden Markov model-based fault diagnostics method in speed-up and speed-down process for rotating machinery. Mechanical Systems and Signal Processing, Vol. 19(2), pp. 329-339.