Hidden Markov Models (HMM)

M. Dudochkin, mdudochk@cs
Y. Lakhtin, ylakhtin@cs

PDF-XCHANGE Click to buy NOW! www.docu-track.com

PDF-XCHANGE Click to buy NOW! www.docu-track.com

The maintenance.

- Introduction.
- Cases of application Hidden Markov Models.
- Aspects of model construction.
- Example and review task.
- Forward-backward algorithms.
- Viterbi Algorithm.
- Baum-Welsh Algorithm.
- Practical example of using HMM.
- Literature list.

PDF-XCHANGE Click to buy NOW! www.docu-track.com

Introduction

Hidden Markov models are one of ways of mathematical model reception of some observable signal. HMM carries to a class of stochastic models. Stochastic models tries to characterize only static properties of a signal, not possessing the information on its specific properties.

In a basis of stochastic models the assumption that:

- the signal can be described by some parametrical casual process;
- parameters of this process can be precisely enough estimated by some, quite certain way is necessary.

For the first time hidden Markov models have been applied by Rabiner (speech recognition,1993).

Cases of application Hidden Markov Models.

Hidden Markov Models are normal for applying, when there are many data sets of small volume. Thus it is supposed, that all sets begin with some fixed condition and the probability of value depends basically on number of that position in a set.

Applications of Hidden Markov Models:

- Speech recognition;
- Image recognition;
- Biocomputer science (Research of fibers and DNA);
- Compression and decompression of audio and video signals;
- And other cases.

PDF-XCHANGE Click to buy NOW! www.docu-track.com

Let's accept following designations:

M - Number of various observable objects (for example, spheres of different color);

$V = \{v_1, \ldots, v_M\}$ - Set of all possible observable objects (for example, spheres of color v_1, color v_2, etc.);

N - Number of conditions of model (for example, drawers in which multi-coloured spheres lay);

$S = \{S_1, \ldots, S_N\}$ - Set of conditions of model (for example, a drawer number S_1, number S_2, etc.);

q_t - Condition in which there is a model during the moment of time (i.e. q_t - one of S_i);

o_t - The object observable during the moment of time ***t*** (i.e. o_t - one of objects v_i);

$O = \{o_1, \ldots, o_T\}$ - Observable sequence;

T - Length of observable sequence;

$\pi = \{\pi_i\}$ - Distribution of probabilities of a choice of an initial condition, i.e. $\pi_i = P(q_1 = S_i)$ - probability of that during the initial moment of time $t = 1$ system is in S_i position;

a_{ij} -Probability of transition from a condition S_i to condition S_j - conditional probability $a_{ij} = P(q_t = S_j \; q_{t-1} = S_i)$; it is considered that it does not depend on time;

$A = \{a_{ij}\}$ - Matrix of probabilities of transition - a square matrix ***N********N***;

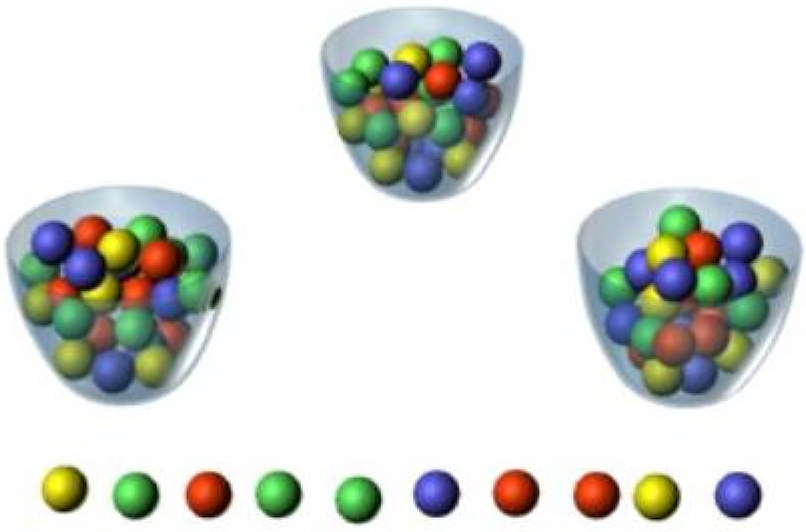
$b_j(k)$ - The probability of that in a condition S_j is observed object v_k, i.e. $b_j(k) = P(o_t = v_k \mid q_t = S_j)$ - a matrix N*M.

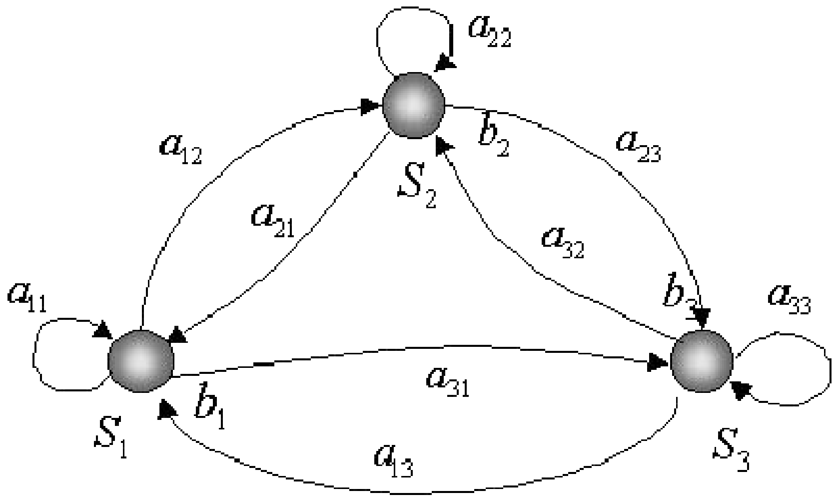
PDF-XCHANGE Click to buy NOW! www.docu-track.com

Hidden Markov Model λ we shall name a set $\lambda = (A, B, \pi)$. We shall show, how the model λ can generate sequence $O = o_1, o_2, \ldots, o_T$ (for example, we should choose ***T*** spheres from ***N*** drawers). On a first step we should choose an initial condition $q_1 = S_i$ (the first drawer) in conformity with distribution of probabilities π and to choose object o_1 (it will be a sphere of color v_k with probability $b_i(k) = P(o_1 = v_k \mid q_1 = S_i)$). Further we pass in any condition $q_2 = S_j$ in conformity with probability a_{ij}: $a_{ij} = P(q_2 = S_j \mid q_1 = S_i)$(we pass to a drawer number S_j). In this condition we choose object o_2 (We choose o_2 - a sphere of color v_l with probability $b_j(l) = P(o_2 = v_l, q_2 = S_j)$). Having executed ***T*** steps of the described process, we shall construct sequence $O = o_1, o_2, \ldots, o_T$ which we shall name observable sequence.

An example

Thus the sequence of conditions ***S*** in which the choice of objects was made, it does not interest us. It the name “Hidden" Markov model also speaks - the sequence of conditions from us "is hidden". The model is " a black box " - after performance of the set quantity of steps it gives out a certain sequence $O = o_1, o_2, \ldots, o_T$.





$$O = \{Y, G, R, G, G, B, R, R, Y, B\}$$

$$S = \{2, 1, 1, 3, 2, 2, 2, 3, 3, 1\}$$

Review task

Let the sequence of supervision $O=\{o_1,\dots,o_T\}$ and model $\lambda=(A,B,\pi)$ are given. How to calculate probability $P(O|\lambda)$ of occurrence of sequence of supervision for the set model?

Solution:

The formula for probability calculation of sequence occurrence of conditions Q for model λ is resulted:

$$P(Q|\lambda)=\pi_{q_1}a_{q_1q_2}a_{q_2q_3}\dots a_{q_{T-1}q_T}$$

The probability of occurrence of the set sequence of supervision for this fixed sequence of conditions is equal to:

$$P(O|Q,\lambda)=b_{q_1}(o_1)b_{q_2}(o_2)\dots b_{q_T}(o_T)$$

For Markov models occurrence of some concrete sequence of conditions and occurrence of supervision sequence are independent events.

$$P(O|\lambda)=\sum_{Q}P(O|Q,\lambda)P(Q|\lambda)=\sum_{Q}\pi_{q_1}b_{q_1}(o_1)a_{q_1q_2}b_{q_2}(o_2)a_{q_2q_3}\dots a_{q_{T-1}q_T}b_{q_T}(o_T)$$

PDF-XCHANGE Click to buy NOW! www.docu-track.com

PDF-XCHANGE Click to buy NOW! www.docu-track.com

Forward-backward algorithms.

So more effective algorithm of probability calculation for which there are two updatings, equivalent on computing expenses - algorithm of a direct course and algorithm of reverse motion refers to. These algorithms differ with a choice of a leading variable, direct or return which is more preferable in each concrete case.

Forward algorithm.

We shall enter a direct variable $\alpha_t(i)$ which we shall define for the set model λ as probability value of that by the moment of time ***t*** the sequence $o_1 o_2 \dots o_t$ was observed, and during the moment ***t*** the system is in a condition S_i :

$$\alpha_t(i) = P\left(o_1 o_2 \dots o_t\, ,\, q_t = S_i \,\middle|\, \lambda\right)$$

Values of a direct variable are calculated in conformity with following procedure:

1. Initialization $\alpha_1(i) = \pi_i\, b_i(o_1) \quad 1 \le i \le N$

2. For all $t = 1, 2, \dots, T-1 \quad 1 \le j \le N$: $\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i)\, a_{ij}\right] b_j(o_{t+1})$

3. Calculation of required probability: $P\left(O \,\middle|\, \lambda\right) = \sum_{i=1}^{N} \alpha_T(i)$

PDF-XCHANGE Click to buy NOW! www.docu-track.com

Backward algorithm.

Let's enter a return variable $\beta_t(i)$ which we shall define as conditional probability of supervision of sequence $o_{t+1}\, o_{t+2}, \ldots, o_T$ since the moment ***t+1*** up to ***T*** provided that during the moment of time ***t*** the system is in a condition S_i :

$$\beta_t(i) = P\left(o_{t+1}\, o_{t+2} \ldots o_T \mid q_t = S_i,\ \lambda\right)$$

Values of a return variable are from following parities:

1. Start value $\beta_T(i) = 1 \quad 1 \le i \le N$

2. For all $t = T-1, T-2, \ldots, 1 \quad 1 \le i \le N$: $\beta_t(i) = \sum_{j=1}^{N} a_{ij}\, b_j(o_{t+1})\, \beta_{t+1}(j)$

3. Probability calculation: $P(O \mid \lambda) = \sum_{i=1}^{N} \pi_i\, b_i(o_1)\, \beta_1(i)$

Viterbi Algorithm.

This algorithm is a variant of a dynamic programming method for Markov circuits. It consists of direct and return passes.

Let's enter following variables:

$$\delta_t(i) = \max_{q_1 q_2 \ldots q_{t-1}} P\left(q_t = S_i \mid q_1 q_2 \ldots q_{t-1}, o_1 o_2 \ldots o_t, \lambda\right)$$

Making sense the maximal probability of that at the set supervision till the moment ***t*** the sequence of conditions will come to the end during the moment of time ***t*** in a condition S_i, and also a variable $\psi_t(j)$ for storage of the arguments maximizing $\delta_t(j)$.

1) Initialization $\delta_1(i) = \pi_i b_i(o_1)$ $1 \le i \le N$ $\psi_1(i) = 0$

2) Inductive transition $\delta_t(j) = \max_{1 \le i \le N}\left[\delta_{t-1}(i) a_{ij}\right] b_j(o_t)$ $1 \le j \le N$ $2 \le t \le T$ $\psi_t(j) = \arg\max_{1 \le i \le N}\left[\delta_{t-1}(i) a_{ij}\right]$

3) Stop $P^* = \max_{1 \le i \le N}\left[\delta_T(i)\right]$- the greatest probability of sequence supervision $o_1 o_2 \ldots o_T$ which is reached at passage of a certain optimum sequence of conditions $Q^* = (q_1^*, \ldots, q_T^*)$ for which by the present moment last condition is known only: $q_T^* = \arg\max_{1 \le i \le N}\left[\delta_T(i)\right]$

4) Restoration of optimum sequence of conditions (return pass).

$$q_t^* = \psi_{t+1}(q_{t+1}^*) \qquad t = T-1, T-2, \ldots, 1$$

Baum-Welsh Algorithm (1)

Let's enter a variable

$$\xi_t(i,j) = P\left(q_t = S_i\, ,\, q_{t+1} = S_j \,\middle|\, O, \lambda\right)$$

Which is probability of that at the set sequence of supervision ***O*** the system during the moments of time ***t*** and ***t+1*** will be accordingly in conditions S_i and S_j. Using the direct and return variables certain above, it is possible to write down:

$$\xi_t(i,j) = \frac{\alpha_t(i)\, a_{ij}\, b_j(o_{t+1})\, \beta_{t+1}(j)}{P(O \mid \lambda)} = \frac{\alpha_t(i)\, a_{ij}\, b_j(o_{t-1})\, \beta_{t-1}(j)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i)\, a_{ij}\, b_j(o_{t+1})\, \beta_{t+1}(j)}$$

Let's enter the following variable which is being aposterior by probability of that at the set sequence of supervision ***O*** the system during the moment of time ***t*** will be in a condition S_i:

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j)$$

The entered sizes possess following properties:

$\sum_{t=1}^{T-1} \gamma_t(i)$ – Expected number of transitions from a condition S_i

$\sum_{t=1}^{T-1} \xi_t(i,j)$ – Expected number of transitions from a condition S_i to S_j;

Baum-Welsh Algorithm (2)

On the basis of these properties formulas of reassessment of parameters of Markov model are received:

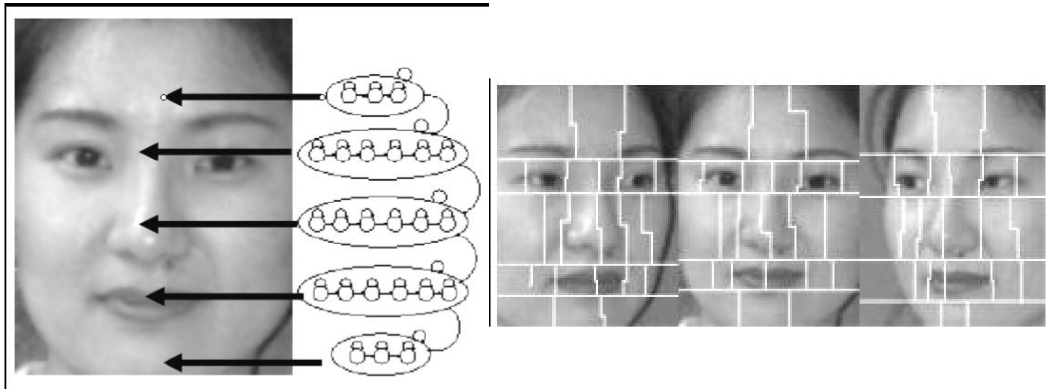
$$\pi_i^* = \gamma_1(i) \qquad a_{ij}^* = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \qquad b_j^*(k) = \frac{\sum_{t=1,\ o_t=k}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)}$$

During application of these formulas there can be only two cases:

1. $\lambda = \lambda^*$ - Extremum point
2. $P(O|\lambda^*) > P(O|\lambda)$ - Plausibility of occurrence of the given sequence of supervision for model with the overestimated parameters above, than for initial model.

Practical example of using HMM.

1) Uniform splitting of entrance vectors on conditions
2) Initialization of model
3) Application of splitting ***Viterbi*** for built in HMM. At this stage there is a redistribution of entrance vectors on conditions.
4) We spend an estimation of the received model, and or it is come back to the previous step, or we speak, that the required model is constructed. Actually there is a segmentation of the image: entrance a vector are divided into groups, i.e. each vector concerns to some inwardness.



The scheme of segmentation of the image on the basis of HMM and the Structural attributes of the image allocated by HMM

Process of recognition in Hidden Markov Models

1) System moves on an input some image
2) System builds of it entrance sequence
3) The System builds probabilities of construction of such entrance sequence all models
4) The most probable answer and probability of conformity stands out

Literature:

1. L. Rabiner, B.-H. Juang. Fundamentals of Speech Recognition. Prentice Hall, 1995, 507 p.
2. X. D. Huang, Y. Ariki, M. A. Jack. Hidden Markov Models for Speech Recognition. Edinburgh University Press, 1990, 275 p.
3. C. D. Manning, H. Schutze. Foundations of Statistical Natural Language Processing. MIT Press, 1999, 680 p
4. http://www.caip.rutgers.edu/~lrr/Reprints/tutorial%20on%20hmm%20and%20applications.pdf
5. http://www.cs.cmu.edu/afs/cs/project/space/www/hmm/hmm.html
6. http://www.accenture.com/xdoc/en/services/technology/publications/hmmtutorialpart2.pdf

PDF-XCHANGE Click to buy NOW! www.docu-track.com

Thanks for attention

Thanks for attention