Janusz BOBULSKI
CZESTOCHOWA UNIVERSITY OF TECHNOLOGY, INSTITUTE OF COMPUTER AND INFORMATION SCIENCE

Selection of parameters of HMM

PhD Eng. Janusz Bobulski

Recieved the M.Sc. degree in Computer Sciences from Czestochowa University of Technology in 2000 and PhD degree in biometrics in 2004. Since 2004 he is assistant professor at the ICIS. His current research interests are focused on image processing, biometrics, multimedia.

e-mail: januszb@icis.pcz.pl.

Abstract

Hidden Markov models are widely applying in data classification. They are using in many areas. The choice of parameters of HMM is very important because of efficiency hole identification system. Individual parameters should be matched individually for each system in the experiment way..

Keywords: image processing, hidden Markov models HMM

Dobór parametrów HMM

Streszczenie

Ukryte modele Markowa (ang. Hidden Markov Models – HMM) są szeroko stosowane do klasyfikacji dany w wielu dziedzinach. Dobór Parametrów HMM jest bardzo ważny ze względu za skuteczność systemu identyfikacji. Poszczególne parametry powinny być dobierane indywidualnie dla każdego systemu w sposób eksperymentalny.

Słowa kluczowe: przetwarzanie obrazów, ukryte modele Markowa, UMM

1. Introduction

Hidden Markov models (HMM) are known in theory as stochastic finite automatons. Every condition is connected with the event possible to be observed – in the case of discrete Markov model. HMM is „double embedded stochastic process", where the elementary stochastic process is hidden, possible to be observed only through the set of other stochastic processes that produce sequences of observations. [1, 2].

The observation determined at t time is denoted by o_t. The observation is a vector, components of which are usually factors of discrete signal spectrum (DTF – Discrete Transform Fourier), linear prediction, wavelet transform etc. The sequence of observations in the period from t = 1 to t = T is marked as $O=(o_1,o_2,...,o_T)$.

Building a model representing face image consists in creation a separate λ model for each image M (or its fragment). Observations sequences obtained from face image M are used in the process of model creation (known as learning or training). Created model is a generator of observations, whose similarities to the observations obtained from recognized face image constitute a base of personal identification.

The identification task, for given set of models $\{\Lambda_i\}$ created for images of faces and placed in the basis, consists in determination of probability concerning generation of O observations sequences through given model, and thereafter comparison of these values.

2. Parameters of hidden Markov models

HMM is a finite automaton, which may be treated as a generator of casual observations sequences. The change of model's conditions, appearing in each step, is described in homogeneous Markov chain, characterized through $A = [a_{ij}]$ matrix of probabilities of transitions between conditions. Decomposition of q_0 conditions probabilities for t = 0 will be denoted by $\pi = [\pi_1, \pi_2,..., \pi_N]$ [3].

Generation of observations (outputs) by the model is random and generally is characterized through the vector of decompositions of observations probabilities $B = [b_i(o_t)]$ with the following components:

$$b_i(o_t) = P(o_t \mid q_t = i) \quad \textit{for } i = 1,2,...,N. \qquad (1)$$

These decompositions, known as outputs, may be discrete decompositions on finite set of output symbols or n-dimensional functions of probability density. Let us assume that o_t is an any vector in E^n euclidean space, then $b_i(o_t)$ is interpreted as o_t observations generation probability through the model being in i condition.

Hidden Markov model is usually marked as $\lambda = (\pi, A, B)$ triple, where π, A, B values are called model's parameters.

Constructing of an automated personal identification system requires creation of tools enabling realization of two fundamental processes:

- training, consisting in estimation of parameters of HMM models set with the help of teaching observations from images;
- recognition, consisting in determination of the model representing face image with the highest probability of generation of observation obtained from unknown test image.

Let's consider a method of an automatic personal recognition, as an example – faces images present people, and set of images creates personal data base. One should make – in the training process - an extraction of $O = (o_1,o_2,...,o_T)$ sequences of observations for each face image from the training set and should use them for HMM building – separately for each image (or its part). The extraction of observations sequence from the image of the person being recognized is made in the process of recognition. Next there is determined the probability of that sequence generation through previously built Markov models for faces from data base. There probabilities serve to assign recognised image to determined model (the person from the base), or they are basis of recognition process [4, 5].

Defining of HMM enables the specification of chain topology and the estimation of probabilities of transitions between conditions and parameters of output decompositions for each conditions. In practice, there are created, in the beginning, simple HMM with output transitions in the form of singular density function. Then, they are replaced with more and more combined model, e.g. with output decompositions in the form of composition of density function.

The goal of training process is to obtain a set of models, which – according to approved criterion, are well-matched to given trainers. Estimation of HMM parameters boils to the optimisation process, where mostly approved criterion, because of the existence of effective algorithms of the solutions, is the criterion of maximum probability [1, 6].

An output decomposition in the simplest constant HMM for *i-th* condition has the form of multi-dimensional Gaussian density function:

$$b_i(o_t) = N(o_t, \mu_i, \Sigma_i) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_i|}} \exp\left(-\frac{1}{2}(o_t - \mu_i)' \Sigma_i^{-1} (o_t - \mu_i)\right), \qquad (2)$$

where n is the dimension of o_t observation vector,

Then each condition in all HMM models is characterized through:

- the vector of μ_i average values,
- Σ_i covariance matrix

Discrete HMM models have output transitions in the form of discrete probability transitions.

Discrete output transition in i-th observation condition o_t has a form:

$$b_i(o_t) = P_i[v(o_t)], \qquad (3)$$

where:

- $v(o_t)$ – the symbol of observations from v-elements set, being the result of o_t observation vector quantization,
- $P_i[v(o_t)]$ – the probability of generation v symbol in i condition.

Thus, discrete output transition is given through the table of similarities in observation symbols. Each symbol has its representative in the form of an observation vector. The set of possible observation vectors is called observation space. Representatives are choosen in that way to cover evenly the observation space and they are usually determined through grouping and averaging observation vectors extracted from training data. That means, that set of symbols is created in the process of observation vectors quantization. The observation vector is attributed to the observation symbol, of which representative is the closest to this vector.

Discrete output transition in i-th condition for observations with derived indenpendt streams $o_t = (o_{t1}, o_{t2}, ..., o_{tS})$ has the following form:

$$b_i(o_t) = \prod_{s=1}^{S} \{P_{is}[v_s(o_{ts})]\}^{k_s}, \qquad (4)$$

where k_s is weight coefficient assigned to s-th stream.

Characteristics' extraction from face images should be made before the process of HMM models' creation or in course of it, as well as before the proper recognition process. It is realized through the processing of face digital images with the help of appropriate software tools (e.g. wavelet transform). There should be accomplished a vector quantisation of obtained characteristics (observation vectors) for HMM models with discrete output decompositions.

Another fundamental stage of personal identification system building is to create hidden Markov models for individual face images. That requires determination of HMM parameters: specification of topology and estimation of parameters' values in the training process.

Specification of topology of HMM model consists in determination the number of its conditions, or the size of A possibilities matrix of transitions between conditions and forms of output transitions – elements of $B = [b_i(o_t)]$ vector. Prototype HMM models are defined in that order.

Training process runs in few stages and it consists in assigning elements of A matrix, π vector of initial probabilities and $B = [b_i(o_t)]$ vector of output transitions.

First initial HMM model is created, i.e. initial values of this model's parameters are determined. That may be done in the following way:

- Select them randomly
- Determine all parameters alike

The process of models' parameters re-estimation is realized after creation of initial set of these models. Re-estimation of HMM parameters may be accomplished twofold:

1) By means of Baum-Welch procedure, using „forward - backward" algorithm;
2) By means of Viterbi procedure, using „best path" algorithm.

3. Defining of prototype HMM models

First step in the process of HMM creation is definition of one or more prototype models. Values of these models' parameters are not important, they are mostly used for determination of topology of appropriate HMM models.

Defining of single HMM model consists in following data specification:

- name or symbol of model λ,
- type of the model (constant or discrete),
- size and contents of observation vector O,
- size and contents of the matrix of observations generation probability B,
- number of data streams S, their size and k_s ($s = 1,2,...,S$) weight coefficients (the sum of all streams' sizes should be equal to the size of an observation vector),
- number of state N,
- transition matrix A, through giving non-zero values to matrix elements all over transitions between conditions are possible. Values of left matrix elements should be zeroth (rows of transition matrix should sum up to one).

4. Estimation of HMM parameters

Basic problem in the process of HMM model creation is the estimation of its parameters. The estimation problem, for given sequence (or set of sequences) of observations $O=(o_1,o_2,...,o_T)$, consists in determination of values of parameters for $\lambda = (\pi, A, B)$ model. That means, at these assumptions for output decompositions, the necessity of average value vectors determination in the training process, as well as variation vectors or matrices of covariance of density function for normal decompositions, being components of the compositions.

The estimation of parameters is realized in the training process of earlier defined HMM models, where the sequence of observation vectors is used.

Basic rule at creation of HMM models is to treat these models as generators of observation vectors. Each observation sequence used during in the training time, constitutes an output for conditions of HMM model, the parameters of which are to be determined. If the condition generated by each of observation vectors in training data was known, then there are possible to determine:

- Unknown average values and variations of that condition's output decomposition, through the use of all vectors connected with this condition,
- Transition matrix, on the basis of the number of transitions between conditions.

Re-estimation of model's parameters

Re-estimation of model's parameters runs similarly as an initiation. The difference pertains to the application of earlier determined initial HMM model in the input, as well as use of Baum-Welch algorithm in the process of parameters' estimation: „forward-backward" algorithm is used at determination of probability of being in each condition. That probability is thereafter used at creation of weighted averages for HMM parameters.

Viterbi algorithm determines „hard" decision of which condition generates specified observation vector from training observation. Baum-Welch algorithm gives „soft" decision, determining only the probability of generation of observations through the condition [7,8].

4. Conclusions

Selection problem of appropriate parameters of hidden Markov model is important, taking into consideration the efficiency of whole identification system. Too high number of states and observations would lengthen training and testing times, but it won't improve system's efficiency. Therefore individual parameters should be matched individually for each system in the experiment way.

5. References

[1] Murphy K..: Hidden Markov Model (HMM) Toolbox for Matlab, http://www.ai.mit.edu/~murphyk/Software/HMM/hmm.html, 2004.
[2] Rabiner L. R.: A Tutorial on Hidden Markov Models and Selected Application in Speech Recognition, Proc. IEEE 77, pp 257-285, 1998.
[3] Li X., Parizeau M., Plamondon R.: Training Hidden Markov Models with Multiple Observations – A Combinatorial Method, IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 22 No.4, pp 371-377, 2000.
[4] Samaria F., Fallside F.: Face Identification and Feature Extraction Using Hidden Markov Models, Image Processing: Theory and Applications, 1993.
[5] Samaria F, Young S.: HMM-based Architecture for Face Identification, Image and Vision Computing, Vol. 12 No 8 October , pp 537-543, 1994.
[6] Kanungo T.: Hidden Markov Model Tutorial, www.cfar.umd.edu/~kanungo, 2003.
[7] Aas K., Eikvil L.: Text Page Recognition Using Gray-level Features and Hidden Markov Models, Pattern Recognition, Vol. 29, pp 977-985, 1996.
[8] Forney G.D.: The Viterbi Algorithm, Proc. IEEE, Vol. 61 No. 3, pp 268-278, 1973.

Artykuł recenzowany