UDC 681.3.06

T. Gryshchuk, Cand. Sc. (Eng.)

DERIVIATION OF CHARACTERISTIC OBSERVATION OF HIDDEN MARKOV MODEL

The process of derivation of voice command characteristic observation on the base of trained Hidden Markov Model. This information can be used for speech recognition process efficiency rising. An idea of estimation function is given, a maximum of which is a characteristic observation. Recommendations for initial approximation selection are given.

Keywords: *characteristic observation, Hidden Markov Model, speech recognition, estimation function.*

Due to the rapid evolution of computer engineering, voice interfaces became more popular. This type of interfaces makes the process of communication with computer easier and leads to the increase of the number of users who use the voice controlled software. Gradually, such type of software comes from application software into system software group. As a result, there exist the following requirements to such software: minimum memory capability and high performance [1].

The main task to be solved in speech communication systems is the task of evaluation of probability of input signal belonging to one of the voice interface commands. Figure 1 shows the simplified scheme of the automatic classifier.

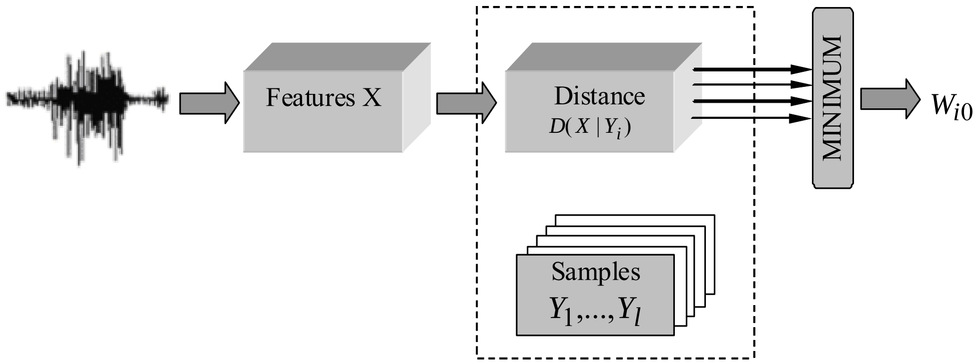


Fig. 1. Automatic classifier structure

The approach based on Hidden Markov Models (HMM) mathematical apparatus is considered nowadays as a classical and universally recognized approach to the speech commands classification. This approach assumes that not the speech signal patters themselves but Hidden Markov Models parameter sets (which are as a matter of fact “generators” of the signal observation vectors) are stored in the memory. Informally speaking, observation is the sequence of vectors, coordinates of which correspond to the time and frequency characteristics of the speech signal in each analysis window (approximately 256 samples). Classification task becomes in fact the task of determining the probabilities of input observation generation by the models, which correspond to the input command language phrase [2, 3].

The first stage of the recognition system building is the stage of HMMs training, i.e., determination of the parameters of the distribution law in each node of a model. This paper is devoted to the process which is inverse comparing to training – receiving of the sample (that is characteristic) speech command observation on the basis of the trained model. Receiving of the characteristic model observation would allow not only to solve the task of simplifying the computation of the speech commands classification process, but also to have additional grounds for solving the tasks of adaptation and creation of the models with noise.

Problem Definition

This article considers the problem inverse to the problem of HMM creation according to Baum–Welch [3] algorithm.

> There is a Hidden Markov Model. It is necessary to find the observation, which gives the maximum in the given model by Baum–Welch.

More formally, let M be a Hidden Markov Model. It is necessary to find the observation x', which can be described by the formula:

$$x' = \arg\max_{x} P\left(x \mid M\right). \tag{1}$$

Characteristic Observation Detection Method

First of all, let us give some necessary designations and definitions of the terms, which will be used in this paper.

Observation vectors will be designated in the following way:

$$\overline{x} = x_1, x_2, ..., x_n,$$

where n is the observation vectors coordinates (parameters) number.

Definition 1 (Forward probabilities):

$$\alpha_1(1) = 1$$

$$\alpha_j(1) = a_{1j}, \quad \text{for j = 2,...,N – 1}$$

$$\alpha_j(t) = \sum_{i=2}^{N-1} \alpha_i(t-1)\, G_i(x_{t-1}) \cdot a_{ij}\ , \text{for j = 2,...,N – 1, t = 2,...,} T_\chi$$

$$\alpha_N(T_\chi) = \sum_{j=2}^{N-1} \alpha_j(T_\chi) G_j(x_{T_\chi}) \cdot a_{jN}.$$

where a_{ij} - transitive probability of transition from node i to node j;

N - model nodes number;

$G_j(\overline{x})$ - distribution function, associated with the emissive node j;

T_χ - expectation of the lengths of the observations, generated by the model M:

$$T_\chi = \sum_{T=1}^{\infty} P(\chi(T) = N) \cdot T \ \text{ or } \ T_\chi = \sum_{T=1}^{\infty} \alpha_N(T) \cdot T.$$

Definition 2 (Backward probabilities):

$$\beta_N(T_\chi) = 1$$

$$\beta_j(T_\chi) = a_{jN}, \text{ for j = 2, ... ,N –1}$$

$$\beta_i(t) = \sum_{i=2}^{N-1} a_{ij} G_j(x_{t+1}) \beta_j(t+1), \text{ for i = 2,...,N – 1, t = 1,...,} T_\chi \text{ -1,}$$

$$\beta_1(1) = \sum_{j=2}^{N-1} a_{1j} G_j(x_1) \cdot \beta_j(1).$$

Now we have the expression, which shows the main idea of the task solving:

$$P\left(x \mid M\right) = \sum_{j=2}^{N-1} \alpha_j(t) G_j(x_t) \cdot \beta_j(t) \tag{2}$$

Expression (2) is true for arbitrary t = 1, ... , T_χ .

So we can gradually enhance the Baum–Welch score by maximum detection:

$$\arg\max_{x_t} P\left(x_t \mid M\right).$$

For t = 1 get: $P\left(x \middle| M\right) = \sum_{j=2}^{N-1} \alpha_j(1) G_j(x_1) \cdot \beta_j(1)$, and for t = T_χ get: $P\left(x \middle| M\right) = \sum_{j=2}^{N-1} \alpha_j(T_\chi) G_j(T_\chi) \cdot a_{jN}$.

To evaluate the quality of the observation score $x = x_1, x_2, ..., x_{T\chi}$ let's consider:

$$e(j,t) = \alpha_j(t) \cdot \beta_j(t),$$

where j = 2,...,N – 1, t = 1,...,T_χ .

Baum–Welch score determines the most probable node, in which observation vector is found in the time moment t [3]. Basing on this score we'll introduce the estimation function as a sum of distribution laws of all the emission nodes of the model:

$$E(x_t, t) = e(2,t) G_2(x_t) + ... + e(j,t) G_j(x_t) +$$
$$... + e(N-1,t) G_{N-1}(x_t)$$

So for any t, t = $1,2,...,T_\chi$ we have the estimation formula $E(x_t, t)$ for the score optimization in x_t .

On the upper level of specification the algorithm of characteristic observation receiving has the following form:

Alg Get Observation

Get Init Observation	' get characteristic time $\mathbf{T_\chi}$, Initial observation $\mathbf{x_1, ..., x_{T\chi}}$ and its score $\mathbf{P_{new}}$ by Baum-Welch
Repeat	
P $\longleftarrow$ $\mathbf{P_{new}}$	
Get Backward Probabilities	' get all $\boldsymbol{\beta_j}$ **(t)**
Get First Forward Probabilities	' get $\boldsymbol{\alpha_j}$ **(1) =** $\mathbf{a_{1j}}$
Loop ' by **t** from **1** to $\mathbf{T_\chi}$	
Get Estimation Function	' get **E** $\mathbf{(x_t,t)}$
Get New Observation Vector	' get $\mathbf{x'_t}$
Get New Forward Probabilities	' get $\boldsymbol{\alpha_j}$ **(t+1)**, if **t** < $\mathbf{T_\chi}$ or $\boldsymbol{\alpha_N}$ $\mathbf{(T_\chi)}$, if **t** = $\mathbf{T_\chi}$
end	
$\mathbf{P_{new}}$ $\longleftarrow$ $\boldsymbol{\alpha_N}$ $\mathbf{(T_\chi)}$	
until $\mathbf{\mid P - P_{new} \mid < \varepsilon}$	

Let us consider the procedures of this algorithm in more details.

Get Init Observation

Here we get the characteristic time, initial characteristic observation and its Baum–Welch score.

Get Backward Probabilities

Here we count the value of $\beta_j(t)$ for this model.

Get First Forward Probabilities

We find the value of $\alpha_j(1) = a_{1j}$ for all the emission nodes.

Get Estimation Function

In this procedure it is necessary to get the estimation function $E(x_t, t)$ to determine the vector x_t. This estimation function is specified with the help of the coefficients:

$$\alpha_j(t) \cdot \beta_j(t) = e(j,t) .$$

$$E(x_t, t) = \sum_{j=2}^{N-1} e(j,t) G_j(x_t) .$$

Get New Observation Vector

The meaning of this procedure is determining of x'_t,

$$x'_t = \arg\max_{x_t} E(x_t, t) .$$

Determining of such functions maximum will be discussed below.

Get New Forward Probabilities

This procedure detects new forward probabilities taking into account that there was found a new observation vector x'_t:

$$\alpha_j(t+1) = \sum_{i=2}^{N-1} \alpha_i(t) G_i(x'_t) \cdot a_{ij}$$

for $j = 2,...,N-1$, $t = 1,...,T_\chi - 1$;

$$\alpha_N(T_\chi) = \sum_{i=2}^{N-1} \alpha_i(T_\chi) G_i(T_\chi) \cdot a_{iN} .$$

Note, that forward probabilities on this cycle step, which corresponds to the time moment t, are determined only for the time moment $t+1$ (for $t < T_\chi$).

If the cycle step corresponds to the time moment $t = T_\chi$, then only $\alpha_N(T_\chi)$ is determined.

Determination of the Maximum of the Function with the Different Constituents Positions

To detect observation vectors it is necessary to determine the maximum of the following function:

$$F(x) = \gamma_2 G_2(x) + ... + \gamma_j G_j(x) + ... + \gamma_{N-1} G_{N-1}(x) ,$$

where $G_j(x) = G_j(x_1,...,x_n)$ is a distribution function, associated with the j node;

$\gamma_j = \alpha_j(t) \cdot \beta_j(t)$.

In the general case it is necessary to find the maximum of the function, which is the sum of the normal distribution laws and has several local extremums. Figure 2 shows the example of such function.

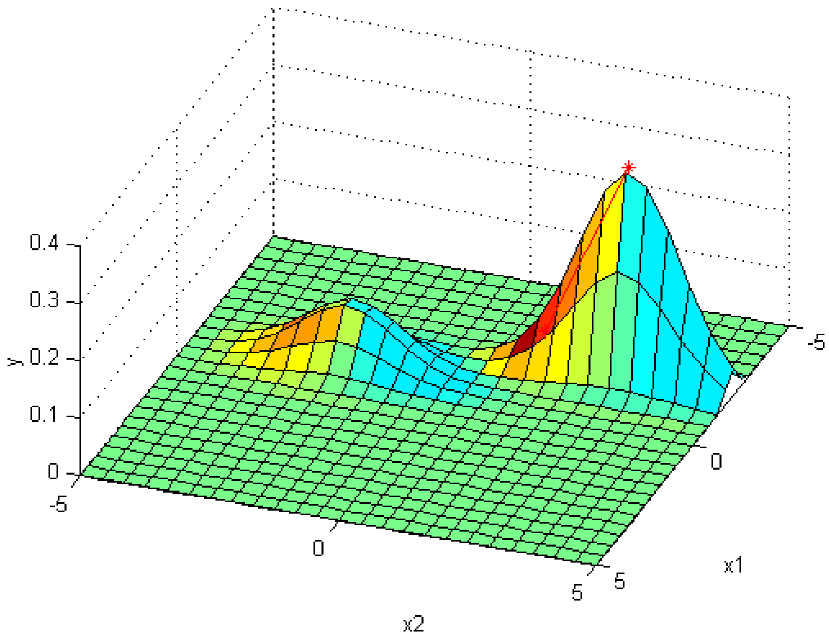


Fig. 2. Example of the model distribution function for 2 nodes and 2 coordinates

The peculiarity of this task is that global maximum is not always reached. That is why we select the x^o observation as an initial approximation:

$$x^o = \frac{1}{W}\left(W_2 \cdot c_2 + ... + W_{N-1} \cdot c_{N-1}\right),$$

where $W_j = F(c_j) = \sum_{i=2}^{N-1} \gamma_i G_i(c_{j1},...,c_{jn}),$

$$W = \sum_{j=2}^{N-1} W_j .$$

Let us consider the following equations set:

$$\begin{cases} \dfrac{\partial E(x_1,...,x_n,t)}{\partial x_1} = 0 \\ ----------- \\ \dfrac{\partial E(x_1,...,x_n,t)}{\partial x_j} = 0 \\ ----------- \\ \dfrac{\partial E(x_1,...,x_n,t)}{\partial x_n} = 0 \end{cases}.$$

Having done the appropriate transformations, we get the equation set for determining the observation vector which maximizes the estimation function:

$$x_i = \frac{1}{\sum_{j=2}^{N-1} e_{ji}(x)} \cdot \sum_{j=2}^{N-1} e_{ji}(x) \cdot c_{ji} \,,$$

where $i = 1,2,\ldots,n$.

$$e_{ji}(x) = \frac{\gamma_j}{\sigma_{ji}^2} \cdot G_j(x) \,,$$

σ_{ji} – mean square deviation for the coordinate i of the distribution function $G_j(x)$, associated with the j node.

c_{ji} – mathematical expectation for the coordinate i of the distribution function $G_j(x)$.

It is possible to solve this equation set by any iteration method of the numeric mathematics.

Experimental Investigations

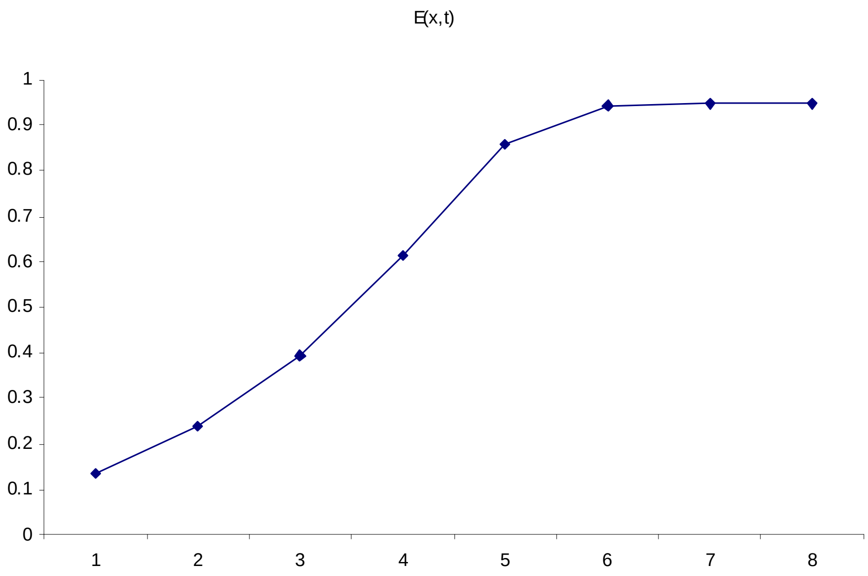
The author of this article performed some experimental investigations of the developed algorithm. Here is the example. Table shows the parameters of the Hidden Markov Model with three emission nodes, for which it is necessary to find the characteristic observation from 2 coordinates with the length of $T_\chi = 9$.

Table

HMM Parameters

Mathematical expectation	Mean square deviation	Transitive probabilities
0 0	0 0	0 1 0 0 0
3.56 1.2	0.6275 0.2115	0 0.6 0.4 0 0
3.5 1.23	0.6169 0.2168	0 0 0.5 0.5 0
2.499 1.189	0.4405 0.2096	0 0 0 0.8 0.2
0 0	0 0	

Figure 3 shows the dynamics of estimation function increase and the value of the resulting characteristic observation.



Results

3.5600	1.2000
3.5137	1.2227
2.7548	1.2039
2.5408	1.1919
2.5055	1.1895
2.5000	1.1891
2.4991	1.1890
2.4990	1.1890
2.4990	1.1890

Fig. 3. Efficiency function value on each iteration

Conclusion

The main scientific result given in this article consists in the development of the method of HMM characteristic observation determination. Experimental investigations proved the conformity of the developed method. Further investigations in this subject can consist in the possibility of characteristic observations usage to solve the tasks of performance enhancement of the recognition and adaptation processes, and creation of models with noise.

REFERENCES

1. Борисов М. ПО для распознавания речи // Мир ПК. – 1999. – №3. – С. 124 – 129.
2. Normandin Y. Hidden Markov Models, Maximum Mutual Information Training, and the Speech Recognition Problem. – Montreal, Canada: Punlished by the Department of Electrical Engineering, McCall University, 1991. – 266 p.
3. Young S., Kershaw D., Odell J., Ollason D., Valthev V., Woodland P. The HTK Book (for HTK Version 3.0). – 2000. – 278 p.

Tetyana Gryshchuk – Senior lecturer of the Department for Computer Control Systems.
Vinnytsia National Technical University.