Adjustment of Observation Probabilities during the Lifetime of Viterbi Algorithm in Unstable Environments

Nader Rezazadeh
Department of Computer, Science and Research Branch, Islamic Azad University Qazvin, Iran
Email: naderrezazadeh1984@gmail.com

Omid Sojodishijani
Department of Computer and Information Technology Engineering-Qazvin Branch, Islamic Azad University, Qazvin, Iran
Email: O_sojoodi@m.ieice.org

***Abstract*—Typically in learning and determining the parameters in various applications of hidden Markov model, like the decoding problem solving in Viterbi algorithm, statistics such as the average are used. In this context, to obtain the average, the stable hypotheses for the data generation process are considered. While in an unstable environment, model parameters such as the value of probability of observing event parameter generated by the state, changes directly between the successive events. For this purpose in this article an adjuster parameter of event probability has been provided in order to adjust and change the parameters after each event during the lifetime of Viterbi algorithm. Test results on the real data sets show the superior performance of the proposed method in terms of accuracy than the other methods.**

***Index Terms*—Statistical Adjustment, Hidden Markov Model, Viterbi algorithm**

I. INTRODUCTION

Hidden Markov Model [1], is nowadays one of the most widely used statistical modeling techniques. Elliptical and flexible network structure of this model is the reason for its abundant usage and popularity.

The common parametric structure of Hidden Markov Model is defined in the form of $\lambda = (A, B, \pi)$ so that. Fig. 1, shows the hidden Markov model with N states and M symbols of observation.

B is a N×M matrix and each of its elements $b_i(k)$ represents the probability of observing symbol kin state of j. Also A is a N×M matrix and each of its elements a_{ij} represents the probability of transition from state i to j. Also π is the initial probabilities matrix of states

All the mentioned parameters is calculated using the Eq. (1), Eq. (2), Eq. (3) [2], [3]:

$$A = \{a_{ij}\}, a_{ij=}P\left(q_{t+1} = S_j \middle| q_t = S_i\right) \tag{1}$$

$$B = b_j(k), b_j(k) = P\left(q_t = V_k \middle| q_t = S_j\right) \tag{2}$$

Manuscript received March 1, 2014; revised May 5, 2014.

$$\pi = \{\pi_i\}, \pi_i = P(q_1 = S_i) \tag{3}$$

$$1 \le i, j \le N, 1 \le k \le M$$

The O_t and V_k variables are respectively, the name of observation symbol in t th time and k th observation symbol.

In order to use the hidden Markov Model with the statistical parameters of λ = (A, B, π), three pivotal models are considered. These three models are evaluation, decoding and learning [2], [3].

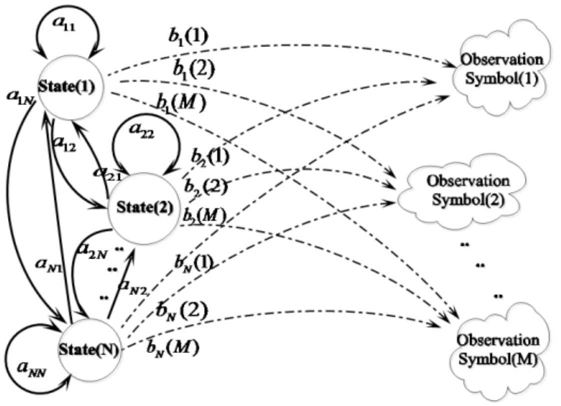

Figure 1. The general structure of hidden Markov model consists of M observation symbols and N states

Proportional to each of these models structure, an algorithm is also proposed to solve it, including the forward and backward algorithm to solve the problem of evaluation, Viterbi algorithm [4], in order to solve the problem of decoding and Baum-Welch algorithm [5], which uses the maximum similarity criterion to solve the problem of education. In all of the above algorithms, even the newer versions of them such as Evidence Feed-Forward Hidden Markov Model, due to ignoring the continuous changes in parameter of observation symbol k possibility by – $b_i(k)$ – i state and the use of a stable hypothesis for the data generating process, the value of mentioned parameter probability during the lifetime of algorithm (and the increase of observations event) is far

doi: 10.12720/ijsps.2.1.23-30

from its true value. In this paper, a non-parametric adaptive method is proposed for update the parameter $b_j(k)$ during the lifetime of Viterbi algorithm. Using this method in Viterbi algorithm, the decoding problems in unstable environments which need to adjust the parameters to the environmental conditions are done with high precision. In addition to comparison and the use of Viterbi algorithm based on the basic model, comparing and using of Evidence Feed-Forward Hidden Markov Model and implementation of adaptive algorithm on them, has been also addressed. This paper is presented as follows: The Evidence Feed-Forward Hidden Markov Model has been addressed in Section 2. In section 3 the proposed algorithm along with the way of parameters' adjustment is presented. In Section 4, the scaling of variance value of parameter $b_i(k)$ probability has been addressed. Chapter 5 shows the experiments and their results on real data and the performance of proposed method compared with other methods and the final section is devoted to the conclusions of the paper.

II. Evidence Feed-Forward Hidden Markov Model

Evidence Feed-Forward Hidden Markov Model is a new model for development of Hidden Markov Models which provides feasibility of calculates the probability of transition between observations symbols in all states of the model. Mentioned connections are in fact the probability of transmission from origin observation symbol $O_t = V_h$ to destination observation symbol $O_t = V_k$, in state $q_t = S_i$. For example, in Fig. 2 an example of Evidence Feed-Forward Hidden Markov Model can be seen with assuming X as the states of rainfall and the lack of rainfall and observations symbols X as umbrella and without umbrella [6].

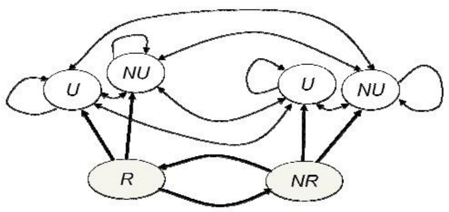

Figure 2. The structure of hidden Markov models based on evidence feed-forward, with 2 states and 2 observation symbols [6]

Symmetrical arrangement of observations connections may bring the ambiguity that observations symbols generated by each state is independent of the others, while this arrangement actually is only for reducing the complexity of shape and drawing the possibility symbol of transition between generated observations symbol in a common state in the mentioned model.

Each independent model of Evidence Feed-Forward Hidden Markov Model method is defined in the form of $\lambda = (A, B, C, \pi)$, By assuming M observation symbols and N state, B is a $N \times M$ matrix and each of its elements $b_j(k)$ represents the value of observation symbol probability k in state j. A is also a $N \times M$ matrix and each of its elements a_{ij} shows the probability of transition from state i to j. Also π_i is the value of initial probability of state i. Also C is a $M \times M \times N$ matrix which in simpler form is characterized by N matrix of $M \times M$ as Eq. (4) [6]:

$$C_1 = \begin{bmatrix} C_1(1,1) & \dots & C_1(1,M) \\ \dots & \dots & \dots \\ C_1(M,1) & \dots & C_1(M,M) \end{bmatrix}$$

$$C_2 = \begin{bmatrix} C_2(1,1) & \dots & C_2(1,M) \\ \dots & \dots & \dots \\ C_2(M,1) & \dots & C_2(M,M) \end{bmatrix}$$

$$C_N = \begin{bmatrix} C_N(1,1) & \dots & C_N(1,M) \\ \dots & \dots & \dots \\ C_N(M,1) & \dots & C_N(M,M) \end{bmatrix}$$

where $C_i(V_h, V_k)$, is Equal With:

$$C_i(V_h, V_k) = P(q_{t+1} = S_{i,V_h} | q_t = S_{i,V_k}) \quad (5)$$

Parameter $q_{t=}S_{i,V_k}$ refers to those data of *i* th state at the time t which have produced V_k.

In Eq. (5) in order to calculate the parameter $C_i(V_h, V_k)$, dividing the frequency of locating the symbol V_k, V_h generated in state *i* – in the form of t and t +1 sequence – to the total number of observation symbol members V_h in the statistical data is used.

In general, each element $C_i(V_h, V_k)$ is the possibility rate of V_h placed as the subsequent observation symbol in state *i*, assuming the V_k placed as the current observation symbol of this state.

All issues raised for hidden Markov model, can also be considered for Evidence Feed-Forward Hidden Markov Model.

The difference between algorithms based on hidden Markov model and algorithm based on Evidence Feed-Forward Hidden Markov Model is in the method of calculate the auxiliary variables in order to solve each of the three pivotal problems.

In the next chapter the new changes in calculation method of auxiliary variables in Viterbi algorithm based on Evidence Feed-Forward Hidden Markov Model will be discussed.

III. Adjustive Viterbi Algorithm in Unstable Environments

In an unstable database, the values of $b_i(k)$ - symbol observation probabilities k generated by state i - is constantly changing. These changes are affected by the impact of corresponding observation symbol event and the probability of observing other symbols in this event. However, here determining the factor of impact in the statistical modeling problems is not discussed but only the average of total possibility changes between the consecutive events of observation symbol k generated from state i is important. This factor is calculated as a matching parameter based on possibility changes during the training data events. This rate of change can be regular or irregular. In most of the natural events these changes have a minimum regularity. Because these events have cause and effect and this causes regular rate of change. In the following the Viterbi algorithm is presented with adjust and update the observations symbol parameter generated by the states of $b_i(k)$ in the algorithm lifetime.

Therefore first, the average parameter of total probability changes of $b_i(k)$ is calculated during the consecutive events (η_k^i)as the adjuster parameter.

A. *Calculate the η_k^i Parameter*

The adjustment of the observation symbol probability during the learning process needs to calculate the statistical outcome of this parameter changes in consecutive events. Therefore, first the three-dimensional matrix B`, which indicates the values of observation symbol probability k generated by state i at the beginning of the process until the event number r is calculated by the Eq. (6).

$$B` = \{b_i^r(k)\}, b_i^r(k) = P(O_t = V_k | q_t = S_i,$$
$$b_i^1(k), b_i^2(k), \ldots, b_i^{r-1}(k))$$
$$1 \le r \le P_k^i - 1,\ 1 \le t \le Occurance\ (i, k, r)$$

where

$$1 \le i \le N,\ 1 \le k \le M \qquad (6)$$

is equal to the number of observation symbol events k generated by state i in the statistical data bank. The parameter of *Occurrence* (*i*, *k*, *r*) represents the time or turn of r th observation symbol event k produced by state i. As can be seen all the observation symbols with the same state are involved in the parameter changes between two consecutive events. In this case, the statistical outcome average parameter of observation symbol probability changes k generated by state i, which in this article is briefly referred as η_k^i, by using the mentioned parameters consecutive events is calculated as Eq. (7).

$$\eta_k^i = Mean\ Of((b_i^{r+1}(k) - b_i^r(k)) = \frac{1}{P_k^i - 1} \times \sum_{r=1}^{P_k^i - 1} (b_i^{r+1}(k) - b_i^r(k)) \qquad (7)$$
$$1 \le i \le N,\ 1 \le k \le M$$

In order to extract the average of probability changes $b_i(k)$ between events V_k, first, the structure of the probability distribution of Markov functions is addressed. Based On Eq. (2), In a Markov B Parameter probability Distribution, unlike the Poisson distribution, changes between the two events V_k are not independent and dependent on observations with the same state.

$$B = \{b_j(k)\},\ \ b_j(k) = P(O_t = V_k | q_t = S_j)$$

where

$$S_j = \{S_{j,1}, S_{j,2}, \ldots, S_{j,M}\} \rightarrow$$
$$\rightarrow b_j(k) = P(O_t = V_k | q_t = S_{j,1} \cup q_t = S_{j,2} \cup ..$$
$$\ldots q_t = S_{j,M}). 1 \le j \le N \qquad (8)$$

Parameter $q_t=S_{i,k}$ refers to those data of i th state at the time t which have produced V_k. For example, in the statistical data for sports match (Iran - Taiwan), in the Asian men championship game of basketball in Philippines, the equation of observation symbol probability of Dunk (cascade) by Iran team in the first quarter of game, parameter b_{Iran}("Dunk"), is calculated as Eq. (9):

$$b_{IRAN}(\text{"Dunk"}) = P(O_t = \text{"Dunk"} | q_t = S_{IRAN}) =$$
$$P(O_t = \text{"Dunk"} | q_t = S_{IRAN,Dunk} \cup q_t = S_{IRAN,Shoot}$$
$$\cup\, q_t = S_{IRAN,Penalty}) \qquad (9)$$

The Eq. (9) shows the use of observations symbols with the state (shot and penalty) in calculating the probability of $b_i(k)$. According to the formula (10), the exact value of b_{Iran}('Dunk') in each turn of the lifetime of mentioned statistical database is shown in Fig. 3.

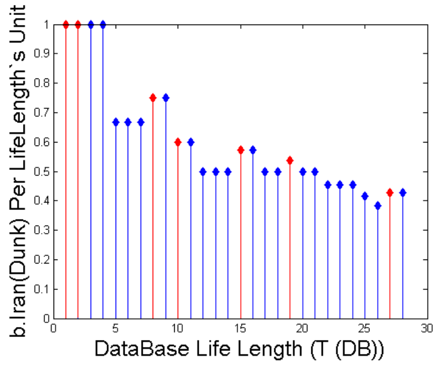

Figure 3. The changes of parameter b_{Iran}("*Dunk*") during the event V_k="*Dunk*", in the lifetime of statistical data

During the event V_k="Dunk", the statistical data b_{Iran}("Dunk") is shown in red. When the non-adaptive Viterbi algorithm, based on Eq. (2) assigns the constant 0.641 for the probability of b_{Iran}("Dunk") in the first quarter of statistical data related to the mentioned sport competition, indeed, the varied and different values can be seen in the lifetime of mentioned parameter statistical data.

Taking into account the non-independence of observations symbol probability generated by the states in the competition (Iran - Taiwan) and similar environments, based on Eq. (8) And Eq. (9), calculate the average of parameter $b_i(k)$ changes during the events of parameter V_k should be done by considering the event changes of all observations symbols with the same state between the events V_k. Therefore the calculation of statistical adjuster parameters of η_k^i is completely based on Markov probability distribution and non-independent (non-Poisson).

B. *Adaptive Viterbi Algorithm*

In this section, with respect to the adaptive parameter η_k^i, the modified Viterbi algorithm equipped with update will be provided. Conventional Viterbi algorithm has four main sections as follows:

1- Initialize 2- Recursion: calling the recursive variables in order to calculate the probability of the maximum corresponding state 3- Termination: adjustment of the final conditions and 4- Trace Back: search to backward in order to find the optimal states sequence.

In the proposed algorithm, the order of observations symbol update generated by the states is adjusted according to the movement direction of Viterbi algorithm ($\delta_t(j)$) and its all lateral calls. Table I, shows the meeting of observations symbols generated by states.

TABLE I. THE EVENT ORDER OF "OBSERVATIONS SYMBOLS GENERATED BY THE STATES" AT THE TIME T`

Current Time	auxiliary Variables Called By Main auxiliary Variable	$b_i(k)$ Order
t=t`	$\delta_{t`}(1)=MAX_i[\delta_1(i)a_{i2}]b_1(O_{t`})$	$b_1(O_{t`})$
	$\delta_{t`}(2)=MAX_i[\delta_1(i)a_{i2}]b_2(O_{t`})$	$b_2(O_{t`})$
	$1 \le j \le N$ ...	$b_j(O_{t`})$
	$\delta_{t`}(N)=MAX_i[\delta_1(i)a_{i2}]b_N(O_{t`})$	$b_N(O_{t`})$

Initialize:
$\delta_{1(j)}=\pi_j b_j(o_1),\ \psi_1(j)=0$
$Update(1,o_1),\ \ 1 \le j \le N$
Recursion:
$For(t=2; t \le t; t=t+1)$
$\delta_t(j)=Max_i(i)[\delta_{t-1}(i)a_{ij}]b_j(o_t)$
$\psi_t(j)=argMax_i[\delta_{t-1}(i)a_{ij}],\ 1 \le j \le N$
$Update(t,o_t);$
End;
Termination:
$\Delta^*=Max_i[\delta_T(i)],\ x_T^*=argMax_i[\delta_T(i)];\ 1 \le i \le N$
TraceBack:
$For(t=T-1, t \ge 1; t=t-1)$
$x_t^*=\psi_{t+1}(x_{t+1}^*);\ X^*=\{x_1^*, x_2^*, \ldots, x_T^*\};$
End;
Function $Update(t,o_t)$
$For(t`=1; t` \le T; t`=t`+1)$
$if(t`==t)\ Observation=O_t;\ End;$
$For(i=1; i \le N; i=i+1)$
$For(k=1; k \le M; k=k+1)$
$if(o_t==k)\ b_i^{Adjusted}(k)=b_i(k)+\eta_k^i;$
$Else\ \ b_i^{Adjusted}(k)=b_i(k);\ \ End;$
$Sum(i,k)=b_i^{Adjusted}(k)+Sum(i,k);$
End;
End; $Controller=1;$
$For(i=1; i \le N; i=i+1)$
$For(k=1: k \le M; k=k+1)$
$if(Sum(i,k) \le 1)\ \ b_i^{Adjusted}(k)=$
$\frac{1}{Sum(i,k)} \times b_i^{Adjusted}(k);$
$Else\ \ b_i^{Adjusted}(k)=$
$\frac{Sum(i)}{1} \times b_i^{Adjusted}(k);\ \ End$
$if\left(\left(b_i^{Adjusted}(k) \le MinProbability(i,k)\right)\right)$
$\cup b_i^{Adjusted}(k) \ge MaxProbability(i,k)))$
$Controller=0;\quad End;$
End;
End;
$For(i=1: i \le N; i=i+1)$
$For(k=1: k \le M; k=k+1)$
$if(controller=1)\ \ b_i(k)=b_i^{Adjusted}(k);\ \ End;$
End;
End;
End;

Algorithm 1. The integrated and adaptive viterbi algorithm

In Algorithm 1, based on observation symbol order generated by the state at the time, as specified in Table I and the type of observation symbol, the Update function is called and the adjustment act is done. The pseudo-code of proposed algorithm is shown in the following:

Parameter P_k^i is equal to the number of events $b_i(k)$ in corresponding educational database.

The parameters MaxProbability (i, k) and MinProbability (i, k), are, respectively, maximum and minimum values of the observation symbol probability of k event produced by state i between $b_i^r(k)$ s, $1 \le r \le P_k^i$. These parameters have the role of control parameters and specify the permissible values of adjusted parameter.

Finally, the normalization along with mentioned control parameters which occurs proportional with the value of each of the probability data, causes maintaining the status quo in Eq. (10) and (11) for the values of adjusted probability of $b_i^{Adjusted}(k)$.

$$b_i^{Adjusted}(k) \ge 0 \qquad (10)$$

$$\sum_{k=1}^{M} b_i^{Adjusted}(k) = 1 \qquad (11)$$

$$1 \le k \le M,\ 1 \le i \le N$$

C. The Adjustment of Adaptive Viterbi Algorithm Based on the Evidence Feed-Forward Model

The difference of Adaptive Viterbi algorithm based on the Evidence Feed-Forward Hidden Markov Model and Hidden Markov Model is in the calculation process of auxiliary variable $\delta_t(i)$

The use of argument $C_i(O_{t-1}, O_t)$ in calculation of auxiliary variable $\delta_t(i)$ in the Viterbi algorithm based on the Evidence Feed-Forward Model is in order to more accurate estimate of decoding problem answer. In the following, the observation symbol will update at any time by function Update(t, o_t) to use in the adaptive algorithm.

auxiliary variable $\delta_t(i)$ Calculation, and send the problem observation to function Update(t, o_t) in the initialization and recursive steps is observed in order to update and adjust $b_i(k)$ Along with the main stages 3 and 4, Presented in Algorithm 2.

InitiaLize:
$\underset{1 \le i \le N}{\delta_1(i)}=\pi_i b_i(O_1)$
$\underset{1 \le i \le N}{\psi_1(i)}=0;$
$Update(t,O_t);$
Recursion:
$\delta_t(j)=\max_{1 \le i \le N}[\delta_{t-1}a_{ij}b_j(O_t)C_i(O_{t-1},O_t)]$
$\psi_t(j)=\underset{1 \le i \le N}{\text{Arg}}\ \text{Max}_i\ [\delta_{t-1}(i)a_{ij}],\ 1 \le j \le N$
$Update(t,O_t);$
TraceBack:
$For(t=T-1, t \ge 1; t=t-1)$
$x_t^*=\psi_{t+1}(x_{t+1}^*);\ X^*=\{x_1^*, x_2^*, \ldots, x_T^*\};$
End;

Algorithm 2. The main four stages in adjustive Evidence Feed Forward HMM

Difference of adjustive Evidence Feed Forward HMM and adjustive HMM, is In calculating the auxiliary

Variables Designing Updates Arguments And Update Function Structure As adjustive HMM is Like the Recursion Step, on Algorithm 1, Due To Retrieve in Back Tracking Step, We Need To keep the Track of the arguments which maximized $\delta_t(j)$ for each t and j, auxiliary variable $\Psi_t(j)$ is containing the name of most likely state lead to the state j at time t, assuming $(V_k = O_t|S_j)$.

D. *Scaling the Variance Rate of Probability Changes of Parameter $b_i(k)$ between the Events V_k*

One of the study goals is to compare the variance of parameter $b_i(k)$ sequence during the events V_k in the statistical data to the possible maximum and minimum of mentioned variance in environment. This goal is to investigate the association between increasing and decreasing of sequence changes of parameter $b_i(k)$ during the events V_k in statistical data and the average of estimate accuracy of adaptive and non-adaptive algorithms.

Also, by calculating the variance of the parameter $b_i(k)$ sequence during the events V_k in the statistical data, the instability rate of mentioned sequence in the modeling environment can be obtained [7].

Highest variance in possibility changes of $b_i(k)$ during events V_k will occur when the total changes are completely without purpose and randomly and without any symptom in orientation. In this case, the total changes in the upside path will be equal to the total changes in downside (resultant of changes will be zero).

$$\sum_{r=1}^{P_k^i-1} b_i^{r+1}(k) - b_i^r(k) = 0 \rightarrow$$

$$\frac{1}{P_k^i-1} \times \sum_{r=1}^{P_k^i-1}\left(b_i^{r+1}(k) - b_i^r(k)\right) = 0 \quad (12)$$

In following we apply the mentioned conditions exist in the equation of the variance of parameter $b_i(k)$sequence during the events V_k in statistical data, in the Eq. (13).

$$\frac{1}{P_k^i-1} \times \sum_{r=1}^{P_k^i-1}\left(\left(b_i^{r+1}(k) - b_i^r(k)\right) - \left(\frac{1}{P_k^i-1} \times \sum_{r=1}^{P_k^i-1}\left(b_i^{r+1}(k) - b_i^r(k)\right)\right)\right)^2$$

$$= \frac{1}{P_k^i-1} \times \sum_{r=1}^{P_k^i-1}\left(\left(b_i^{r+1}(k) - b_i^r(k)\right) - 0\right)^2$$

$$= \frac{1}{P_k^i-1} \times \sum_{r=1}^{P_k^i-1}\left(\left(b_i^{r+1}(k) - b_i^r(k)\right)\right)^2 \quad (13)$$

According to the observations, the output variance of the parameter $b_i(k)$, during the events V_k at the maximum state of possibility changes of $b_i(k)$, during the events V_k is equal to the mean squared of sequence changes of parameter $b_i(k)$ during the events V_k, and parameter r is the event counters of observation symbol V_k generated by state I, which $1 \le r \le P_k^i$.

Based on the obtained results, the output variance of the parameter $b_i(k)$ during events V_k at the maximum state in Eq. (12) And Eq. (13) will form the function SVD_k^i by Eq. (14), which on its basis can distinguish the value of probability variance $b_i(k)$ during the events V_k at statistical data in accordance with respective zero and one range.

$$SVD_k^i = \begin{cases} if\left(\frac{1}{P_k^i-1} \times \sum_{r=1}^{P_k^i-1}\left(b_i^{r+1}(k) - b_i^r(k)\right)^2 \neq 0\right) \\ SVD_k^i = \sum_{r=1}^{P_k^i-1}\left(\left(b_i^{r+1}(k) - b_i^r(k)\right) - \frac{1}{P_k^i-1} \times \sum_{r=1}^{P_k^i-1}\left(b_i^{r+1}(k) - b_i^r(k)\right)\right)^2 \times \frac{1}{\sum_{r=1}^{P_k^i-1}\left(b_i^{r+1}(k) - b_i^r(k)\right)^2} \\ if\left(\frac{1}{P_k^i-1} \times \sum_{r=1}^{P_k^i-1}\left(b_i^{r+1}(k) - b_i^r(k)\right)^2 == 0\right) \\ SVD_k^i = 0 \end{cases} \quad (14)$$

According to Eq. (14), in exchange for the highest values, if the numerator is equal to mean squared of changes in the sequence of parameter $b_i(k)$ during the events V_k, will give us the output 1, also in exchange for the least changes, the values of mentioned sequence will have the same level as average which as the average value being decrease in numerator, the numerical value will tends to zero.

Only when the value of the parameter $b_i(k)$ during the events V_k are equal the square of sequence changes of parameter $b_i(k)$ during the events V_k, (the denominator), will be equal to zero. Since in this case, the rate of change is equal to zero constant, the function output is also zero in this case, which in order to avoid ambiguity in Eq. (15), the mentioned state is independently included in the equation.

Based on the linear gradient of function SVD_k^i and the equation of function, the mentioned function can be traced as you can see in Fig. 4.

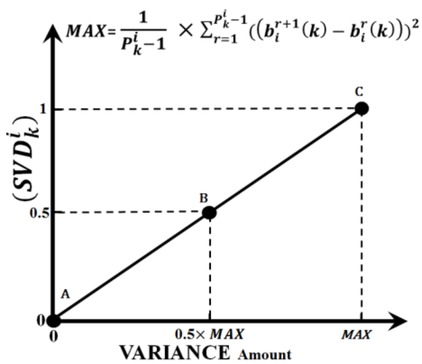

Figure 4. Graph of function 〖SVD〗_k^i output based on the variance of probability bi (k), during events V_k, in statistical data

Like the point A and point C, which are obtained in exchange for the entry of minimum and maximum possible value SVD_k^i in the corresponding environment, the other function points such as point B, could be calculated considering the line slope of the diagram of

Fig. 5 which is equal to $\frac{1}{\text{MAX}}$, that are equivalent to the inverse parameter value of the squares sum of the parameter $b_i(k)$ sequence changes during the events V_k in statistical data.

Due to the structure of the adaptive parameter, the adjuster section in algorithm 1, can be represented as Eq. (15):

$$b_i^{Adjusted}(k) = b_i(k) + \eta_k^i$$
$$= b_i(k) + \frac{1}{P_k^i - 1} \times \sum_{r=1}^{P_k^i - 1} (b_i^{r+1}(k) - b_i^r(k))$$
$$1 \le i \le N, 1 \le k \le M \qquad (15)$$

Along with the increase in the variance of the probability $b_i(k)$ during the events V_k, the adaptive parameter (η_k^i) in adjuster section of proposed algorithm grown smaller and the final answer of observation symbol probability event V_k produced by state i in adaptive algorithm is get closer to the answer of non-adaptive algorithm, namely $b_i(k)$. The flexibility of the proposed algorithm will cause that even in statistical environments with very high random changes that the possibility of statistical adjuster parameter extraction is lower, the average accuracy of adaptive algorithm estimation never be less than the non-adaptive Viterbi algorithm.

IV. Experimental Results

In order to investigate the estimation ability of adaptive algorithms compared to non-adaptive ones, three sets of experiments with different characteristics in the terms of use of parameters have been adopted.

A. *Model Environments*

This experiment has been implemented in two different environments.

1-Environment 1

This environment includes statistical data of Asian championship Games men's basketball, 2013 - Philippine, knockout stage, Semifinal, game of (Iran - Taiwan) [8].

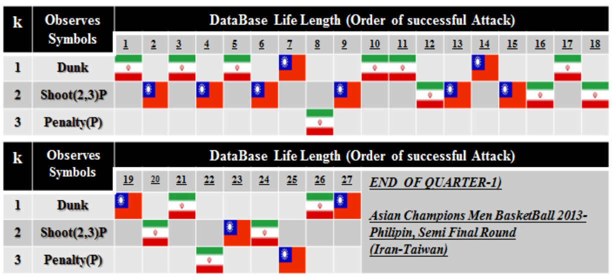

Figure 5. Showing the data bank, asian basketball championship games, 2013 philippine, semifinals, (Iran - Taiwan)

The observation symbols are all the techniques exist in order to earn points. According to basketball rules [9] will include: 1- Entering the ball into the basket with jumping and cascade which is specific to basket, (Dunk); 2- entering the ball into the basket with 2-point or 3-point shots Shoot, (2P, 3P); 3- the penalty kick that at least one of its throw will be scored, (Penalty). States based on the alphabet is divided into two main categories. 1 - Iran: All players who have produced the above observations symbols are included in this case. 2 - Taiwan: All players who have produced the above observations symbols are included in this case. The order of successful attacks has the role of statistical data lifetime ($T_{DataBase}$) and all data are defined on this parameter. Fig. 5 shows the statistical data in the first quarter of the game.

2-Environment 2

This environment includes Data Bank volleyball match (Brazil - Russia), in the final stage of the London 2013 Olympic Games ([10]).

Observation symbols are included common techniques available in order to earn points, which are set based on the volleyball rules [11]; this observation symbols are:

1- Types of Attack (Spike); 2- Types of Service (Service); 3- Types of Defense directly lead to points (Block); 4- Mistakes or errors, which will lead directly to earn points (Opponent Team Error). Fig. 6 shows the statistical data in the first Set of the game.

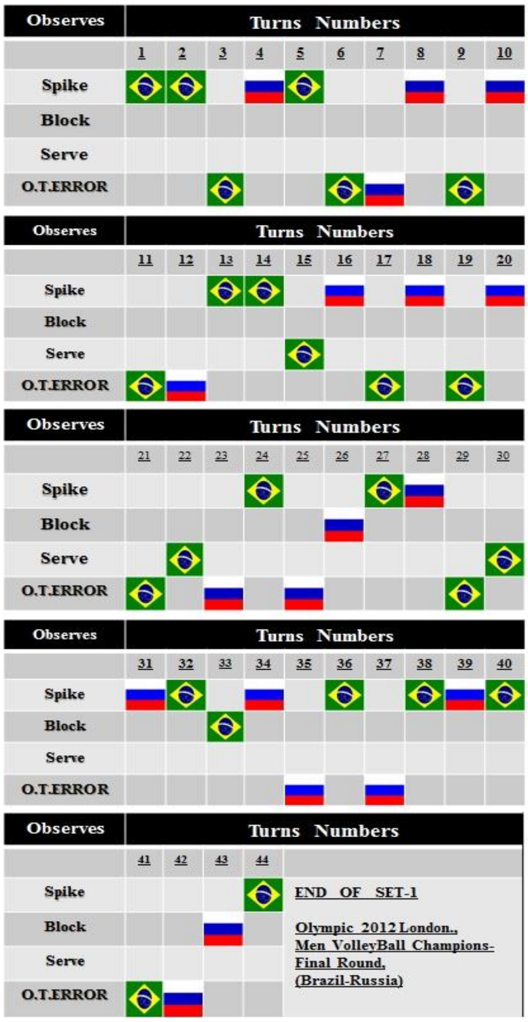

Figure 6. Showing the Data Bank, 2012 London Olympics Games, Volley Ball, Final Match (Brazil-Russia)

At Following, To Scaling the random changes in the environments, variance rate of the probability $b_i(k)$ changes during the events V_k, in the two Environments, 1 and 2, using Eq. (14), Calculated and is given in Table II.

TABLE II. CALCULATED SVD^i_k VALUES, AT E-1 (ENVIRONMENT 1) AND E- 2 (ENVIRONMENT 2)

SVD^i_k	Value In E-2	Value In E-1
SVD^1_1	0.365	0.173
SVD^1_2	0.409	0.218
SVD^1_3	0.312	0.327
SVD^1_4	0.431	--
SVD^2_1	0.233	0.187
SVD^2_2	0.339	0.118
SVD^2_3	0.433	0.094
SVD^2_4	0.156	--

In order to calculate the scale average of the probability $b_i(k)$ variance during the events V_k between calculated scales within M×N type of observation symbol produced by state, Eq. (16) is used.

$$\overline{SVD^i_k} = \frac{\sum_{i=1}^{N}\sum_{k=1}^{M} SVD^i_k}{M \times N} \qquad (16)$$

M is the number of observations symbols and N is the number of existing states, by calculating the mentioned parameter according to Eq. (16), which is equal to 0.1612 in environment 1, and is equal to 0.3347 in Environment 2, it can be argued that on average, the variance scale of the probability $b_i(k)$ during the events V_k in Environment 2 is greater than those Environment in 1

B. Design Decoding Problems

In Order To Challenging Of proposed algorithm Estimation Ability, faced on Statistical test datasets,

Decoding Problems in Various Lengths are presented, that In each test runs. represented decoding problem's length -T- is, as following:

T = 3 For Test1
T = 5 For Test2
T = 7 For Test3
T = 9 For Test4

Data Test Format, AT every test turn, is formed, according to Eq. (17)

$$(X,Y) = \{(X_1,Y_1),(X_2,Y_2),\ldots,(X_T,Y_T)\},$$
$$\{(X_2,Y_{t`+2}),(X_{t`+3},Y_{t`+3}),\ldots,(X_{t`+T+1},Y_{t`+T+1})\},$$
$$\{(X_{t`+1},Y_{t`+1}),(X_{t`+2},Y_{t`+2}),\ldots,(X_{T_{T.Database}},Y_{T_{T.Database}})\}$$

where

$$t` = T_{T.Database} - T \qquad (17)$$

$T_{T.DateBase}$ is Test Statistical Data Set Life Length.

C. Test Experiments

In the following series of experiments, in experiment No. 2, the length of decoding problem assumed to be constant, and we compare the accuracy rate of adaptive and non-adaptive algorithm and medium scale of probability variance $b_i(k)$ during the events V_k in statistical test data for corresponding environment.

As stated earlier, with increasing the length of decoding problem which is proportional with increase of event V_k, decreasing trend of accuracy estimation of the Viterbi algorithm is accelerated. Proposed algorithm moves towards more precision of parameter $b_i(k)$ and lowering increasing acceleration of accuracy reduction, by adjust the possibility changes of $b_i(k)$ during the events V_k. Fig. 7 is formed in order to compare the average of estimation accuracy of conventional Viterbi algorithm and adaptive Viterbi algorithm at Environment 1.

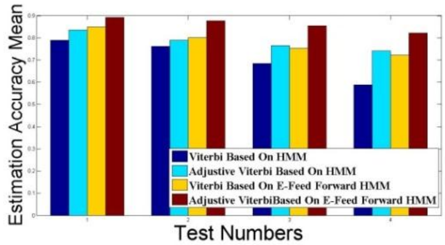

Figure 7. Average estimation accuracy of adaptive and non-adaptive algorithms, in quad experiments with different problem length in environment 1

In the following, the series of quad experiments with the equal lengths of decoding problems with experiment 1 is applied in Environment 2. The graph in Fig. 8 is formed in order to compare the average of estimation accuracy of conventional Viterbi algorithm and adaptive Viterbi algorithm.

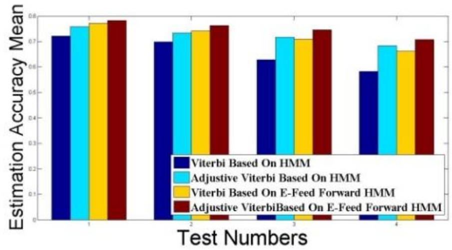

Figure 8. The average of estimation accuracy of adaptive and non-adaptive algorithms, in quad experiments with different length of problem in the environment 2

According to the results of experiments, the average of estimation accuracy of the non-adaptive Viterbi algorithms and adaptive Viterbi algorithm in Environment 2, are closer to each other. Due to the variance increment in possibility changes of $b_i(k)$ during the events V_k, in Environment 2, these results are normal.

In the following, the third experiment will be formed by assuming a constant decoding problem length of T=6 and in order to investigate the relationship of the variance scale average of probability changes of $b_i(k)$ during the events V_k in statistical environment of competition and the average accuracy of estimation.

TABLE III. CORRESPONDENCE TABLE OF THE COMPETITION NUMBER (ENVIRONMENT) AND THE AVERAGE OF VARIANCE SCALE OF THE PROBABILITY CHANGES OF $b_i(k)$ DURING THE EVENTS V_k

$\overline{SVD^i_k}$	Game Name	N
0.417	Miami Heat-Brooklyn Nets	1
0.388	Milwaukee Bucks-Boston Celtics	2
0.312	Toronto Raptors-Atlanta Hawks	3
0.236	Detroit Pistons-Memphis Grizzlies	4
0.209	Dallas Mavericks-Houston Rockets	5
0.182	Portland Blazers-Denver Nuggets	6
0.136	New Orleans Pelicans-Orlando Magic	7
0.104	Utah Jazz-Phoenix Suns	8

To this end, 8 games of the games played series in the first week of November 14, 2013, of men's basketball leagues [12] NBA has been selected. How to extract the symbols of observations and states is precisely in the form of (Iran - Taiwan) competition. In order to establish a correspondence between the number of game (environment) and parameter $\overline{\mathrm{SVD}}_k^i$ use the Table III. the competitions are sorted based on the numerical value of mentioned parameter and in the descending direction.

The term $\overline{\mathrm{Scale}}(\mathrm{D})$ refers to the parameter $\overline{\mathrm{SVD}}_k^i$ in the selected competition. In the following, the accuracy of answering each of the adaptive and non-adaptive algorithms against the decoding problems with constant length (T=6) is calculated in octoploid environment, and is drawn in Fig. 9.

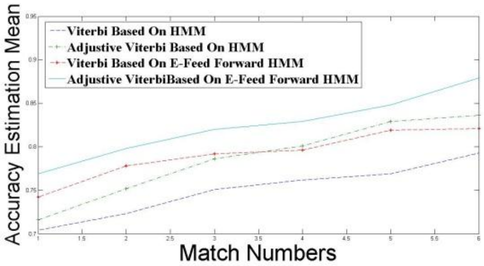

Figure 9. The average of estimation accuracy in adaptive and non-adaptive Viterbi algorithm, assuming a constant length of problem (T = 6) in the octoploid environment

By comparing the parameter $\overline{\mathrm{SVD}}_k^i$ and the average of estimation accuracy of adaptive and non-adaptive Viterbi algorithm, the crucial role of the variance of probability $b_i(k)$ during the events V_k in the environment and the average of estimation accuracy of the algorithm, can be achieved. Also increase in the performance of the adaptive Viterbi algorithm and adaptive Viterbi algorithm based on the Hidden Markov Model and based on the Event Feed-Forward Model, compared to its non-adaptive types during reduction of the variance mean of the probability $b_i(k)$ during the events V_k is evident in the Environment of competition statistical data in Fig. 9

V. Conclusion

The proposed algorithm is an adaptive algorithm with the non - parametric and statistical nature which causes to obtain more accurate values of the parameter $b_i(k)$ in the Viterbi algorithm by applying the adjustment in values of probability $b_i(k)$ after each event V_k in the lifetime of Viterbi algorithm through the parameter η_k^i, which is equal to the average of total possibility changes of $b_i(k)$ during the event V_k in the training statistical data.

The calculation of statistical adjuster parameter η_k^i, completely based on Markov probability distribution is non-independent (non-Poisson) and applying it in the algorithm considering the control parameters in order to receive values are allowed. Also, the use of Evidence Feed-Forward Hidden Markov Model, due to increasing the ability to estimate observations has increased the estimation, general search and understanding the adaptive algorithm of observations communication.

According to the experiments, despite the performance of the non-adaptive algorithm being close to its adaptive ones during the increment of probability variance changes of $b_i(k)$ during the events V_k (increment of Suddenly changes of $b_i(k)$ during the events V_k in statistical data), according to the interpretations of Eq. (12) with the assumption that the most probability changes of $b_i(k)$ during the events V_k are adjusted in statistical data, and considering the flexibility and reduction in adjustment step in Eq. (15) in the adaptive algorithm, the average of estimation accuracy of the mentioned algorithm even in the worst conditions, will never be lower than non-adaptive Viterbi algorithm.

References

[1] L. R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proc. of the IEEE*, vol. 77, no. 2, pp. 257-286, 1989.

[2] A. Wilson and A. Bobick, “Parametric hidden Markov models for gesture recognition,” *IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 21, no. 9, pp. 884-899, 1999.

[3] R. S. Feisal, “An evaluation of HMM-based techniques for the recognition of screen rendered text,” in *Proc. International Conference on Document Analysis and Recognition*, 2011, pp. 1520-5363.

[4] A. Bouloutas, G. W. Hart, and M, Schwartz, “Two Extensions of viterbi algorithm,” *IEEE Transactions on Information Theory*, vol. 37, no. 2, pp. 430-436, 1991.

[5] P. M. Baggenstoss, “A modified baum-welch algorithm for hidden Markov models with multiple observation spaces,” vol. 9, no. 4, pp. 411-416, May 2001.

[6] M. D. Rose, C. Wagner, and P. Frederick, “Evidence feed forward hidden Markov model: A new type of hidden Markov model,” *International Journal of Artificial Intelligence and Applications (IJAIA)*, vol. 2, no. 1, pp. 92-111, Jan. 2011.

[7] D. A. Freedman and K. W. Wachter, “Methods for census 2000 and statistical adjustments,” *Social Science Methodology. Sage*, pp. 232-245, 2007.

[8] Asian Volleyball Championship Games Dataset. [Online]. Available: http:// FIBA_Asia_Championship

[9] Basketball Basis Rules. [Online]. Available: http://www.fiba.com/pages/eng/fc/baskbasi /basir ule.asp

[10] London Olympic Volleyball Games Dataset. [Online]. Available: http://www.volleywood.net/volleyball-tournaments/2012-london-olympics-volleyball-tournaments

[11] Volleyball Basis rules. [Online]. Available: http://www.fivb.org/en/volleyball /Basic_Rules.asp

[12] NBA Basketball Games Dataset. [Online]. Available: HTTP://www.nbafile.com/category/nba-classic-games

Nader Rezazadeh received Software Engineering from the University of Science and Research Branch, Islamic Azad university Lahijan, Iran. presently he is Student of MSc in Artificial Intelligence, Department of Computer and Information Technology Engineering, Qazvin Branch, Islamic Azad University ,His research interests Artificial Intelligence Especially in Pattern Recognition, Data Mining And Statistical data Modeling.

Omid Sojodishijani received the PhD degree in artificial intelligence from the University Putra Malaysia (UPM) in 2011. He is an assistant professor in the faculty of Computer and Information Technology Engineering, QIAU, Iran. His current research activity addresses adaptation and learning in non-stationary environments and intelligent data processing.