Research Article

A Novel Method for Decoding Any High-Order Hidden Markov Model

Fei Ye^{1,2} and Yifei Wang³

¹ Computational Experiment Center for Social Science, Nanjing University, Nanjing 210093, China

² School of Mathematics and Computer Science, Tongling University, Tongling, Anhui 244061, China

³ Department of Mathematics, Shanghai University, Shanghai 200444, China

Correspondence should be addressed to Fei Ye; postyf@163.com

Received 9 August 2014; Revised 19 October 2014; Accepted 11 November 2014; Published 23 November 2014

Academic Editor: Weiming Xiang

Copyright © 2014 F. Ye and Y. Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes a novel method for decoding any high-order hidden Markov model. First, the high-order hidden Markov model is transformed into an equivalent first-order hidden Markov model by Hadar's transformation. Next, the optimal state sequence of the equivalent first-order hidden Markov model is recognized by the existing Viterbi algorithm of the first-order hidden Markov model. Finally, the optimal state sequence of the high-order hidden Markov model is inferred from the optimal state sequence of the equivalent first-order hidden Markov model. This method provides a unified algorithm framework for decoding hidden Markov models including the first-order hidden Markov model and any high-order hidden Markov model.

1. Introduction

Hidden Markov models are powerful tools for modeling and analyzing sequential data. For several decades, hidden Markov models have been used in many fields including handwriting recognition [1–3], speech recognition [4, 5], computational biology [6, 7], and longitudinal data analysis [8, 9]. Past and current developments on hidden Markov models are well documented in [10, 11]. A hidden Markov model comprises an underlying Markov chain and an observed process, where the observed process is a probabilistic function of the underlying Markov chain [12]. Given a hidden Markov model, an efficient procedure for finding the optimal state sequence is of great interest in the real-world applications. In the traditional first-order hidden Markov model, the Viterbi algorithm is utilized to recognize the optimal state sequence [13]. Like the Kalman filter, the Viterbi algorithm tracks the optimal state sequence with a recursive method.

In recent years, the theory and applications of high-order hidden Markov models have been substantially advanced, and high-order hidden Markov models are known to be more

powerful than the first-order hidden Markov model. There are two basic approaches to study the algorithms of high-order hidden Markov models. The first one is called the extended approach, which is to extend directly the existing algorithms of the first-order hidden Markov model to high-order hidden Markov models [14–16]. The second one is called the model reduction method, which is to transform a high-order hidden Markov model to an equivalent first-order hidden Markov model by some means and then to establish the algorithms of the high-order hidden Markov model by using standard techniques applicable to the first-order hidden Markov model [17–20].

In this paper, we propose a novel method for decoding any high-order hidden Markov model. First, the high-order hidden Markov model is transformed into an equivalent first-order hidden Markov model by Hadar's transformation. Next, the optimal state sequence of the equivalent first-order hidden Markov model is recognized by the existing Viterbi algorithm of the first-order hidden Markov model. Finally, the optimal state sequence of the high-order hidden Markov model is inferred from the optimal state sequence of the equivalent first-order hidden Markov model.

2. High-Order Hidden Markov Model and Hadar's Transformation

Initially suppose two processes $\{\tilde{q}_t\}$ and $\{o_t\}$ are defined on some probability space $(\Omega, \mathcal{F}, P)$, where t is an integer index. $\tilde{q}_t$ takes values in a finite set $\tilde{S} = \{s_0, s_1, \dots, s_{N-1}\}$, and o_t takes values in a finite set $\tilde{V} = \{v_1, v_2, \dots, v_M\}$. Without loss of generality, the elements of $\tilde{S}$ can be denoted by $\{0, 1, \dots, N-1\}$. A high-order hidden Markov model is defined as follows.

Definition 1 (see [18, 20]). A high-order hidden Markov model is a doubly stochastic process with an underlying state process that is not directly observable but can be observed only through another stochastic process that is called the observation process. The observation process is governed by the hidden state process and produces the observation sequence. The state process and observation process, respectively, satisfy the following conditions.

- (a) The hidden state process $\{\tilde{q}_t\}$ is a homogeneous Markov chain of order n , that is, a stochastic process that satisfies

$$P(\tilde{q}_t | \{\tilde{q}_l\}_{l < t}) = P(\tilde{q}_t | \{\tilde{q}_l\}_{l=t-n}^{t-1}). \quad (1)$$

- (b) The observation process $\{o_t\}$ is governed by the hidden state process according to a set of probability distributions that satisfy

$$P(o_t | \{o_l\}_{l < t}, \{\tilde{q}_l\}_{l \leq t}) = P(o_t | \{\tilde{q}_l\}_{l=t-(m-1)}^t). \quad (2)$$

To model the high-order hidden Markov model, the following parameters are needed.

- (1) State transition probability distribution:

$$\tilde{a}_{i_n \dots i_1 i_0} = P(\tilde{q}_t = i_0 | \tilde{q}_{t-1} = i_1, \dots, \tilde{q}_{t-n} = i_n). \quad (3)$$

- (2) Symbol emission probability distribution:

$$\tilde{b}_{i_{m-1} \dots i_0}(\ell) = P(o_t = v_\ell | \tilde{q}_t = i_0, \dots, \tilde{q}_{t-(m-1)} = i_{m-1}). \quad (4)$$

- (3) Initial state probability distribution:

$$\tilde{\pi}_{i_r \dots i_1} = P(\tilde{q}_1 = i_1, \dots, \tilde{q}_{2-r} = i_r), \quad (5)$$

where $r = \max\{n, m\}$, $i_0, i_1, \dots, i_r \in \tilde{S}$, and $v_\ell \in \tilde{V}$. We call $\tilde{A} = \{\tilde{a}_{i_n \dots i_1 i_0}\}$, $\tilde{B} = \{\tilde{b}_{i_{m-1} \dots i_0}(\ell)\}$, and $\tilde{\pi} = \{\tilde{\pi}_{i_r \dots i_1}\}$. For convenience, we use the compact notation $\tilde{\lambda} = (\tilde{\pi}, \tilde{A}, \tilde{B})$ to indicate the complete parameters of the high-order hidden Markov model.

Definition 2 (see [18]). Let

$$f: \hat{S} = \{0, 1, \dots, N-1\}^r \longrightarrow S = \{0, 1, \dots, N^r-1\} \quad (6)$$

be the mapping of any base N number to its decimal value; that is, if $[i_0, i_1, \dots, i_{r-1}] \in \hat{S}$, then

$$f([i_0, i_1, \dots, i_{r-1}]) = \sum_{l=0}^{r-1} i_l N^{r-1-l}. \quad (7)$$

Definition 3 (see [17]). Any two models λ_1 and λ_2 are defined as equivalent if

$$P(O | \lambda_1) = P(O | \lambda_2) \quad (8)$$

for any arbitrary observation sequence O . In other words, two models are only considered equivalent if they yield the same likelihood, regardless of the specific observation sequence.

Based on Definition 2, we set

$$q_t = f([\tilde{q}_t, \tilde{q}_{t-1}, \dots, \tilde{q}_{t-(r-1)}]) = \sum_{l=0}^{r-1} \tilde{q}_{t-l} N^{r-1-l}. \quad (9)$$

Since $\tilde{q}_t, \tilde{q}_{t-1}, \dots$, and $\tilde{q}_{t-(r-1)}$ take values in the set $\tilde{S} = \{0, 1, \dots, N-1\}$, q_t takes values in the set $S = \{0, 1, \dots, N^r-1\}$. Moreover, it is easy to see that the inverse transformation can be implemented as follows:

$$\begin{aligned} \tilde{q}_t &= \left\lfloor \frac{q_t}{N^{r-1}} \right\rfloor, \\ \tilde{q}_{t-1} &= \left\lfloor \frac{q_t - \tilde{q}_t N^{r-1}}{N^{r-2}} \right\rfloor, \\ &\vdots \\ \tilde{q}_{t-(r-1)} &= q_t - \tilde{q}_t N^{r-1} - \dots - \tilde{q}_{t-(r-2)} N. \end{aligned} \quad (10)$$

Remark 4. The function f is a one to one correspondence between the set $\hat{S} = \{0, 1, \dots, N-1\}^r$ and the set $S = \{0, 1, \dots, N^r-1\}$.

Proposition 5 (see [18]). Let $a_{ij} = P(q_{t+1} = j | q_t = i)$ for any $i, j \in S$. If $[i/N] \neq j - [j/N^{r-1}]N^{r-1}$, then $a_{ij} = 0$; that is, a transition from $q_t = i$ to $q_{t+1} = j$ is impossible.

Lemma 6. Let $q_t = \sum_{l=0}^{r-1} \tilde{q}_{t-l} N^{r-1-l}$; then the process $\{q_t\}$ forms the first-order homogeneous Markov chain.

Proof. Without loss of generality, we may assume that $q_t = i$ and $q_{t+1} = j$, where $i, j \in S$.

First, we consider the case that $[i/N] \neq j - [j/N^{r-1}]N^{r-1}$. By Proposition 5, it is easy to see that

$$P(q_{t+1} = j | q_t = i, q_{t-1} = \dots) = P(q_{t+1} = j | q_t = i) = 0. \quad (11)$$

Next, we consider the case that $[i/N] = j - [j/N^{r-1}]N^{r-1}$. Since $q_t = i$ and $q_{t+1} = j$, it follows from (10) that

$$\begin{aligned} \tilde{q}_{t+1} &= i_0, \\ \tilde{q}_t &= i_1, \\ &\vdots \\ \tilde{q}_{t+1-(r-1)} &= i_{r-1}, \\ \tilde{q}_{t-(r-1)} &= i_r, \end{aligned} \quad (12)$$

where $i_0, i_1, \dots, i_{r-1}, i_r \in \tilde{S}$. Moreover, we have

$$\begin{aligned}
& P(q_{t+1} = j \mid q_t = i, q_{t-1} = \dots) \\
&= P(\tilde{q}_{t+1} = i_0, \tilde{q}_t = i_1, \dots, \\
&\quad \tilde{q}_{(t+1)-(r-1)} = i_{r-1} \mid \tilde{q}_t = i_1, \dots, \\
&\quad \tilde{q}_{t-(r-2)} = i_{r-1}, \tilde{q}_{t-(r-1)} = i_r, \tilde{q}_{t-r} = \dots) \\
&= P(\tilde{q}_{t+1} = i_0 \mid \tilde{q}_t = i_1, \dots, \\
&\quad \tilde{q}_{t-(n-2)} = i_{n-1}, \tilde{q}_{t-(n-1)} = i_n) \\
&= P(\tilde{q}_{t+1} = i_0 \mid \tilde{q}_t = i_1, \dots, \\
&\quad \tilde{q}_{t-(r-2)} = i_{r-1}, \tilde{q}_{t-(r-1)} = i_r) \\
&= P(\tilde{q}_{t+1} = i_0, \tilde{q}_t = i_1, \dots, \\
&\quad \tilde{q}_{t-(r-2)} = i_{r-1} \mid \tilde{q}_t = i_1, \dots, \\
&\quad \tilde{q}_{t-(r-2)} = i_{r-1}, \tilde{q}_{t-(r-1)} = i_r) \\
&= P(q_{t+1} = j \mid q_t = i).
\end{aligned} \tag{13}$$

Through the above analysis, we derive that

$$P(q_{t+1} \mid \{q_l\}_{l < t+1}) = P(q_{t+1} \mid q_t). \tag{14}$$

Analogously, it is easy to see that

$$P(q_{t+1} = j \mid q_t = i) = P(q_t = j \mid q_{t-1} = i). \tag{15}$$

Therefore, the process $\{q_t\}$ forms the first-order homogeneous Markov chain. $\square$

Lemma 7. *The two processes $\{q_t\}$ and $\{o_t\}$ form the first-order hidden Markov model.*

Proof. Without loss of generality, we may assume that $o_t = v_\ell$ and $q_t = i$, where $v_\ell \in \tilde{V}$ and $i \in S$. Since $q_t = i$, it follows from (10) that

$$\begin{aligned}
& \tilde{q}_t = i_1, \\
& \tilde{q}_{t-1} = i_2, \\
& \vdots \\
& \tilde{q}_{t-(r-1)} = i_r,
\end{aligned} \tag{16}$$

where $i_1, i_2, \dots, i_r \in \tilde{S}$. Moreover, we have

$$\begin{aligned}
& P(o_t = v_\ell \mid \{o_l\}_{l < t}, \{q_l\}_{l < t}, q_t = i) \\
&= P(o_t = v_\ell \mid \{o_l\}_{l < t}, \tilde{q}_t = i_1, \\
&\quad \tilde{q}_{t-1} = i_2, \dots, \tilde{q}_{t-(r-1)} = i_r, \\
&\quad \tilde{q}_{t-(r-2)} = \dots) \\
&= P(o_t = v_\ell \mid \tilde{q}_t = i_1, \\
&\quad \tilde{q}_{t-1} = i_2, \dots, \tilde{q}_{t-(m-1)} = i_m) \\
&= P(o_t = v_\ell \mid \tilde{q}_t = i_1, \\
&\quad \tilde{q}_{t-1} = i_2, \dots, \tilde{q}_{t-(r-1)} = i_r) \\
&= P(o_t = v_\ell \mid q_t = i).
\end{aligned} \tag{17}$$

Analogously, it is easy to see that

$$P(o_t = v_\ell \mid q_t = i) = P(o_{t-1} = v_\ell \mid q_{t-1} = i). \tag{18}$$

Combining these with Lemma 6, we prove that the two processes $\{q_t\}$ and $\{o_t\}$ form the first-order hidden Markov model. $\square$

Remark 8. Hadar and Messer [18] had also mentioned the fact that the two processes $\{q_t\}$ and $\{o_t\}$ form the first-order hidden Markov model, but they did not discuss and prove it in detail.

To model the first-order hidden Markov model $\{q_t, o_t\}$, the following parameters are needed.

(1) State transition probability distribution:

$$a_{ij} = P(q_{t+1} = j \mid q_t = i). \tag{19}$$

(2) Symbol emission probability distribution:

$$b_i(\ell) = P(o_t = v_\ell \mid q_t = i). \tag{20}$$

(3) Initial state probability distribution:

$$\pi_i = P(q_1 = i), \tag{21}$$

where $i, j \in S$ and $v_\ell \in \tilde{V}$. We call $A = \{a_{ij}\}$, $B = \{b_i(\ell)\}$, and $\pi = \{\pi_i\}$. For convenience, we use the compact notation $\lambda = (\pi, A, B)$ to indicate the complete parameters of the first-order hidden Markov model.

Proposition 9 (see [18]). *Let $i = f([i_1, \dots, i_r])$ and $j = f([i_0, \dots, i_{r-1}])$ for any $i_0, i_1, \dots, i_r \in \tilde{S}$; then*

$$\tilde{\pi}_{i_r \dots i_1} = \pi_i, \quad \tilde{a}_{i_r \dots i_0} = a_{ij}, \quad \tilde{b}_{i_r \dots i_1}(\ell) = b_i(\ell). \tag{22}$$

Lemma 10. *Let $O = o_1 \dots o_T$ be any arbitrary observation sequence; then*

$$P(O \mid \tilde{\lambda}) = P(O \mid \lambda). \tag{23}$$

That is, the high-order hidden Markov model $\{\tilde{q}_t, o_t\}$ is equivalent to the first-order hidden Markov model $\{q_t, o_t\}$.

Proof. For any $i_{2-r}, \dots, i_1, \dots, i_{T-(r-1)}, \dots, i_T \in \tilde{S}$, let

$$j_1 = f([i_1, \dots, i_{2-r}]), \dots, j_T = f([i_T, \dots, i_{T-(r-1)}]). \quad (24)$$

By Proposition 9, we have

$$\begin{aligned} P(O | \tilde{\lambda}) &= \sum_{i_{2-r}=0}^{N-1} \cdots \sum_{i_T=0}^{N-1} P(o_1 \cdots o_T, \tilde{q}_{2-r} = i_{2-r}, \dots, \tilde{q}_T = i_T | \tilde{\lambda}) \\ &= \sum_{i_{2-r}=0}^{N-1} \cdots \sum_{i_T=0}^{N-1} \tilde{\pi}_{i_{2-r} \cdots i_1} \tilde{b}_{i_{2-r} \cdots i_1}(o_1) \\ &\quad \times \tilde{a}_{i_{2-r} \cdots i_2} \tilde{b}_{i_{3-r} \cdots i_2}(o_2) \cdots \\ &\quad \times \tilde{a}_{i_{T-r} \cdots i_T} \tilde{b}_{i_{T-m+1} \cdots i_T}(o_T) \\ &= \sum_{i_{2-r}=0}^{N-1} \cdots \sum_{i_T=0}^{N-1} \tilde{\pi}_{i_{2-r} \cdots i_1} \tilde{b}_{i_{2-r} \cdots i_1}(o_1) \\ &\quad \times \tilde{a}_{i_{2-r} \cdots i_2} \tilde{b}_{i_{3-r} \cdots i_2}(o_2) \cdots \\ &\quad \times \tilde{a}_{i_{T-r} \cdots i_T} \tilde{b}_{i_{T-r+1} \cdots i_T}(o_T) \\ &= \sum_{j_1=0}^{N^r-1} \cdots \sum_{j_T=0}^{N^r-1} \pi_{j_1} b_{j_1}(o_1) a_{j_1 j_2} b_{j_2}(o_2) \cdots \\ &\quad \times a_{j_{T-1} j_T} b_{j_T}(o_T) \\ &= \sum_{j_1=0}^{N^r-1} \cdots \sum_{j_T=0}^{N^r-1} P(o_1 \cdots o_T, q_1 = j_1, \dots, q_T = j_T | \lambda) \\ &= P(O | \lambda). \end{aligned} \quad (25)$$

Remark 11. Hadar and Messer [18] had also mentioned the fact that the high-order hidden Markov model $\{\tilde{q}_t, o_t\}$ is equivalent to the first-order hidden Markov model $\{q_t, o_t\}$, but they did not discuss and prove it in detail.

3. Methodology

Theorem 12. Let $O = o_1 \cdots o_T$ be any given observation sequence, and assume that $q_t = f([\tilde{q}_t, \tilde{q}_{t-1}, \dots, \tilde{q}_{t-(r-1)}]) = \sum_{l=0}^{r-1} \tilde{q}_{t-l} N^{r-1-l}$; then

$$P(\tilde{q}_{2-r} \cdots \tilde{q}_T | O, \tilde{\lambda}) = P(q_1 \cdots q_T | O, \lambda). \quad (26)$$

Proof. Without loss of generality, let $\tilde{q}_{2-r} = i_{2-r}, \dots, \tilde{q}_T = i_T$, where $i_{2-r}, \dots, i_T \in \tilde{S}$. According to Proposition 9, we have

$$\begin{aligned} P(o_1 \cdots o_T, \tilde{q}_{2-r} = i_{2-r}, \dots, \tilde{q}_T = i_T | \tilde{\lambda}) &= \tilde{\pi}_{i_{2-r} \cdots i_1} \tilde{b}_{i_{2-r} \cdots i_1}(o_1) \tilde{a}_{i_{2-r} \cdots i_2} \\ &\quad \times \tilde{b}_{i_{3-r} \cdots i_2}(o_2) \cdots \tilde{a}_{i_{T-r} \cdots i_T} \tilde{b}_{i_{T-m+1} \cdots i_T}(o_T) \\ &= \tilde{\pi}_{i_{2-r} \cdots i_1} \tilde{b}_{i_{2-r} \cdots i_1}(o_1) \tilde{a}_{i_{2-r} \cdots i_2} \\ &\quad \times \tilde{b}_{i_{3-r} \cdots i_2}(o_2) \cdots \tilde{a}_{i_{T-r} \cdots i_T} \tilde{b}_{i_{T-r+1} \cdots i_T}(o_T) \\ &= \pi_{j_1} b_{j_1}(o_1) a_{j_1 j_2} b_{j_2}(o_2) \cdots a_{j_{T-1} j_T} b_{j_T}(o_T) \\ &= P(o_1 \cdots o_T, q_1 = j_1, \dots, q_T = j_T | \lambda), \end{aligned} \quad (27)$$

where $j_t = f([i_t, \dots, i_{t-(r-1)}])$ for $1 \leq t \leq T$.

On the other hand, it is easy to see that

$$\begin{aligned} P(\tilde{q}_{2-r} = i_{2-r}, \dots, \tilde{q}_T = i_T | O, \tilde{\lambda}) &= \frac{P(O, \tilde{q}_{2-r} = i_{2-r}, \dots, \tilde{q}_T = i_T | \tilde{\lambda})}{P(O | \tilde{\lambda})}, \\ P(q_1 = j_1, \dots, q_T = j_T | O, \lambda) &= \frac{P(O, q_1 = j_1, \dots, q_T = j_T | \lambda)}{P(O | \lambda)}. \end{aligned} \quad (28)$$

Hence, by Lemma 10, we have

$$P(\tilde{q}_{2-r} \cdots \tilde{q}_T | O, \tilde{\lambda}) = P(q_1 \cdots q_T | O, \lambda). \quad (29)$$

□

Theorem 13. Let $O = o_1 \cdots o_T$ be any given observation sequence, and assume that the state sequence $\tilde{Q}^* = \tilde{q}_{2-r}^* \cdots \tilde{q}_T^*$ satisfies

$$\tilde{Q}^* = \underset{\tilde{Q}}{\operatorname{argmax}} P(\tilde{Q}, O | \tilde{\lambda}); \quad (30)$$

that is, the state sequence $\tilde{Q}^* = \tilde{q}_{2-r}^* \cdots \tilde{q}_T^*$ is some optimal state sequence of the high-order hidden Markov model $\{\tilde{q}_t, o_t\}$. Let $q_t^* = f[\tilde{q}_t^*, \dots, \tilde{q}_{t-(r-1)}^*]$ ($1 \leq t \leq T$); then the state sequence $Q^* = q_1^* \cdots q_T^*$ satisfies

$$P(Q^*, O | \lambda) = \max_Q P(Q, O | \lambda). \quad (31)$$

That is, the state sequence $Q^* = q_1^* \cdots q_T^*$ is some optimal state sequence of the first hidden Markov model $\{q_t, o_t\}$.

Proof. By Theorem 12, it is easy to see that

$$\begin{aligned} P(Q^*, O | \lambda) &= P(\tilde{Q}^*, O | \tilde{\lambda}), \\ \max_Q P(Q, O | \lambda) &= \max_{\tilde{Q}} P(\tilde{Q}, O | \tilde{\lambda}). \end{aligned} \quad (32)$$

Meanwhile, we have the equation

$$P(\bar{Q}^*, O | \bar{\lambda}) = \max_{\bar{Q}} P(\bar{Q}, O | \bar{\lambda}). \quad (33)$$

Hence, we drive that

$$P(Q^*, O | \lambda) = \max_Q P(Q, O | \lambda). \quad (34)$$

□

According to Theorem 13, we can know that some optimal state sequence of the high-order hidden Markov model $\{\tilde{q}_t, o_t\}$ is mapped to some optimal state sequence of the first-order hidden Markov model $\{q_t, o_t\}$. Similarly, we can draw the following conclusion.

Theorem 14. Let $O = o_1 \cdots o_T$ be any given observation sequence, and assume that the state sequence $Q^* = q_1^* \cdots q_T^*$ satisfies

$$Q^* = \operatorname{argmax}_Q P(Q, O | \lambda); \quad (35)$$

that is, the state sequence $Q^* = q_1^* \cdots q_T^*$ is some optimal state sequence of the first-order hidden Markov model $\{q_t, o_t\}$. For $1 \leq t \leq T$, let

$$\begin{aligned} \tilde{q}_t^* &= \left\lfloor \frac{q_t^*}{N^{r-1}} \right\rfloor, \\ \tilde{q}_{t-1}^* &= \left\lfloor \frac{q_t^* - \tilde{q}_t^* N^{r-1}}{N^{r-2}} \right\rfloor, \\ &\vdots \\ \tilde{q}_{t-(r-1)}^* &= q_t^* - \tilde{q}_t^* N^{r-1} - \cdots - \tilde{q}_{t-(r-2)}^* N; \end{aligned} \quad (36)$$

then the state sequence $\bar{Q}^* = \tilde{q}_1^* \cdots \tilde{q}_T^*$ satisfies

$$P(\bar{Q}^*, O | \bar{\lambda}) = \max_{\bar{Q}} P(\bar{Q}, O | \bar{\lambda}). \quad (37)$$

That is, the state sequence $\bar{Q}^* = \tilde{q}_1^* \cdots \tilde{q}_T^*$ is some optimal state sequence of the high-order hidden Markov model $\{\tilde{q}_t, o_t\}$.

Remark 15. Combining Theorem 13 with Theorem 14, it is known that there exists a one to one correspondence between the optimal state sequence of the high-order hidden Markov model $\{\tilde{q}_t, o_t\}$ and the optimal state sequence of the first-order hidden Markov model $\{q_t, o_t\}$.

To decode any high-order hidden Markov model $\{\tilde{q}_t, o_t\}$, we transform it into an equivalent first-order hidden Markov model $\{q_t, o_t\}$ by Hadar's transformation. Then, do as follows.

Step 1. Determine some optimal state sequence $Q^* = q_1^* q_2^* \cdots q_T^*$ of the first-order hidden Markov model $\{q_t, o_t\}$ by using the Viterbi algorithm. Without loss of generality, let

$$\begin{aligned} q_1^* &= j_1, \\ q_2^* &= j_2, \\ &\vdots \\ q_T^* &= j_T, \end{aligned} \quad (38)$$

where $j_1, j_2, \dots, j_T \in S$.

Step 2. For $q_T^* = j_T$, using the transformations

$$\begin{aligned} \tilde{q}_t &= \left\lfloor \frac{q_t}{N^{r-1}} \right\rfloor, \\ \tilde{q}_{t-1} &= \left\lfloor \frac{q_t - \tilde{q}_t N^{r-1}}{N^{r-2}} \right\rfloor, \\ &\vdots \\ \tilde{q}_{t-(r-1)} &= q_t - \tilde{q}_t N^{r-1} - \cdots - \tilde{q}_{t-(r-2)} N, \end{aligned} \quad (39)$$

we have

$$\begin{aligned} \tilde{q}_T^* &= \left\lfloor \frac{j_T}{N^{r-1}} \right\rfloor, \\ \tilde{q}_{T-1}^* &= \left\lfloor \frac{j_T - \tilde{q}_T^* N^{r-1}}{N^{r-2}} \right\rfloor, \\ &\vdots \\ \tilde{q}_{T-(r-2)}^* &= \left\lfloor \frac{j_T - \sum_{l=0}^{r-3} \tilde{q}_{T-l}^* N^{r-1-l}}{N} \right\rfloor, \\ \tilde{q}_{T-(r-1)}^* &= j_T - \sum_{l=0}^{r-2} \tilde{q}_{T-l}^* N^{r-1-l}. \end{aligned} \quad (40)$$

Step 3. For $q_1^* = j_1, q_2^* = j_2, \dots, q_{T-1}^* = j_{T-1}$, using the transformation

$$q_t = \sum_{l=0}^{r-1} \tilde{q}_{t-l} N^{r-1-l}, \quad (41)$$

we have

$$\begin{aligned} \tilde{q}_{T-r}^* &= j_{T-1} - \sum_{l=1}^{r-1} \tilde{q}_{T-l}^* N^{r-l}, \\ \tilde{q}_{T-(r+1)}^* &= j_{T-2} - \sum_{l=1}^{r-1} \tilde{q}_{T-1-l}^* N^{r-l}, \\ &\vdots \\ \tilde{q}_{2-r}^* &= j_1 - \sum_{l=1}^{r-1} \tilde{q}_{2-l}^* N^{r-l}. \end{aligned} \quad (42)$$

Combining Step 2 with Step 3, we obtain the state sequence

$$\bar{Q}^* = \bar{q}_{2-r}^* \bar{q}_{3-r}^* \cdots \bar{q}_{T-1}^* \bar{q}_T^*. \quad (43)$$

According to Theorem 14, the above state sequence $\bar{Q}^* = \bar{q}_{2-r}^* \bar{q}_{3-r}^* \cdots \bar{q}_{T-1}^* \bar{q}_T^*$ is some optimal state sequence of the high-order hidden Markov model $\{\bar{q}_t, o_t\}$.

4. Conclusions

In this paper, a novel method for decoding any high-order hidden Markov model is given. Based on this method, the optimal state sequence of any high-order hidden Markov model can be inferred by the existing Viterbi algorithm of the first-order hidden Markov model. This method has universal character for decoding hidden Markov models and provides a unified algorithm framework for decoding hidden Markov models including the first-order hidden Markov model and any high-order hidden Markov model. For instance, the Viterbi algorithm of the first-order hidden Markov model can be easily derived as a special case of our conclusion when $m = n = 1$.

This method we analyzed is practical and valuable in its own right. Future research could use this method for applications in handwriting, speech recognition, speaker recognition, emotion recognition, and so forth.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is supported by the Major Program of the National Natural Science Foundation of China (no. 71390521), the Postdoctoral Science Foundation of China (no. 2014M551565), and the Scientific Research Foundation of Tongling University (no. 2012tlxyrc04).

References

- [1] J. Hu, M. K. Brown, and W. Turin, "HMM based on-line handwriting recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 10, pp. 1039–1045, 1996.
- [2] M. S. Khorsheed, "Recognising handwritten Arabic manuscripts using a single hidden Markov model," *Pattern Recognition Letters*, vol. 24, no. 14, pp. 2235–2242, 2003.
- [3] T. Artières, S. Marukat, and P. Gallinari, "Online handwritten shape recognition using segmental hidden markov models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 205–217, 2007.
- [4] B. H. Juang and L. R. Rabiner, "Hidden Markov models for speech recognition," *Technometrics*, vol. 33, no. 3, pp. 251–272, 1991.
- [5] M. Gales and S. Young, "The application of hidden Markov Models in speech recognition," *Foundations and Trends in Signal Processing*, vol. 1, no. 3, pp. 195–304, 2008.
- [6] A. Löytynoja and M. C. Milinkovitch, "A hidden Markov model for progressive multiple alignment," *Bioinformatics*, vol. 19, no. 12, pp. 1505–1513, 2003.
- [7] L. Regad, F. Guyon, J. Maupetit, P. Tufféry, and A. C. Camproux, "A Hidden Markov Model applied to the protein 3D structure analysis," *Computational Statistics and Data Analysis*, vol. 52, no. 6, pp. 3198–3207, 2008.
- [8] R. M. Altman, "Mixed hidden Markov models: an extension of the hidden Markov model to the longitudinal data setting," *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 201–210, 2007.
- [9] A. Spagnoli, R. Henderson, R. J. Boys, and J. J. Houwing-Duistermaat, "A hidden Markov model for informative dropout in longitudinal response data with crisis states," *Statistics and Probability Letters*, vol. 81, no. 7, pp. 730–738, 2011.
- [10] Y. Ephraim and N. Merhav, "Hidden Markov processes," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1518–1569, 2002.
- [11] J. A. Bilmes, "What HMMs can do," *IEICE Transactions on Information and Systems*, vol. E89-D, no. 3, pp. 869–891, 2006.
- [12] L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *The Annals of Mathematical Statistics*, vol. 37, no. 6, pp. 1554–1563, 1966.
- [13] L. R. Rabiner and B.-H. Juang, "An introduction to hidden Markov models," *IEEE ASSP Magazine*, vol. 3, no. 1, pp. 4–16, 1986.
- [14] J.-F. Mari, J.-P. Haton, and A. Kriouile, "Automatic word recognition based on second-order hidden Markov models," *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 1, pp. 22–25, 1997.
- [15] J.-F. Mari and F. Le Ber, "Temporal and spatial data mining with second-order hidden markov models," *Soft Computing*, vol. 10, no. 5, pp. 406–414, 2006.
- [16] L.-M. Lee, "High-order hidden markov model and application to continuous mandarin digit recognition," *Journal of Information Science and Engineering*, vol. 27, no. 6, pp. 1919–1930, 2011.
- [17] J. A. du Preez, "Efficient training of high-order hidden Markov models using first-order representations," *Computer Speech and Language*, vol. 12, no. 1, pp. 23–39, 1998.
- [18] U. Hadar and H. Messer, "High-order hidden Markov models—estimation and implementation," in *Proceedings of the 15th IEEE/SP Workshop on Statistical Signal Processing (SSP '09)*, pp. 249–252, Cardiff, Wales, September 2009.
- [19] H. A. Engelbrecht and J. A. du Preez, "Efficient backward decoding of high-order hidden Markov models," *Pattern Recognition*, vol. 43, no. 1, pp. 99–112, 2010.
- [20] F. Ye, N. Yi, and Y. F. Wang, "EM algorithm for training high-order hidden Markov model with multiple observation sequences," *Journal of Information and Computational Science*, vol. 8, no. 10, pp. 1761–1777, 2011.