The Interactive Hidden Markov Model with real Practical applied

Ibrahim Z. Chaloob

Ministry of higher Education & Scientific Research

2008-2009

نموذج ماركوف المخفي التفاعل بتطبيق عملي حقيقي

ابراهيم زغيتون جلوب

وزارة التعليم العالي والبحث العلمي

2009-2008

الملخص

نقترح ان نموذج ماركوف المخفي التفاعلي بحالات الانتقال المخفية تعتمد على حالات المشاهدات الحالية، والنموذج (IHHM) هي من النوع العام لـ(HMM)، ونلاحظ ان النوع العام من النموذج (HMM) يختلف عن النوع التقليدية (HMMs)، حيث الحالات المخفية المستقبلية يسيطر عليها من قبل الحالات المخفية السابقة فقط. وسوف نعطي مثالاً للبرهان على النوع (IHMM)، وسنتسع في النتائج للوصول الى النوع العام من (IHMM).

Abstract

I proposed an Interactive Hidden Markov Model (IHMM) where the transitions of hidden states depend on the current observable states. The IHHM is a generalization of the HMM. I note that this kind of HMM is different from classical HMMs where the next hidden states are governed by the previous hidden states only. An example is given to demonstrate IHMM. I'll extend the results to give a general IHMM.

Introduction

Hidden Markov Models (HMMs) are widely used in bioinformatics [1], speech recognition [2] and many other areas [3]. In a HMM, there are two types of states: the observable states and the hidden states. In a HMM, there is no one-to-one correspondence between the hidden states and the observed symbols. It is therefore no longer possible to tell what hidden state the model is in which the observation symbol is generated just by looking at the observation symbol.

Hidden Markov Models

A standard HMM is usually characterized by the following elements [2]:

(i) *N*, the number of states in the model. Although the states are hidden, for many practical applications, very often, there is physical significance to the states. We denote the individual states as

$\quad$ **S** = {S1, S2, . . . , SN}, and the state at the length *t* as *qt*.

(ii) *M*, the number of distinct observation symbols (or state) for the hidden states. The observation symbols correspond to the physical output of the system being modeled. We denote the individual symbols as

$$V = \{v1, v2, \cdots, vM\}$$

(iii) The state transition probability distribution $\quad$ A = {aij}

$\quad$ Where:

$$a_{ij} = P(q_{t+1} = S_j | q_t = S_i), \quad 1 \le i, j \le N.$$

(iv) The observation probability distribution in state *j*, B = {bj(k)}, where

$$b_j(k) = P(O_t = v_k | q_t = S_j), \quad 1 \le j \le N, 1 \le k \le M.$$

v) The initial state distribution *Π = {πi}* where

$$\pi_i = P(q_1 = S_i), \quad 1 \le i \le N.$$

Given appropriate values of *N,M, A,B* and *Π*, the HMM can be used as a generator to give an observation sequence $\quad$ *O = O₁O₂ . . . Oτ*

where each observation O_T is one of the symbols from **V**, and *T* is the number of observations in the sequence. For simplicity, we use the compact notation

$$Λ = (A,B,Π), \quad O = \{O_1O_2O_3 \cdots O_T\}$$

To indicate the complete parameter set of the HMM. According to the above specification, very often a first order Markov process is used in modeling the transitions among the hidden states in a HMM.

Higher-order HMMs

There are three key issues in HHMMs:

Problem 1: Given the observation sequence $O = \{O1O2 \cdots OT\}$ and a HMM, how to efficiently compute the probability of the observation sequence ?

• **Problem 2:** Given the observation sequence $O = \{O1O2 \cdots OT\}$ and a HMM, how to choose a corresponding state sequence $Q = \{Q1Q2 \cdots QT\}$ which is optimal in certain sense?

• **Problem 3:** Given the observation sequence $O = \{O1O2 \cdots OT\}$, how to choose the model parameters in a HMM?

For **Problem 1**, a forward-backward dynamic programming procedure [4] is formulated to calculate the probability of the observation sequence efficiently.

For **Problem 2**, it is the one in which we attempt to uncover the hidden part of the model, i.e., to find the "correct" state sequence. In many practical situations, we use optimality criteria to solve the problem as good as possible.

The most widely used criterion is to find a single best state sequence, i.e., maximize the likelihood $P(Q|\Lambda,O)$. This is equivalent to maximizing $P(Q,O|\Lambda)$ since

$$P(Q|\Lambda, O) = \frac{P(Q, O|\Lambda)}{P(O|\Lambda)}.$$

Viterbi algorithm [5] is a dynamic programming technique for finding this single best state sequence

$$Q = \{Q1, Q2, \cdots, QT\}$$

for the given observation sequence

$$O = \{O1, O2, \cdots, OT\}.$$

For **Problem 3**, we attempt to adjust the model parameters Λ such that $P(O|\Lambda)$ is maximized by using Expectation-Maximization (EM) algorithm.

For a complete tutorial on hidden Markov model, we refer readers to the paper by Rabiner [2] and the book by MacDonald and Zucchini [3].

The DNA Sequence

In the DNA sequence analysis, higher-order Markov models have been used to model the transitions among the observable states, see [6, 7]. An mth order Markov process is a stochastic process where each event depends on the previous m events. It is believed that higher-order Markov model (in the hidden layer) can better capture a number of data sequences such as the DNA sequences. The main aim of this paper is to develop higher-order HMMs (higher-order Markov model for the hidden states). The main difference between the traditional HMM and a higher-order HMM is that in the hidden layer, the state transition probability is governed by the mth order higher-order Markov model

$$a_{i_{t-m+1},\ldots,i_{t+1}} = P\big(q_{t+1} = S_{i_{t+1}}\big| q_t = S_{i_t}, \ldots, q_{t-m+1} = S_{i_{t-m+1}}\big).$$

We assume that the distribution Π of initial m states is given by

$$\pi_{i_1,i_2,\ldots,i_m} = P(q_1 = S_{i_1}, q_2 = S_{i_2}, \ldots, q_m = S_{i_m}).$$

In order to determine whether certain short DNA sequence (a categorical data sequence of four possible categories: A,C,G and T) occurred more often than would be expected by chance, Avery [8] examined the Markovian structure of introns from several other genes in mice. Here we apply our model to the introns from the mouse αA-crystallin gene see for instance. We compare our second-order model with the Raftery's second-order model. The model parameters of the Raftery's model are given in [9].

The results are reported in Table 1.

Table 1. Prediction accuracy in the DNA sequence.

	2-state model	3-state model	4-state model
New Model	0.57	0.49	0.33
Raftery's Model	0.57	0.47	0.31
Random Chosen	0.50	0.33	0.25

The comparison is made with different grouping of states as suggested in [175]. In grouping states 1 and 3, and states 2 and 4 we have a 2-state model.

Our model gives

$$\hat{Q}_1 = \begin{pmatrix} 0.5568 & 0.4182 \\ 0.4432 & 0.5818 \end{pmatrix},$$

$$\hat{Q}_2 = \begin{pmatrix} 0.4550 & 0.5149 \\ 0.5450 & 0.4851 \end{pmatrix}$$

$$\hat{X} = (0.4858, 0.5142)^T, \quad \lambda_1 = 0.7529 \quad \text{and} \quad \lambda_2 = 0.2471.$$

In grouping states 1 and 3 we have a 3-state model. Our model gives

$$\hat{Q}_1 = \begin{pmatrix} 0.5568 & 0.3573 & 0.4949 \\ 0.2571 & 0.3440 & 0.2795 \\ 0.1861 & 0.2987 & 0.2256 \end{pmatrix},$$

$$\hat{Q}_2 = \begin{pmatrix} 0.4550 & 0.5467 & 0.4747 \\ 0.3286 & 0.2293 & 0.2727 \\ 0.2164 & 0.2240 & 0.2525 \end{pmatrix}$$

$$\hat{X} = (0.4858, 0.2869, 0.2272)^T, \quad \lambda_1 = 1.0 \quad \text{and} \quad \lambda_2 = 0.0$$

If there is no grouping, we have a 4-state model. Our model gives

$$\hat{X} = (0.2395, 0.2869, 0.2464, 0.2272)^T, \quad \lambda_1 = 0.253 \quad \text{and} \quad \lambda_2 = 0.747.$$

$$\hat{Q}_1 = \begin{pmatrix} 0.2268 & 0.2987 & 0.2274 & 0.1919 \\ 0.2492 & 0.3440 & 0.2648 & 0.2795 \\ 0.3450 & 0.0587 & 0.3146 & 0.3030 \\ 0.1789 & 0.2987 & 0.1931 & 0.2256 \end{pmatrix},$$

$$\hat{Q}_2 = \begin{pmatrix} 0.1891 & 0.2907 & 0.2368 & 0.2323 \\ 0.3814 & 0.2293 & 0.2773 & 0.2727 \\ 0.2532 & 0.2560 & 0.2305 & 0.2424 \\ 0.1763 & 0.2240 & 0.2555 & 0.2525 \end{pmatrix}$$

When using the expected errors (assuming that the next state is randomly chosen with equal probability for all states) as a reference, the percentage gain in effectiveness of using higher-order Markov chain models is in the 3-state model. In this case, our model also gives a better estimation when compared with Raftery's model. Raftery [10].

I consider use the BIC to weight efficiency gain in terms of extra parameters used. This is important in his approach since his method requires solving a highly non-linear optimization problem.

The complexity of solving the optimization problem increases when there are many parameters to be estimated.

The Interactive Hidden Markov Model

The IHHM is a generalization of the HMM .We note that this kind of HMM is different from classical HMMs where the next hidden states are governed by the previous hidden states only. An example is given to demonstrate IHMM. We then extend the results to give a general IHMM.

Example (case study)

Suppose that we are given a categorical data sequence (in steady state) of volume of transactions as follows:

1, 2, 1, 2, 1, 2, 2, 4, 1, 2, 2, 1, 3, 3, 4, 1.

Here 1=high transaction volume, 2= medium transaction volume, 3=low transaction volume and 4=very low transaction volume.

Suppose there are two hidden states: A (bull market period) and B (bear market period).

In period A, the probability distribution of the transaction volume is assumed to follow

(1/4, 1/4, 1/4, 1/4).

In period B, the probability distribution of the transaction volume is assumed to follow

(1/6, 1/6, 1/3, 1/3).

In the proposed model, we assume that hidden states are unobservable but the transaction volumes are observable. We would like to uncover the hidden state by modeling the dynamics by a Markov chain.

In the Markov chain, the states are   A,B, 1, 2, 3, 4.

We assume that when the observable state is i then the probabilities that the hidden state is A and B are given by αi and $(1 - \alpha i)$, (depending on i) respectively in next time step. The transition probability matrix governing the Markov chain is given by:

$$P_1 = \begin{pmatrix} 0 & 0 & 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 1/6 & 1/6 & 1/3 & 1/3 \\ \alpha_1 & 1-\alpha_1 & 0 & 0 & 0 & 0 \\ \alpha_2 & 1-\alpha_2 & 0 & 0 & 0 & 0 \\ \alpha_3 & 1-\alpha_3 & 0 & 0 & 0 & 0 \\ \alpha_4 & 1-\alpha_4 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Estimation of Parameters

In order to define the IHMM, one has to estimate the model parameters $\alpha 1$, $\alpha 2$, $\alpha 3$ and $\alpha 4$ from an observed data sequence. One may consider the following two-step transition probability matrix as follows:

$$P_1^2 = \begin{pmatrix} \frac{\alpha_1+\alpha_2+\alpha_3+\alpha_4}{4} & 1 - \frac{\alpha_1+\alpha_2+\alpha_3+\alpha_4}{4} & 0 & 0 & 0 & 0 \\ \frac{\alpha_1+\alpha_2}{6}+\frac{\alpha_3+\alpha_4}{3} & 1-\frac{\alpha_1+\alpha_2}{6}-\frac{\alpha_3+\alpha_4}{3} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{6}+\frac{\alpha_1}{12} & \frac{1}{6}+\frac{\alpha_1}{12} & \frac{1}{3}-\frac{\alpha_1}{12} & \frac{1}{3}-\frac{\alpha_1}{12} \\ 0 & 0 & \frac{1}{6}+\frac{\alpha_2}{12} & \frac{1}{6}+\frac{\alpha_2}{12} & \frac{1}{3}-\frac{\alpha_2}{12} & \frac{1}{3}-\frac{\alpha_2}{12} \\ 0 & 0 & \frac{1}{6}+\frac{\alpha_3}{12} & \frac{1}{6}+\frac{\alpha_3}{12} & \frac{1}{3}-\frac{\alpha_3}{12} & \frac{1}{3}-\frac{\alpha_3}{12} \\ 0 & 0 & \frac{1}{6}+\frac{\alpha_4}{12} & \frac{1}{6}+\frac{\alpha_4}{12} & \frac{1}{3}-\frac{\alpha_4}{12} & \frac{1}{3}-\frac{\alpha_4}{12} \end{pmatrix}.$$

One can extract the one-step transition probability matrix of the observable states from P_2^2 as follows:

$$\tilde{P}_2 = \begin{pmatrix} \frac{1}{6}+\frac{\alpha_1}{12} & \frac{1}{6}+\frac{\alpha_1}{12} & \frac{1}{3}-\frac{\alpha_1}{12} & \frac{1}{3}-\frac{\alpha_1}{12} \\ \frac{1}{6}+\frac{\alpha_2}{12} & \frac{1}{6}+\frac{\alpha_2}{12} & \frac{1}{3}-\frac{\alpha_2}{12} & \frac{1}{3}-\frac{\alpha_2}{12} \\ \frac{1}{6}+\frac{\alpha_3}{12} & \frac{1}{6}+\frac{\alpha_3}{12} & \frac{1}{3}-\frac{\alpha_3}{12} & \frac{1}{3}-\frac{\alpha_3}{12} \\ \frac{1}{6}+\frac{\alpha_4}{12} & \frac{1}{6}+\frac{\alpha_4}{12} & \frac{1}{3}-\frac{\alpha_4}{12} & \frac{1}{3}-\frac{\alpha_4}{12} \end{pmatrix}.$$

However, in this case, we do not have a closed form solution for the stationary distribution of the process. To estimate the parameter αi, we first estimate the one-step transition probability matrix from the observed sequence.

This can be done by counting the transition of the states in the observed sequence and we have frequencies:

$$\hat{P}_2 = \begin{pmatrix} 0 & \frac{4}{5} & \frac{1}{5} & 0 \\ \frac{1}{2} & \frac{1}{3} & 0 & \frac{1}{6} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

We expect that

$$\tilde{P}_2 \approx \hat{P}_2$$

And hence αi can be obtained by solving the following minimization problem:

$$\min_{\alpha_i} ||\tilde{P}_2 - \hat{P}_2||_F^2$$

…………..(1)

Subject to $\quad 0 \leq \alpha_i \leq 1.$

Here $||.||_F$ is the Frobenius norm, i.e.

$$||A||_F^2 = \sum_{i=1}^{n} \sum_{i=1}^{n} A_{ij}^2.$$

This is equivalent to solve the following four independent minimization problems (i) - (iv) and they can be solved in parallel. This is an advantage of the estimation method.

We remark that one can also consider other matrix norms for the objective function (1), let us say $||.||_{M1}$ or $||.||_{M\infty}$ and they may result in linear programming problems.

(i) α_1: $\displaystyle\min_{0 \leq \alpha_1 \leq 1} \{(\frac{1}{6} + \frac{\alpha_1}{12})^2 + (\frac{1}{6} + \frac{\alpha_1}{12} - \frac{4}{5})^2 + (\frac{1}{3} - \frac{\alpha_1}{12} - \frac{1}{5})^2 + (\frac{1}{3} - \frac{\alpha_1}{12})^2\}$;

(ii) α_2: $\displaystyle\min_{0 \leq \alpha_2 \leq 1} \{(\frac{1}{6} + \frac{\alpha_1}{12} - \frac{1}{2})^2 + (\frac{1}{6} + \frac{\alpha_1}{12} - \frac{1}{3})^2 + (\frac{1}{3} - \frac{\alpha_1}{12})^2 + (\frac{1}{3} - \frac{\alpha_1}{12} - \frac{1}{6})^2\}$;

(iii) α_3: $\displaystyle\min_{0 \leq \alpha_3 \leq 1} \{(\frac{1}{6} + \frac{\alpha_1}{12})^2 + (\frac{1}{6} + \frac{\alpha_1}{12})^2 + (\frac{1}{3} - \frac{\alpha_1}{12} - \frac{1}{2})^2 + (\frac{1}{3} - \frac{\alpha_1}{12} - \frac{1}{2})^2\}$;

(iv) α_4: $\displaystyle\min_{0 \leq \alpha_4 \leq 1} \{(\frac{1}{6} + \frac{\alpha_1}{12} - 1)^2 + (\frac{1}{6} + \frac{\alpha_1}{12})^2 + (\frac{1}{3} - \frac{\alpha_1}{12})^2 + (\frac{1}{3} - \frac{\alpha_1}{12})^2\}.$

Solving the above optimization problems, we have

$$\alpha_1^* = 1, \quad \alpha_2^* = 1, \quad \alpha_3^* = 0, \quad \alpha_4^* = 1.$$

Hence we have

$$P_2 = \begin{pmatrix} 0 & 0 & 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 1/6 & 1/6 & 1/3 & 1/3 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

And

$$P_2^2 = \begin{pmatrix} 3/4 & 1/4 & 0 & 0 & 0 & 0 \\ 2/3 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 1/6 & 1/6 & 1/3 & 1/3 \\ 0 & 0 & 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}.$$

Extension to the General Case

The method can be extended to a general case of m hidden states and n observable states. I note the one-step transition probability matrix of the observable states is given by

$$\tilde{P}_2 = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1m} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{n1} & \alpha_{m2} & \cdots & \alpha_{nm} \end{pmatrix} \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mn} \end{pmatrix},$$

i.e

$$[\tilde{P}_2]_{ij} = \sum_{k=1}^{m} \alpha_{ik} p_{kj} \quad i, j = 1, 2, \ldots, n.$$

Here we assume that α_{ij} are unknowns and the probabilities p_{ij} are given.

Suppose $[Q]_{ij}$ is the one-step transition probability matrix estimated from the observed sequence. Then for each fixed i, α_{ij}, $j = 1, 2, \ldots, m$ can be obtained by solving the following constrained least squares problem:

$$\min_{\alpha_{ik}} \left\{ \sum_{j=1}^{n} \left(\sum_{k=1}^{m} \alpha_{ik} p_{kj} - [Q]_{ij} \right)^2 \right\}$$

Subject to

$$\sum_{k=1}^{m} \alpha_{ik} = 1$$

And

$$\alpha_{ik} \geq 0 \quad \text{for all } i, k.$$

The idea of the IHMM presented in this paper is further extended to address the following applications and problems in Ching et al. [11].

(i) IHMM is applied to some practical data sequences in sales demand data sequences.

(ii) there are only a few works on modeling the non-linear behavior of categorical time series can be found in literature. In the continuous-state case, the threshold auto- regressive model is a well-known approach.

The idea is to provide a piecewise linear approximation to a non-linear autoregressive time series model by dividing the state space into several regimes via threshold principle.

The IHMM provides a first-order approximation of the non-linear behavior of categorical time series by dividing the state space of the Markov chain process into several regimes.

Reference

1. Koski T (2001) Hidden Markov Models for Bioinformatics, Kluwer Academic Publisher, Dordrecht.

2. Rabiner L (1989) A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Proceedings of the IEEE, 77:257–286.

3. MacDonald I and Zucchini W (1997) Hidden Markov and Other Models for Discrete-valued Time Series, Chapman & Hall, London.

4. Baum L (1972) An Inequality and Associated Maximization Techniques in statistical Estimation for Probabilistic Function of Markov Processes, Inequality,3:1–8.
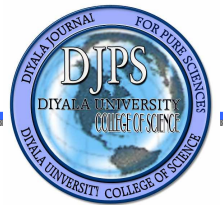
5. Viterbi A (1967) Error Bounds for convolution Codes and an Asymptotically Optimum Decoding Algorithm, IEEE Transaction on Information theory,13:260–269.

6. Borodovskii M, Sprizhitskii A, Golovanov I and Aleksandrov A (1986) Statistical Patterns in Primary Structures of the Functional Regions of Genome in Escherichia coli-, Molecular Biology, 20:826–833.

7. Salzberg S, Delcher S, Kasif S and White O (1998) Microbial gene identification using interpolated Markov models, Nuclei Acids Research, 26:544–548.

8. Avery P (1987) The Analysis of Intron Data and Their Use in the Detection of Short Signals, Journal of Molecular Evolution, 26:335–340.

9. Raftery A and Tavare S (1994) Estimation and Modelling Repeated Patterns in High Order Markov Chains with the Mixture Transition Distribution Model,Journal of Applied Statistics, 43: 179–199.

10.Raftery A (1985) A Model for High-order Markov Chains, Journal of Royal Statistical Society, Series B, 47:528–539.

11. Ching W, Ng M, Fung E and Siu T (2005) An Interactive Hidden Markov Model for Categorical Data Sequences.