



東京大学 工学部 計数工学科/物理工学科

応用音響学：Forward アルゴリズム

嵯峨山 茂樹 <sagayama@hil.t.u-tokyo.ac.jp>

東京大学 工学部 計数工学科

<http://hil.t.u-tokyo.ac.jp/~sagayama/applied-acoustics/>

参考文献

- 中川 聖一 「確率モデルによる音声認識」 コロナ社
- 古井 貞熙 「音声情報処理」 森北出版
- 谷萩 隆嗣 「音声と画像のデジタル信号処理」 コロナ社



HMMの音声認識への応用

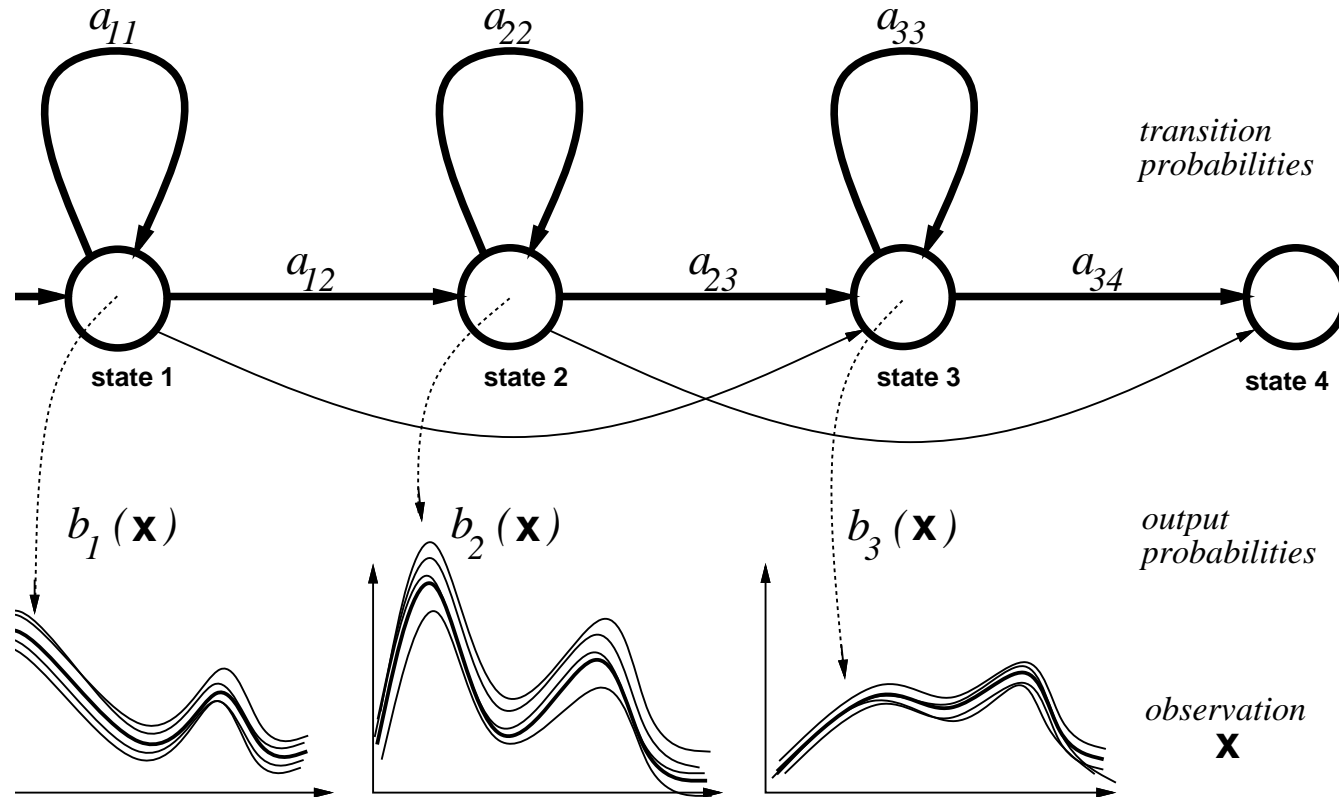
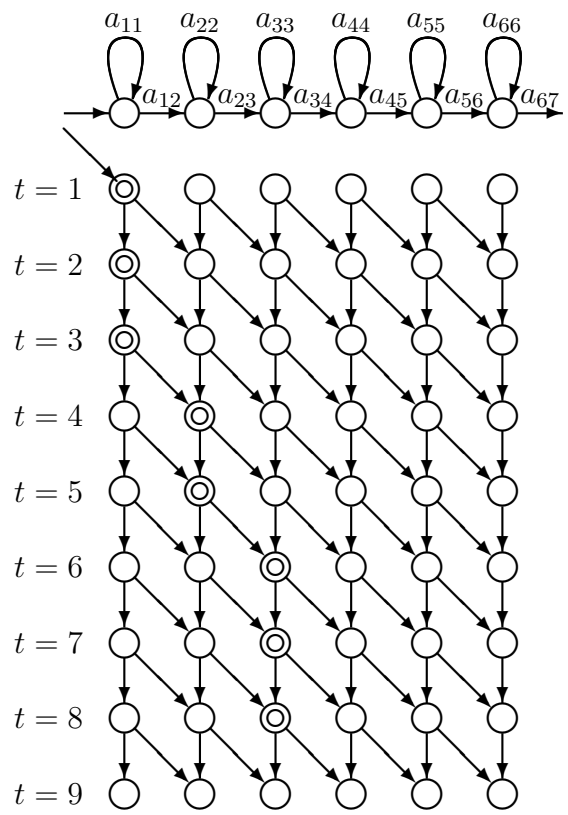


図1. 音声認識のためのHMMの意味



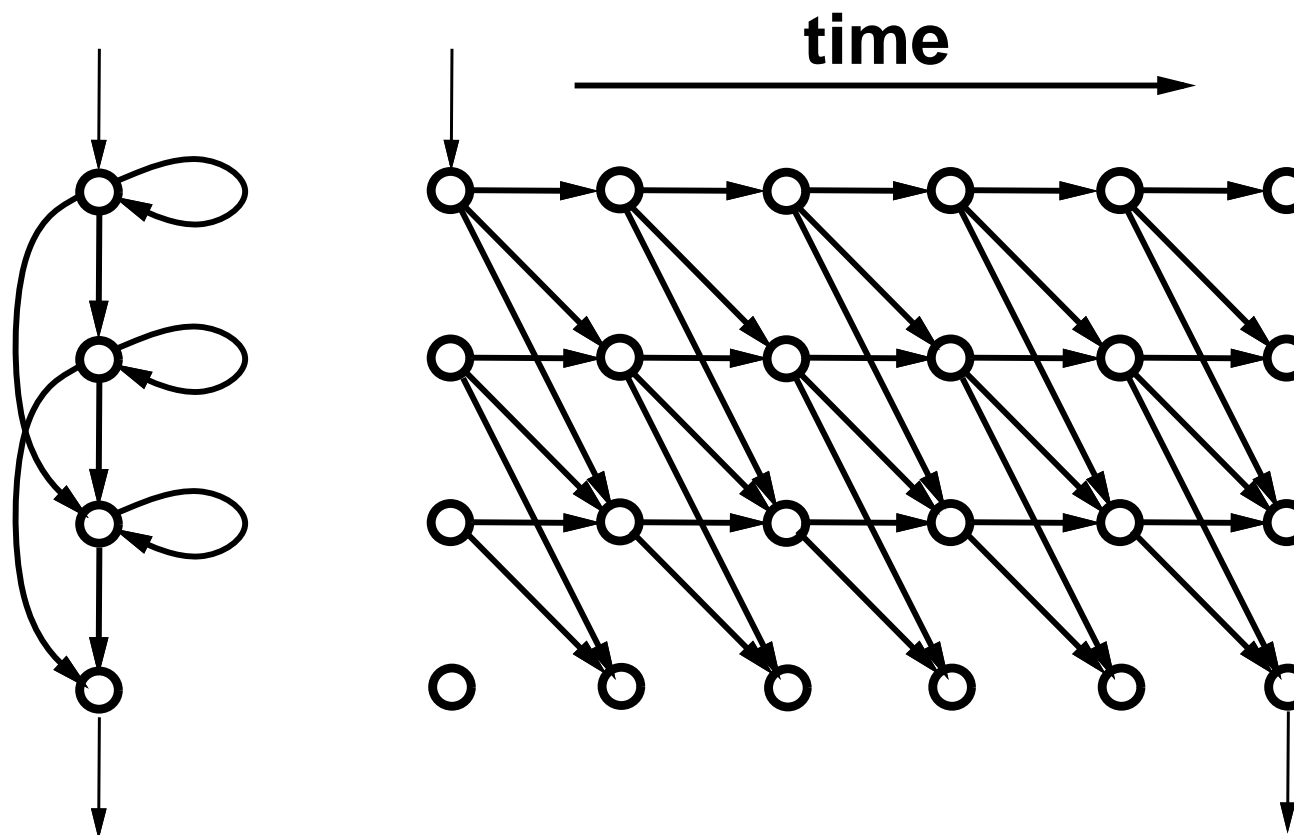
音声認識におけるHMMの利点

- 確率理論に支えられていて理論に乗りやすいこと、
- 尤度(あるいは他の規準)を最大にするようなパラメータ推定を行なう学習法が存在すること、
- 標準パターンではなく分布を表現していること、
- 状態の概念により時間を適度に粗く表現していること、
- フレームごとの入力音声の尤度が得られるのでその積により文全体の尤度が得られること





HMMの計算 1 (確率評価): Forward アルゴリズム



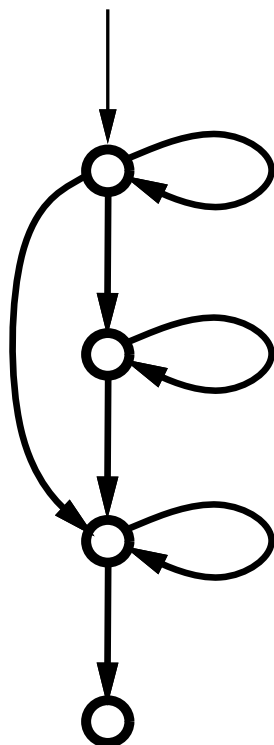
Forward Algorithm

図2. 確率評価: Forward アルゴリズム



HMMの計算 2 (最適経路決定): Viterbi アルゴリズム

structure



trellis

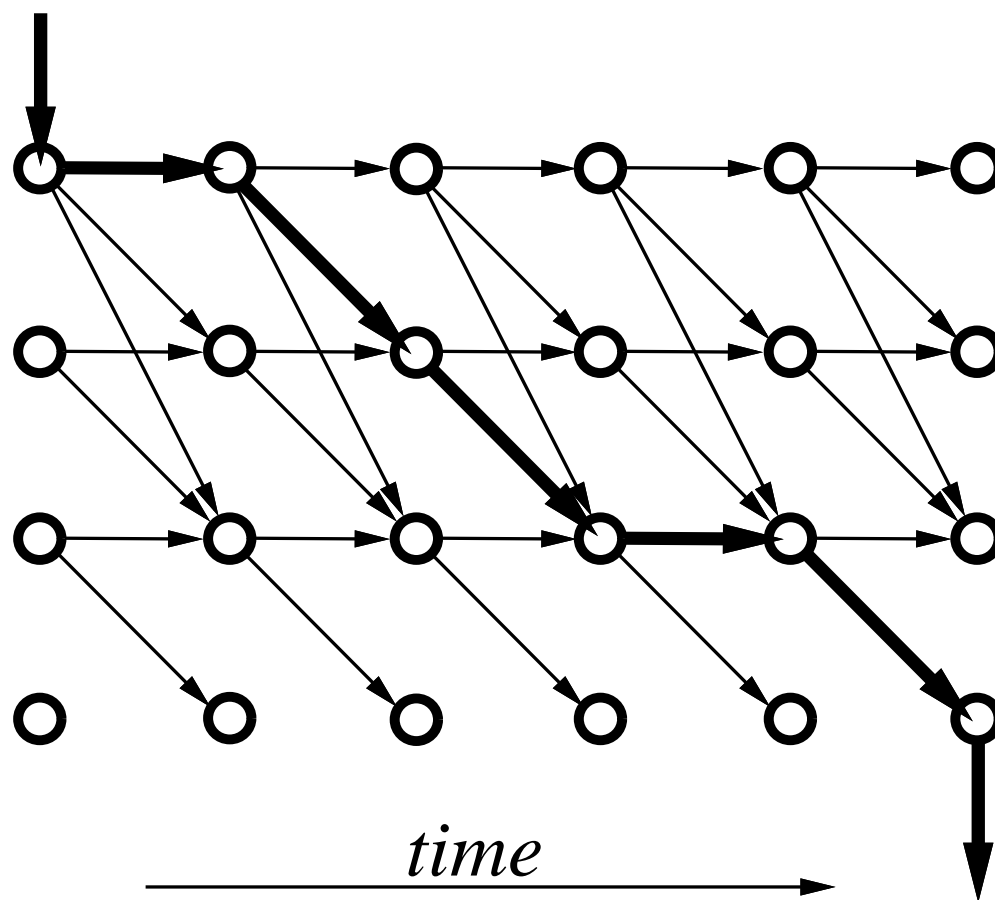


図3. 最適経路決定: Viterbi アルゴリズム



Hidden Markov Model (HMM)

HMMの特徴

- 一般的に大きな変動のある特徴ベクトルの時系列を確率モデルで表現。
- 各単語や音素を標準的な時系列を、DPのように標準パターンとして用いず、確率状態遷移機械(マルコフモデル)で表現する。
- 事後確率 $P(Y|\lambda)$ を求めることを目的とした方法。

HMMの利点

- 特徴ベクトルの時系列パターンの統計的変動を最も吸収するように、モデルのパラメータを推定できる。
- 認識における計算量が少ない。

HMMの問題点

- モデルの設計法が確立されていない。
- 統計的モデルであるので、モデルのパラメータ推定にある程度のサンプルを必要とし、計算量も多い。



隠れマルコフモデル (Hidden Markov Model: HMM)

- 統計的な性質が異なる複数の定常信号源： $b_i(y_t)$
- ある順序で確率的(マルコフ的)に信号源を切替え： a_{ij}

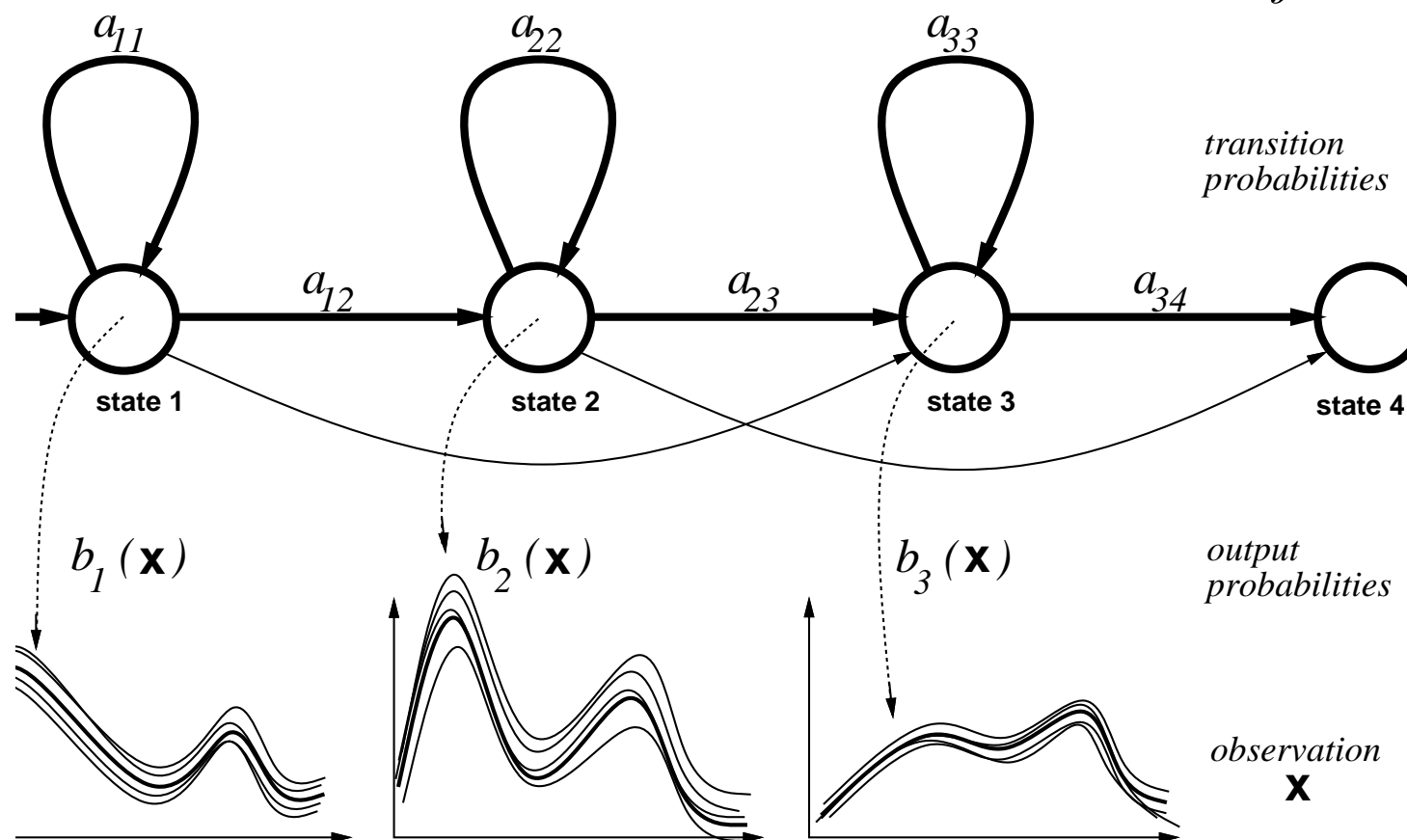


図4. 隠れマルコフモデル(HMM)の例



HMMにおける基本的な計算法

問題 1：確率評価問題

モデル λ に対する観測系列 $Y = (y_1 y_2 \cdots y_T)$ の確率 $P(Y|\lambda)$ をいかに効率よく計算できるか？

但し、状態系列を q_i ($i = 1, 2, \dots, N$)とする。

■ 直接計算

観測系列 Y を出力する可能性のある状態系列 q_i の種類数は、それぞれの時刻 $t = 1, 2, \dots, T$ について N 個の到達可能な状態が存在するので、とりうる状態系列は N^T 個存在する。

計算量： N^T のオーダー！



■ Forward アルゴリズム、backward アルゴリズム

計算量： N^2T のオーダー



Forward アルゴリズム

初期化

フォワード変数(**forward**、前向き) $\alpha(i, t)$ を
モデル λ が与えられたときに、
時刻 t までに部分的な部分観測系列 y_1, y_2, \dots, y_t を出力し、
時刻 t に状態 q_i ($i = 1, 2, \dots, N$) に存在する確率と定義する。

$$\alpha(i, t) = \pi_i \quad (t = 0)$$

逐次計算 (帰納)

$$\alpha(i, t) = \sum_{j=1}^N \alpha(j, t-1) a_{ji} b_{ji}(\mathbf{y}_t) \quad (t = 1, 2, \dots, T)$$

終了

$$P(\mathbf{Y}|\lambda) = \sum_{j=1}^N \alpha(j, T)$$



例題

単語HMMを用いた単語音声認識

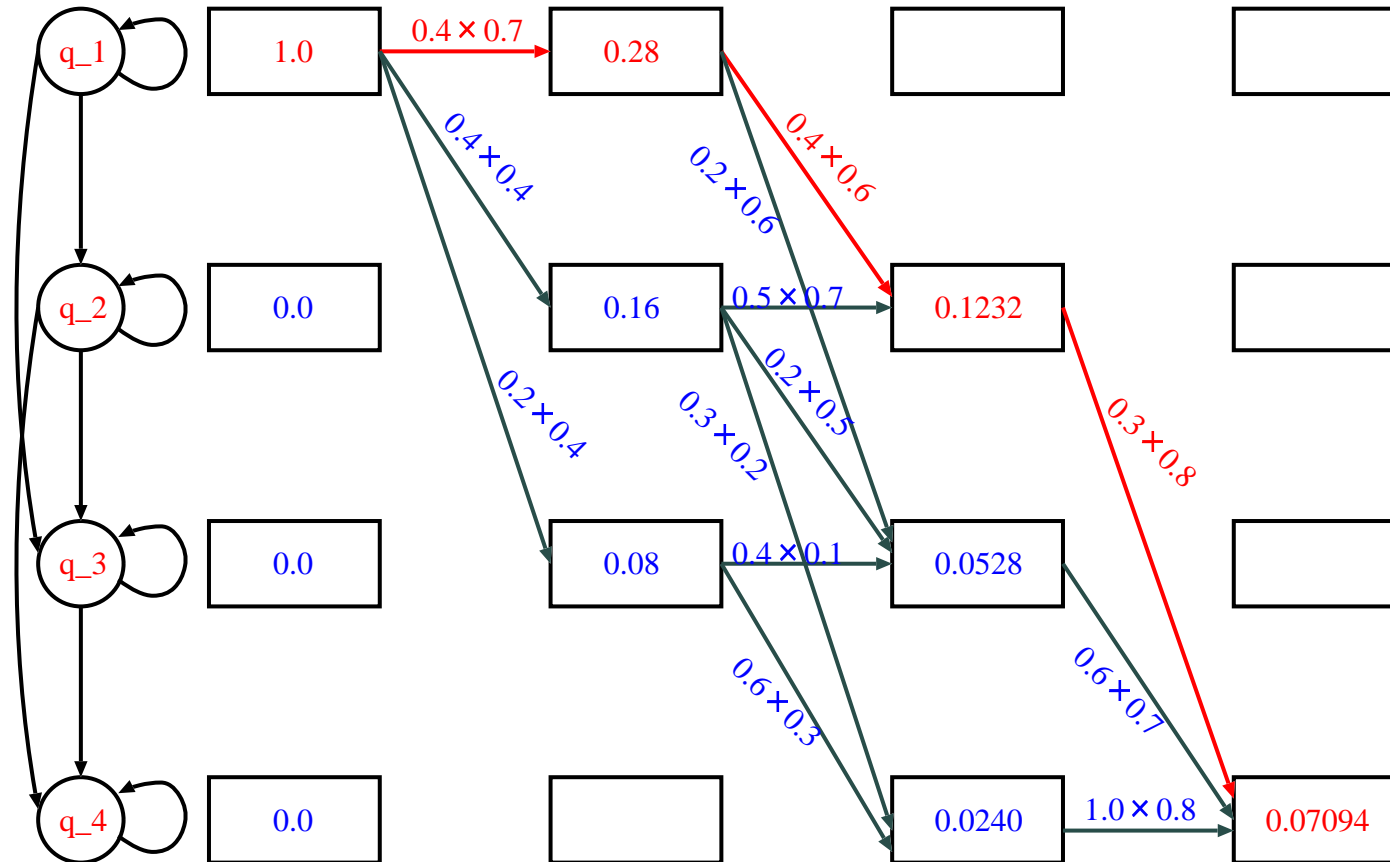


図5. 図3のHMMから系列'aba'が観測される確率を計算する手順

問) この図について、Viterbiパスを求め、Forwardアルゴリズムの結果と一致することを示せ。



HMMの3つの基本問題

モデル $\lambda = (\overset{\text{状態遷移確率行列}}{A}, \overset{\text{観測シンボル確率分布}}{B}, \overset{\text{初期状態分布}}{\pi})$ と観測系列 $Y = (y_1, y_2, \dots, y_T)$ が与えられたとき

問題1 (確率評価)

モデル λ に対する観測系列 Y の確率 $P(Y|\lambda)$ の計算

■ モデル λ が観測系列 Y に対してどの程度適応しているか

⇒ 例：Forward アルゴリズム, backward アルゴリズム

問題2 (最適状態系列)

最適な状態系列 $q = (q_1 q_2 \dots q_T)$ の発見

■ 観測系列 Y がどの状態系列 q から生成されたと考えられるか

⇒ 例：Viterbi アルゴリズム

問題3 (パラメータ推定)

$P(Y|\lambda)$ を最大とするようなモデルパラメータ λ の調整

■ 観測系列 Y を生成するためのパラメータ λ の最適化

⇒ 例：Baum-Welch アルゴリズム



HMMにおける基本的な計算法

問題 1：確率評価問題

モデル λ に対する観測系列 $Y = (y_1 y_2 \cdots y_T)$ の確率 $P(Y|\lambda)$ をいかに効率よく計算できるか？

但し、状態系列を q_i ($i = 1, 2, \dots, N$)とする。

■ 直接計算

観測系列 Y を出力する可能性のある状態系列 q_i の種類数は、それぞれの時刻 $t = 1, 2, \dots, T$ について N 個の到達可能な状態が存在するので、とりうる状態系列は N^T 個存在する。

計算量： N^T のオーダー！



■ Forward アルゴリズム、backward アルゴリズム

計算量： $N^2 T$ のオーダー



Forward アルゴリズム

初期化

フォワード変数 (**forward**、前向き) $\alpha(i, t)$ を
モデル λ が与えられたときに、
時刻 t までに部分的な部分観測系列 y_1, y_2, \dots, y_t を出力し、
時刻 t に状態 q_i ($i = 1, 2, \dots, N$) に存在する確率と定義する。

$$\alpha(i, t) = \pi_i \quad (t = 0)$$

逐次計算 (帰納)

$$\alpha(i, t) = \sum_{j=1}^N \alpha(j, t-1) a_{ji} b_{ji}(\mathbf{y}_t) \quad (t = 1, 2, \dots, T)$$

終了

$$P(\mathbf{Y}|\lambda) = \sum_{j=1}^N \alpha(j, T)$$



認識アルゴリズムの例

単語HMMを用いた単語音声認識

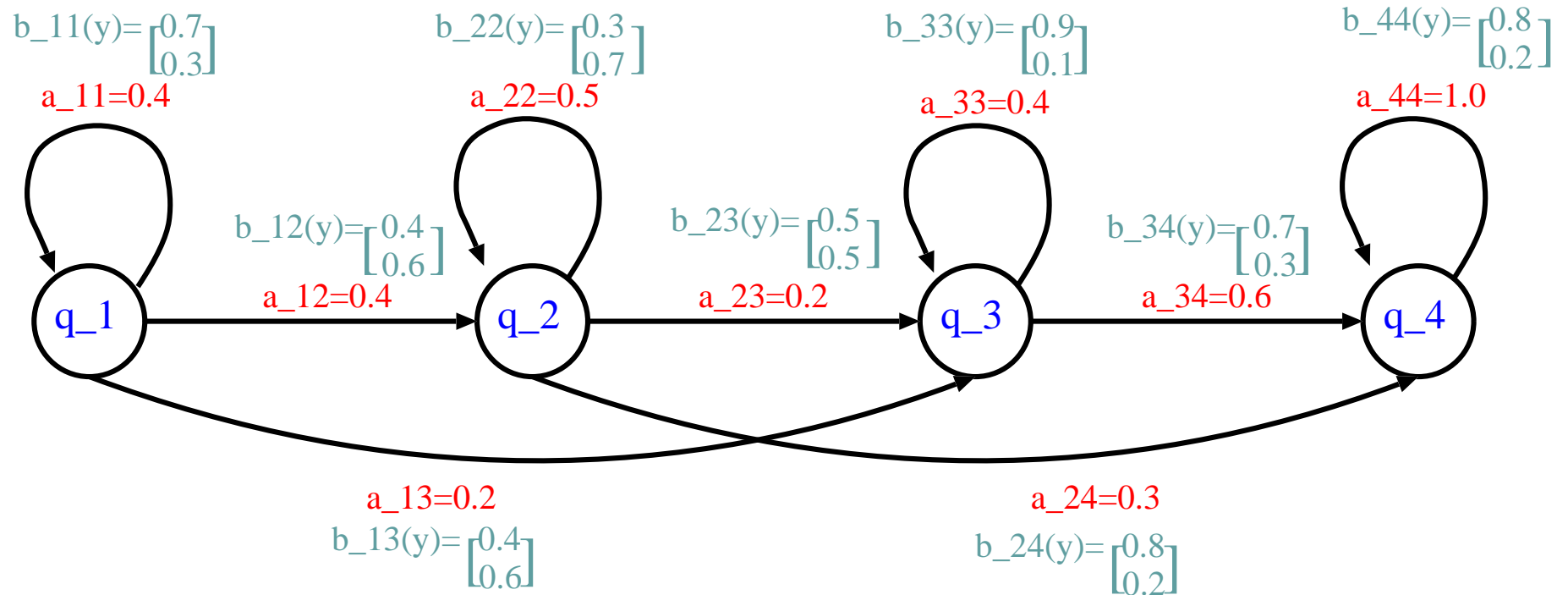


図6. バイキス型HMM

- q_i : 状態
- a_{ij} : 状態遷移行列
- $b_{ij}(Y)$: 出現確率



Forward アルゴリズムの計算手順

単語 HMM を用いた単語音声認識

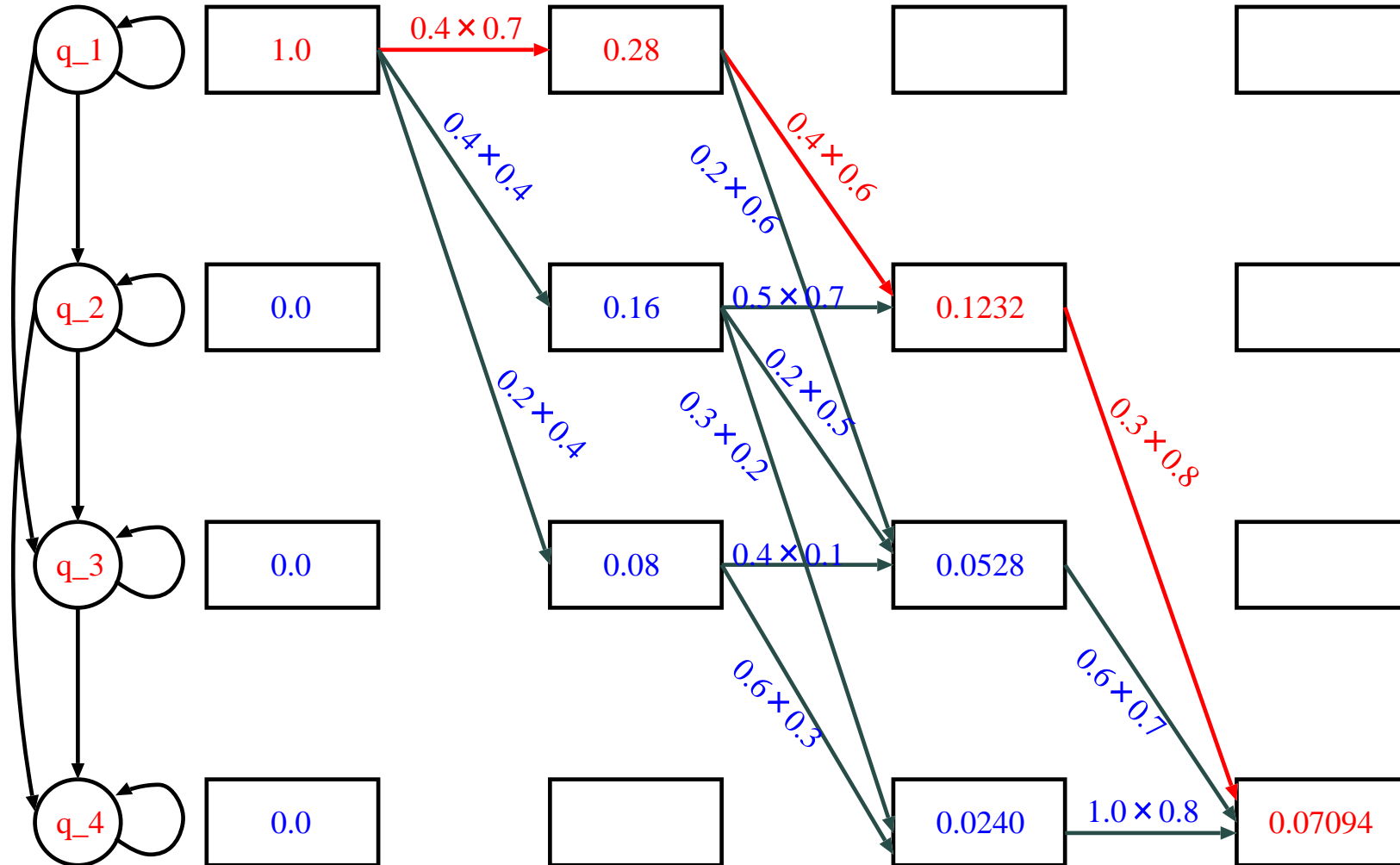


図7. 図3のHMMから系列'aba'が観測される確率を計算する手順



Forward アルゴリズムの計算手順

単語 HMM を用いた単語音声認識

観測系列 'aba' が出力される確率を求める。

この場合の状態遷移系列は、

時間を横に状態を縦に並べた図 4 の平面に於いて、

右上隅から右下隅に至る経路に対応し、次の 7 通りである。

以下の表では直接計算しているが、

Forward アルゴリズムで計算が省略できる。

| $q_1 \rightarrow q_u \rightarrow q_v \rightarrow q_4$ | $P_k = a_{1u} \times b_u(\mathbf{Y}_1) \times a_{uv} \times b_v(\mathbf{Y}_2) \times a_{v4} \times b_4(\mathbf{Y}_3)$ |
|---|---|
| $q_1 \rightarrow q_1 \rightarrow q_2 \rightarrow q_4$ | $P_1 = 0.4 \times 0.7 \times 0.4 \times 0.6 \times 0.3 \times 0.8 = 0.016128$ |
| $q_1 \rightarrow q_1 \rightarrow q_3 \rightarrow q_4$ | $P_2 = 0.4 \times 0.7 \times 0.2 \times 0.6 \times 0.6 \times 0.7 = 0.014112$ |
| $q_1 \rightarrow q_2 \rightarrow q_2 \rightarrow q_4$ | $P_3 = 0.4 \times 0.4 \times 0.5 \times 0.7 \times 0.3 \times 0.8 = 0.013440$ |
| $q_1 \rightarrow q_2 \rightarrow q_3 \rightarrow q_4$ | $P_4 = 0.4 \times 0.4 \times 0.2 \times 0.5 \times 0.6 \times 0.7 = 0.006720$ |
| $q_1 \rightarrow q_2 \rightarrow q_4 \rightarrow q_4$ | $P_5 = 0.4 \times 0.4 \times 0.3 \times 0.2 \times 1.0 \times 0.8 = 0.007680$ |
| $q_1 \rightarrow q_3 \rightarrow q_3 \rightarrow q_4$ | $P_6 = 0.2 \times 0.4 \times 0.4 \times 0.1 \times 0.6 \times 0.7 = 0.001344$ |
| $q_1 \rightarrow q_3 \rightarrow q_4 \rightarrow q_4$ | $P_7 = 0.2 \times 0.4 \times 0.6 \times 0.3 \times 1.0 \times 0.8 = 0.011520$ |

それぞれの確率は上に示す通りで、

$$P(aba|\lambda) = P_1 + P_2 + P_3 + P_4 + P_5 + P_6 + P_7 = 0.070944$$



単語HMMを用いた単語音声認識

認識アルゴリズム

各HMMモデルごとに、観測系列 $Y = (y_1 y_2 \cdots y_T)$ が生起する確率（尤度） $P(Y|\lambda)$ を求め、最大確率（最大尤度）を与えるモデルを選んで、認識結果とする。

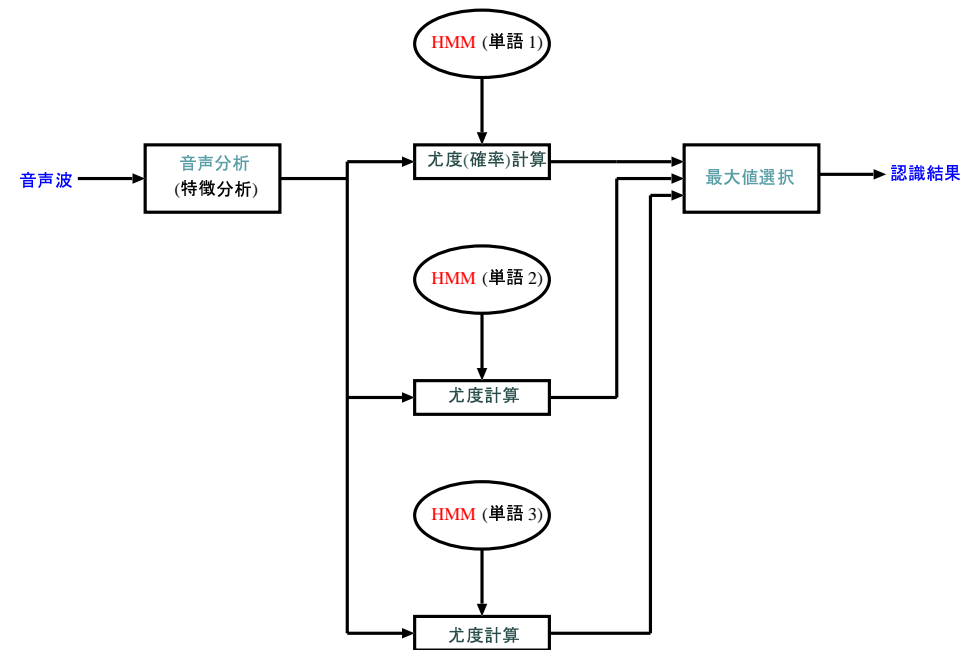


図8. 単語HMMを用いた単語音声認識の方法