

6. 音声の認識：基本的な音響モデル

6.1 音響モデルの単位

6.2 隠れマルコフモデルとは

6.3 隠れマルコフモデルの確率計算

6.4 状態系列の推定

6.5 パラメータの学習

6.1 音響モデルの単位

- 音響モデル p (特徴ベクトル系列|単語列) の条件部
 - 単語列 (=文)
 - 可能な文が列挙できる小さなタスクでは有効
 - 数万語の語彙が必要なディクテーションでは、可能な文の数は膨大になるので、実質的にモデル化は不可能
 - 単語
 - 新しい単語がタスクに追加されるごとに、モデルを作成しなければならない
 - 音素
 - 単語辞書を音素系列で記述することで、大語彙に対応可能
 - 前後の音素情報を組み込んだトライフォンが主流

6.2 隠れマルコフモデルとは

- 音響モデルのためのオートマトンの構造
 - 各状態で確率的にベクトルを出力し、確率的に状態遷移を行うムーア型オートマトン
 - 状態遷移が一方向に限定されているleft-to-right型の構造をもつマルコフモデルと解釈できる

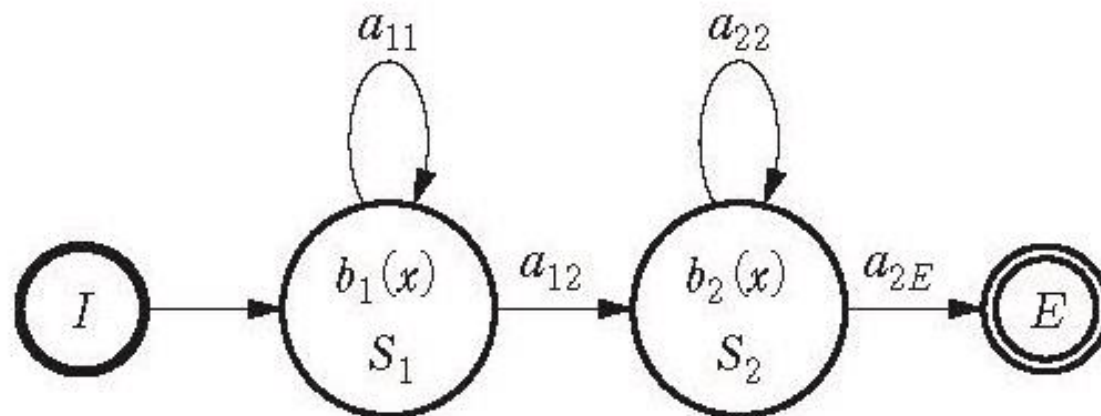


図 6.4

音響モデルのためのオートマトンの構造

6.2 隠れマルコフモデルとは

- 「隠れ」マルコフモデル
 - どの状態からどのベクトルが出力されたかという情報が隠れている

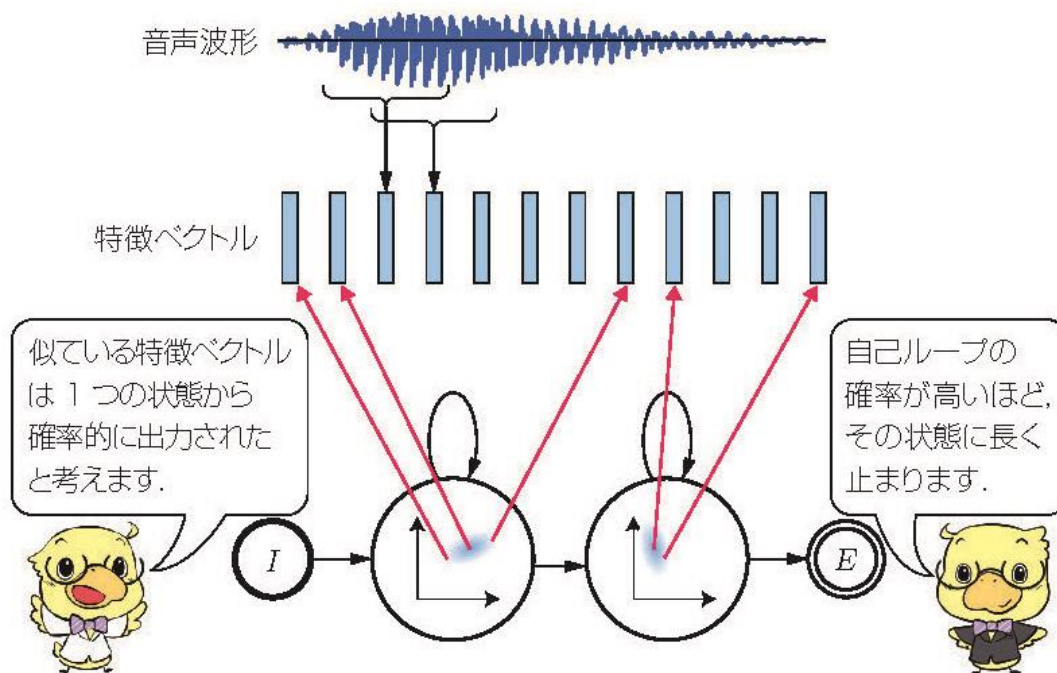


図 6.5 隠れマルコフモデルからの特徴ベクトルの出力

6.3 隠れマルコフモデルの確率計算

- 系列の出力確率

- 例) 図6.4において $\mathbf{x} = x_1 x_2 x_3 x_4 x_5$,
状態遷移 $S_1 S_1 S_1 S_2 S_2 E$ の場合

$$P(\mathbf{x}) = b_1(x_1)a_{11}b_1(x_2)a_{11}b_1(x_3)a_{12}b_2(x_4)a_{22}b_2(x_5)a_{2E}$$

- すべての可能な状態遷移について求め、和を計算
- トレリス計算による効率化 (前向きアルゴリズム)
 - 時刻 t 、状態 i における前向き確率 $\alpha_t(i)$ を、入力の時間単位で順次計算 (m はHMMの状態数)

$$\alpha_t(i) = \left\{ \sum_{j=1}^m \alpha_{t-1}(j) a_{ji} \right\} b_i(x_t)$$

6.3 隠れマルコフモデルの確率計算

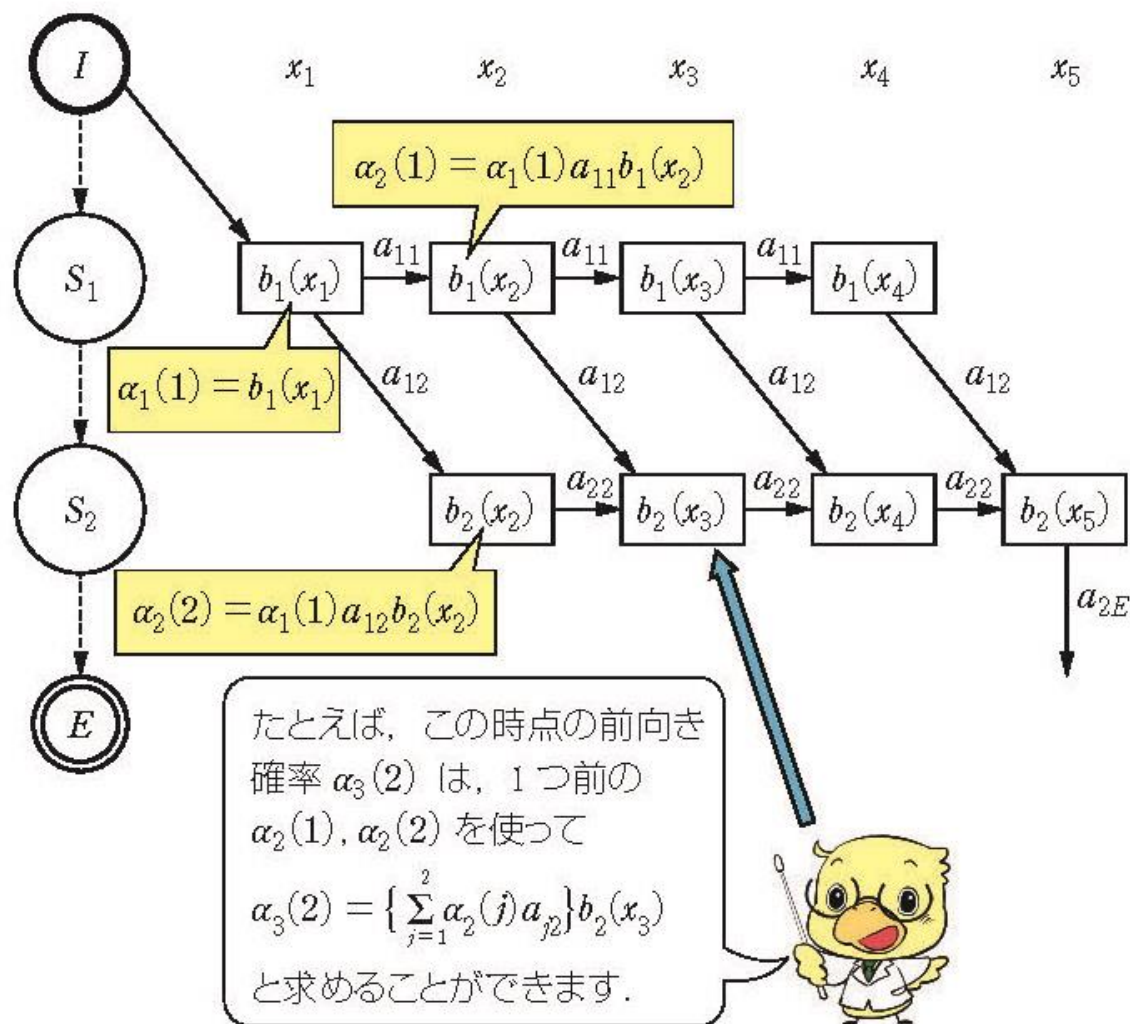


図 6.6 前向きアルゴリズムを用いた確率計算

6.4 状態系列の推定

- ビタビアルゴリズム
 - 前向きアルゴリズムにおける系列の和の計算を、最大値演算に置き換える
 - 最大値を与えた経路を保存しておき、最終状態から逆にたどることで、最も確率の高い経路が得られる

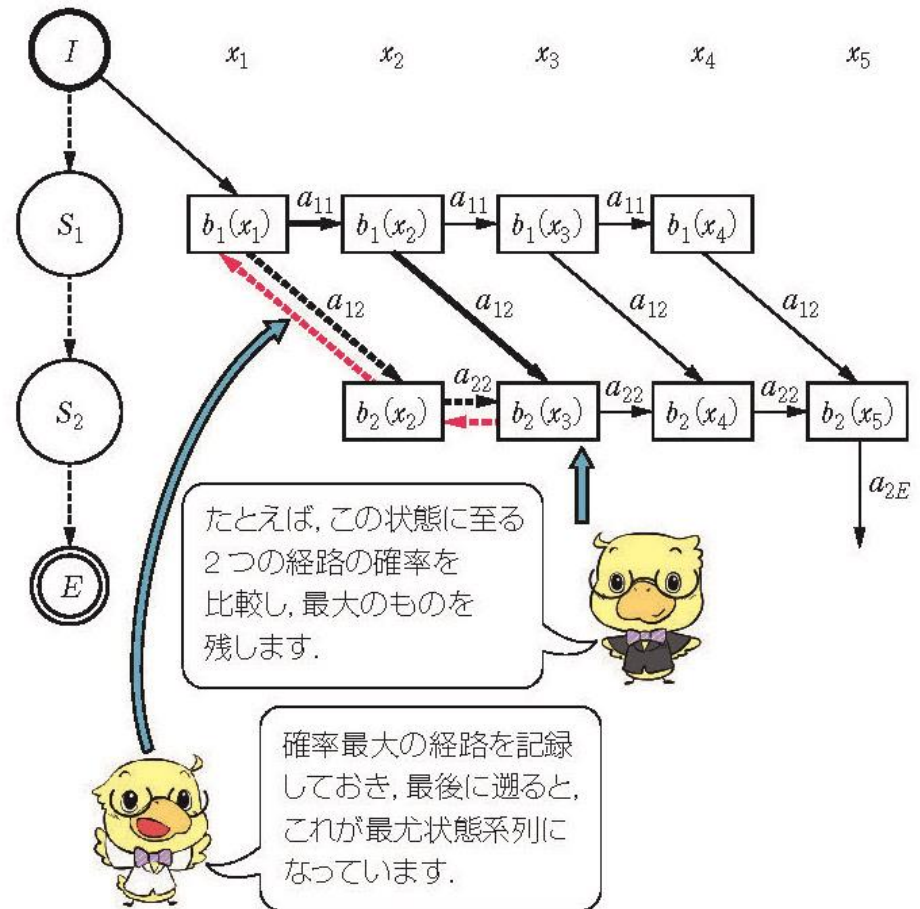


図 6.7 ビタビアルゴリズム

6.5 パラメータの学習

- 状態遷移系列が既知の場合のパラメータ推定
 - 単純な最尤推定

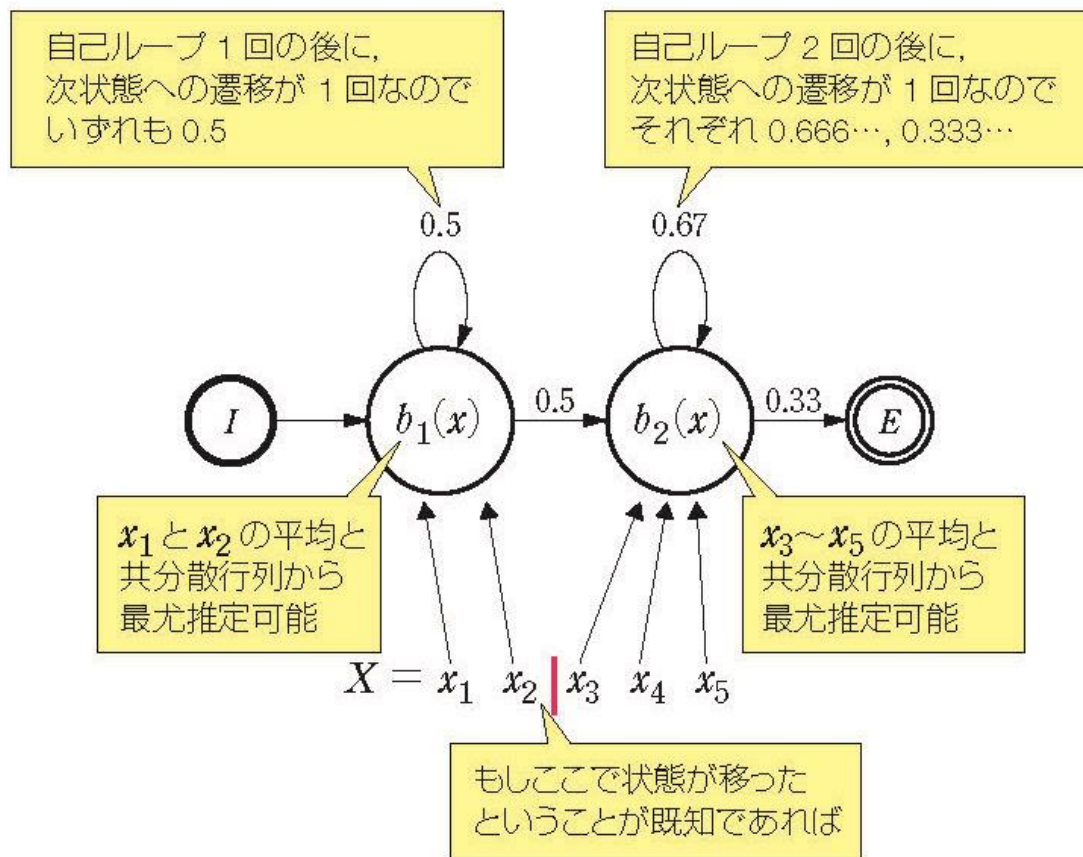


図 6.9 状態遷移系列が既知の場合のパラメータ推定

6.5 パラメータの学習

- 状態遷移系列の確率が既知の場合のパラメータ推定
 - それぞれの最尤推定結果を重み付きでたし合わせる

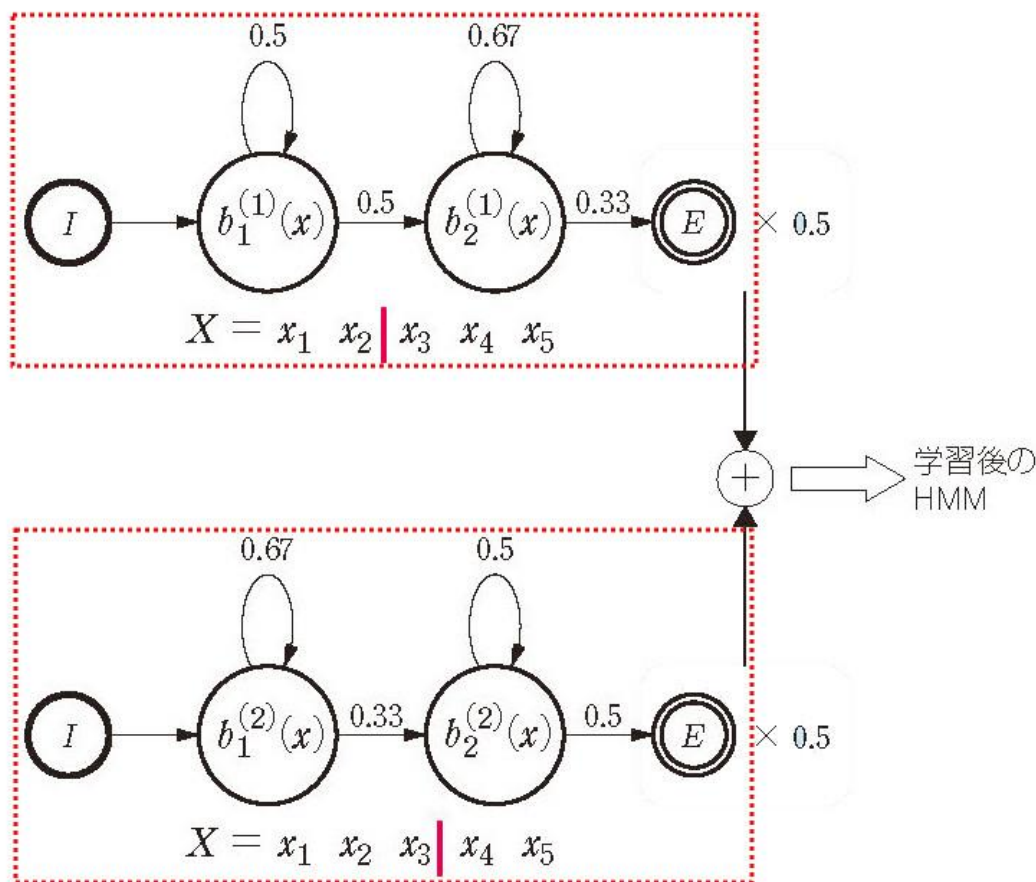
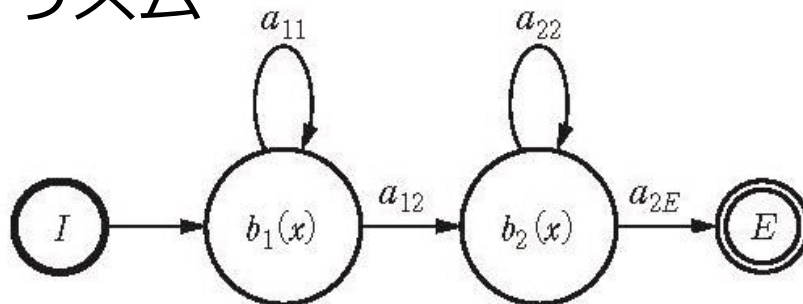


図 6.10

状態遷移系列の確率が既知の場合のパラメータ推定

6.5 パラメータの学習

- 状態遷移系列が未知の場合のパラメータ推定
 - EMアルゴリズム



この手順を、
Mステップの
パラメータが
安定するまで
繰り返します。



Mステップ：Eステップで得られた確率をもとに
図 6.10 の方法で HMM を学習

$$P(x_1 \mid x_2 \ x_3 \ x_4 \ x_5)$$

$$P(x_1 \ x_2 \mid x_3 \ x_4 \ x_5)$$

$$P(x_1 \ x_2 \ x_3 \mid x_4 \ x_5)$$

$$P(x_1 \ x_2 \ x_3 \ x_4 \mid x_5)$$

Eステップ：Mステップで得られた HMM で、
それぞれの状態遷移系列の確率を計算

初期パラメータが
多少でたらめでも、
学習データの値の
偏りによって、
HMMのパラメー
タが徐々にデー
タの統計的性質を反
映させてゆくプロ
セスとみなせる。

図 6.11 EM アルゴリズム