人工知能特論II　第7回

二宮　崇

今日の講義の予定

- 系列ラベリングのためのHMM
  - 解析
    - ビタビアルゴリズム
  - 学習
    - 最尤推定

- 教科書
  - 北研二(著) 辻井潤一(編) 言語と計算4 確率的言語モデル 東大出版会
  - C. D. Manning & Hinrich Schütze “FOUNDATIONS OF STATISTICAL NATURAL LANGUAGE PROCESSING” MIT Press, 1999
  - Christopher M. Bishop “PATTERN RECOGNITION AND MACHINE LEARNING” Springer, 2006

系列ラベリングのためのHMM

HMM FOR SEQUENTIAL LABELING

品詞解析

- 品詞タガー

"I have a pen."

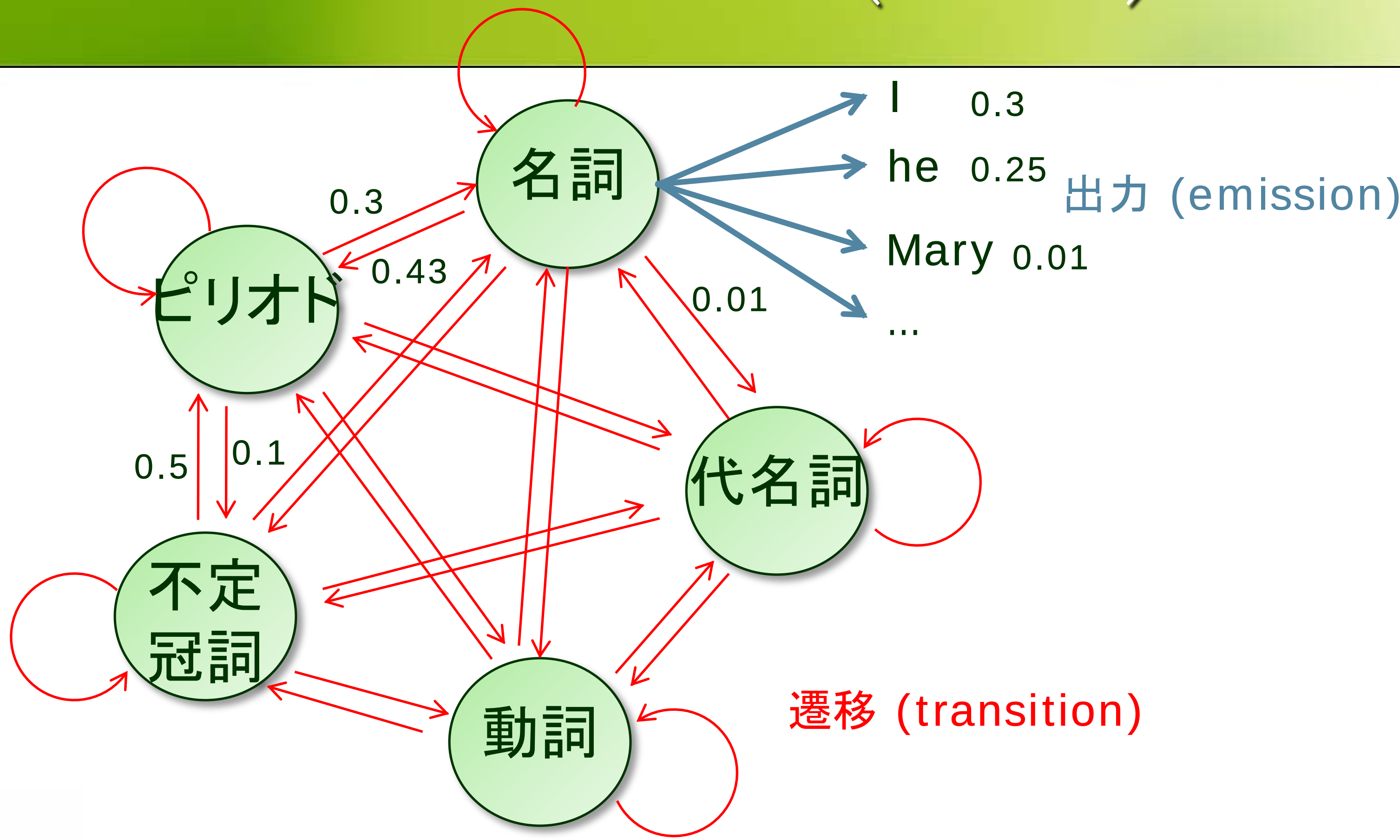
↓ トーカナイザー

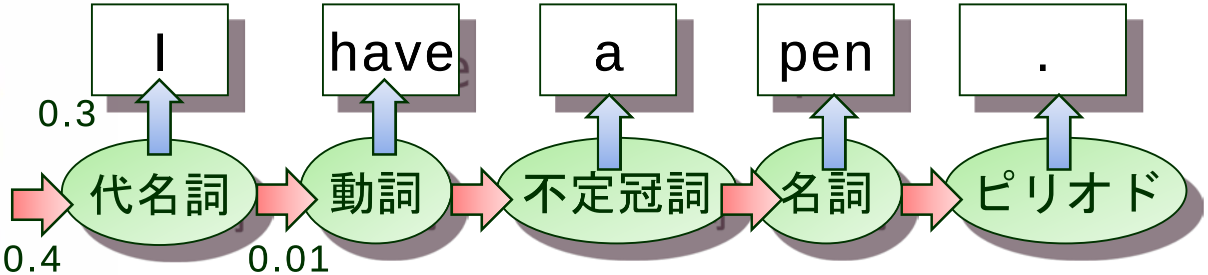
I	have	a	pen	.

↓ POSタガー

I	have	a	pen	.
代名詞	動詞	不定冠詞	名詞	ピリオド

隠れマルコフモデル
Hidden Markov Model (HMM)





隠れマルコフモデル Hidden Markov Model (HMM)

- Q: 状態の有限集合
- Σ: 出力記号の有限集合
- π_q: 文頭が状態qになる確率
  - $\Sigma_{r \in Q}\ \pi_r = 1$
- $a_{q,r}$: 状態qから状態rへの遷移確率
  - $\Sigma_{r \in Q}\ a_{q,r} = 1$
- $b_{q,o}$: 状態qにおける記号oの出力確率
  - $\Sigma_{o \in \Sigma}\ b_{q,o} = 1$

状態記号列の確率と生成確率

- 状態と記号の列が与えられた時

$$\text{状態記号列}: q_1 o_1 q_2 o_2 \cdots q_T o_T$$

$$\begin{aligned} p(q_1 o_1 q_2 o_2 \cdots q_T o_T) &= \pi_{q_1} b_{q_1, o_1} a_{q_1, q_2} b_{q_2, o_2} \cdots a_{q_{T-1}, q_T} b_{q_T, o_T} \\ &= \pi_{q_1} a_{q_1, q_2} \cdots a_{q_{T-1}, q_T} b_{q_1, o_1} b_{q_2, o_2} \cdots b_{q_T, o_T} \\ &= \pi_{q_1} \prod_{t=2}^{T} a_{q_{t-1}, q_t} \prod_{t=1}^{T} b_{q_t, o_t} \end{aligned}$$

- 記号列のみが与えられた時

$$\text{記号列}: o_1 o_2 \cdots o_T$$

$$p(o_1 o_2 \cdots o_T) = \sum_{q_1 \in Q, q_2 \in Q, \cdots q_T \in Q} p(q_1 o_1 q_2 o_2 \cdots q_T o_T)$$ (生成確率)

推論(INFERENCE)
解析(ANALYSIS)
タグ付け(TAGGING)
復号化(DECODING)

解析

- 解析 (入力: $o_1 o_2 \ldots o_T$)

$$\tilde{q}_1 \tilde{q}_2 \cdots \tilde{q}_T = \mathop{\arg\max}_{q_1 \in Q, q_2 \in Q, \cdots q_T \in Q} p(q_1 o_1 q_2 o_2 \cdots q_T o_T)$$

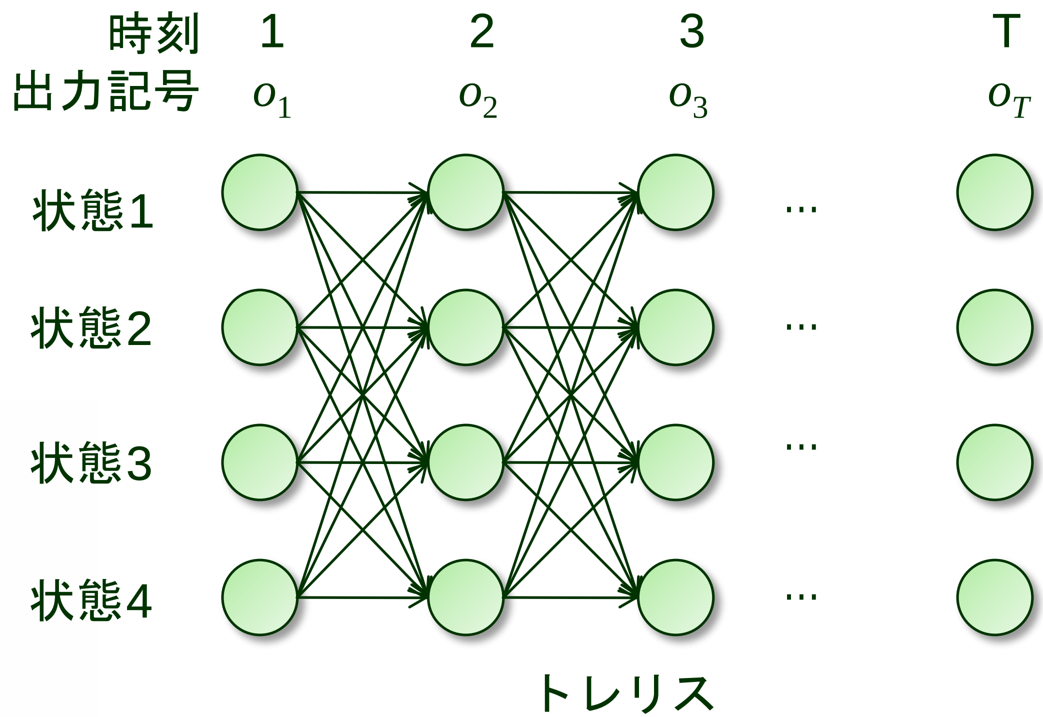
しかし、計算量はO($|Q|^T$)!

効率的な品詞解析: ビタビアルゴリズム

- 動的計画法

$\delta(t, q)$: 時刻t(o_tが出力される時)に状態qとなる状態列の中で最大確率となる列の確率

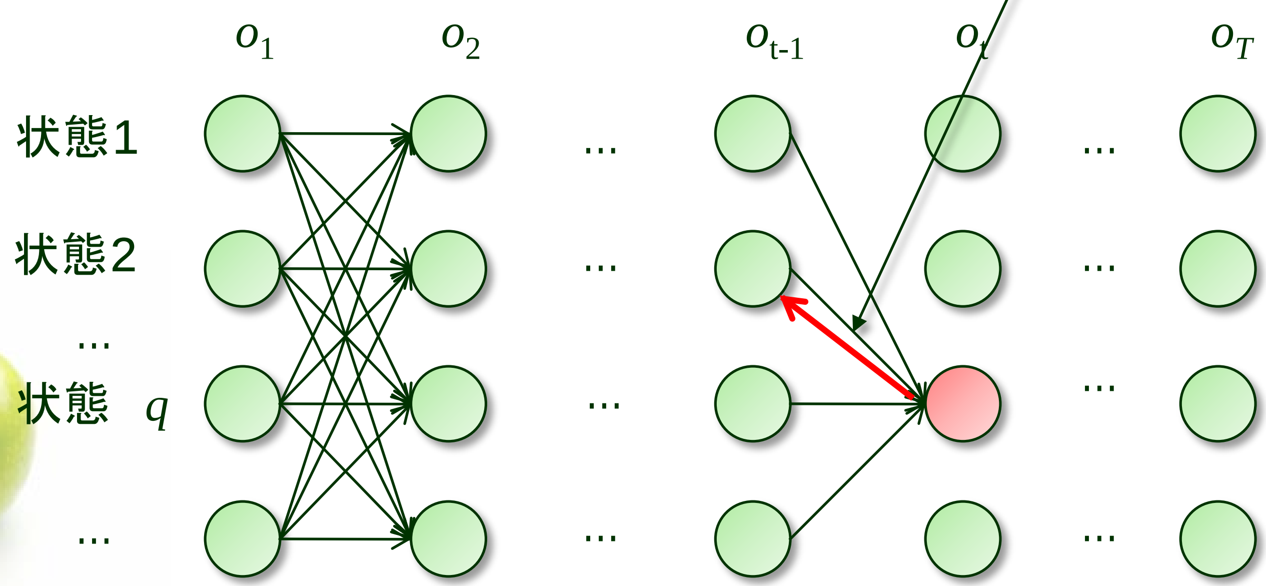
$max_{q \in Q}\ \delta(T, q)$を求めれば良い



効率的な品詞解析: ビタビアルゴリズム

$$
\begin{aligned}
\delta(t,q) &= \max_{q_1\in Q,\cdots,q_{t-1}\in Q} p(q_1 o_1 \cdots q_{t-1} o_{t-1}) a_{q_{t-1},q} b_{q,o_t} \\
&= \max_{q_{t-1}\in Q} \left\{ \max_{q_1\in Q,\cdots,q_{t-2}\in Q} \left\{ p(q_1 o_1 \cdots q_{t-2} o_{t-2}) a_{q_{t-2},q_{t-1}} b_{q_{t-1},o_{t-1}} \right\} a_{q_{t-1},q} b_{q,o_t} \right\} \\
&= \max_{q_{t-1}\in Q} \delta(t-1, q_{t-1}) a_{q_{t-1},q} b_{q,o_t}
\end{aligned}
$$

最大確率で遷移したパス
をバックポインタで保存

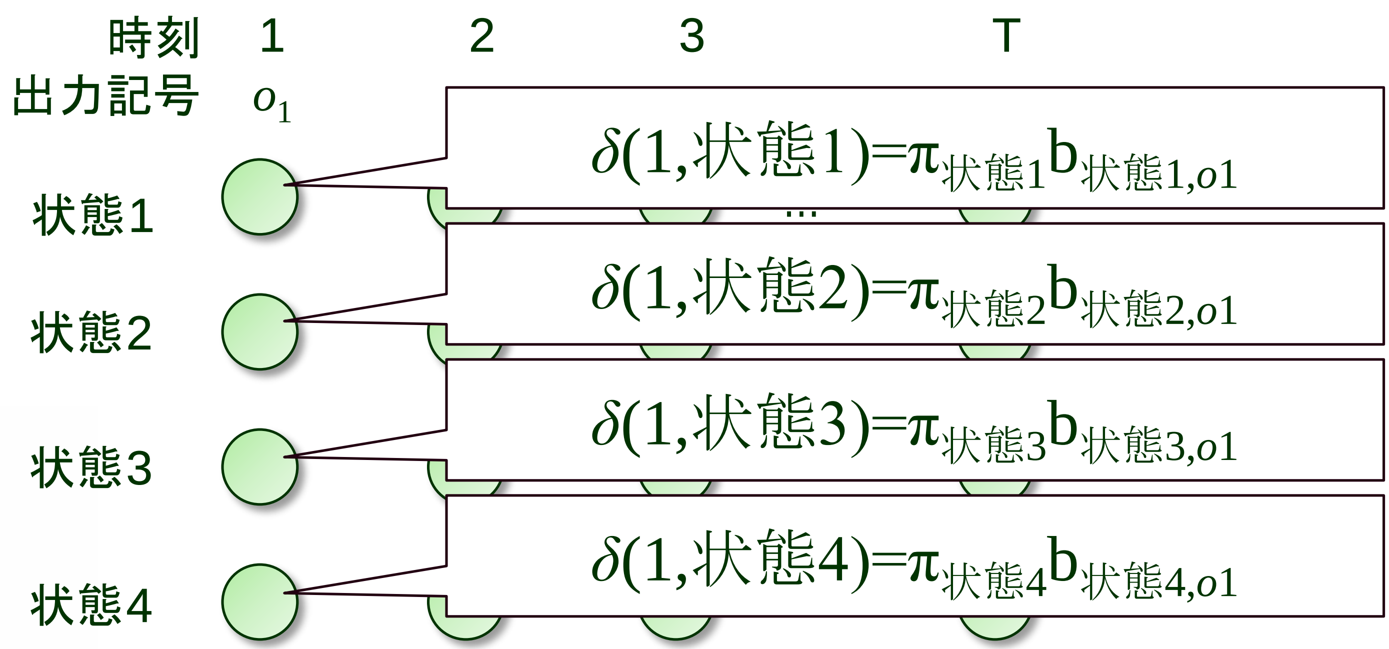


効率的な品詞解析: ビタビアルゴリズム

- 動的計画法

$\delta(t, q)$: 時刻t(o_tが出力される時)に状態qとなる状態列の中で最大確率となる列の確率

左から右に向かって
全ての$\delta(t,q)$を計算



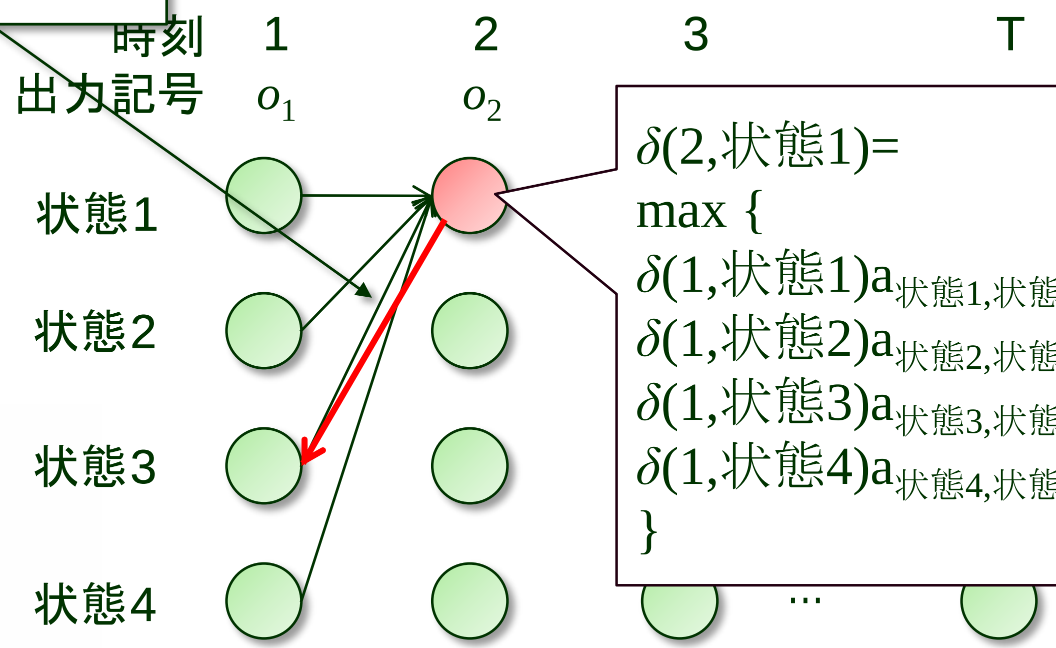
効率的な品詞解析: ビタビアルゴリズム

- 動的計画法

$\delta(t, q)$: 時刻t(o_tが出力される時)に状態qとなる状態列の中で最大確率となる列の確率

最大確率で遷移したパスをバックポインタで保存

左から右に向かって全ての$\delta(t,q)$を計算



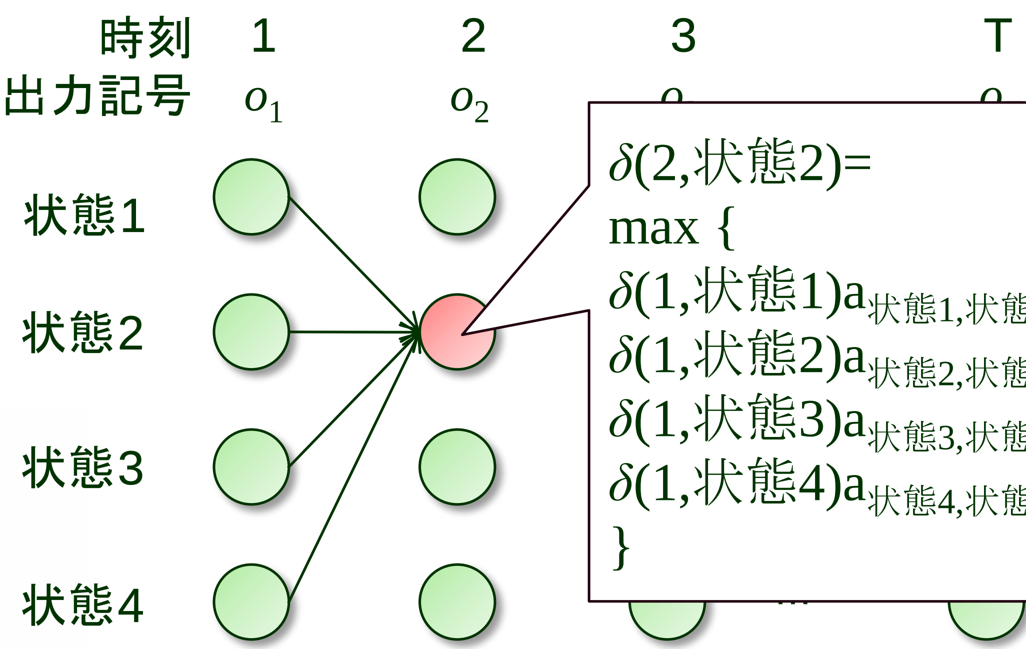
δ(2,状態1)=
max {
δ(1,状態1)$a_{状態1,状態1}b_{状態1,o2}$,
δ(1,状態2)$a_{状態2,状態1}b_{状態1,o2}$,
δ(1,状態3)$a_{状態3,状態1}b_{状態1,o2}$,
δ(1,状態4)$a_{状態4,状態1}b_{状態1,o2}$
}

効率的な品詞解析: ビタビアルゴリズム

- 動的計画法

$\delta(t, q)$: 時刻t(o_tが出力される時)に状態qとなる状態列の中で最大確率となる列の確率

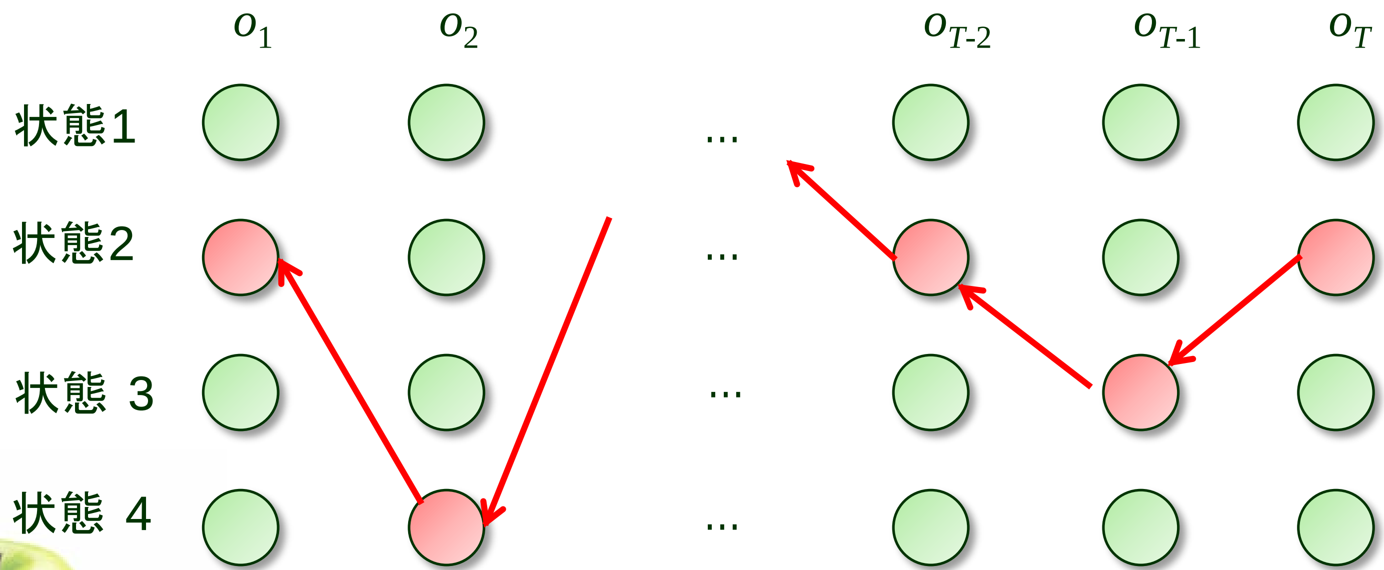
左から右に向かって
全ての$\delta(t,q)$を計算



$$
\begin{aligned}
&\delta(2,\text{状態}2)= \\
&\max\{ \\
&\delta(1,\text{状態}1)a_{\text{状態}1,\text{状態}2}b_{\text{状態}2,o2}, \\
&\delta(1,\text{状態}2)a_{\text{状態}2,\text{状態}2}b_{\text{状態}2,o2}, \\
&\delta(1,\text{状態}3)a_{\text{状態}3,\text{状態}2}b_{\text{状態}2,o2}, \\
&\delta(1,\text{状態}4)a_{\text{状態}4,\text{状態}2}b_{\text{状態}2,o2} \\
&\}
\end{aligned}
$$

効率的な品詞解析: ビタビアルゴリズム

- 最後にバックポインタを辿ることで最大確率となる状態列が得られる



効率的な品詞解析: ビタビアルゴリズム

$\delta[1,q] := \pi[q]b[q, o_1]$ (for all q)

for t =2 to T

  for $q \in Q$

    $\delta[t, q] := \max_{q' \in Q} \{\delta[t-1, q']a[q',q]b[q, o_t]\}$

    $bp[t,q] := \mathrm{argmax}_{q' \in Q} \{\delta[t-1, q']a[q',q]b[q, o_t]\}$

学習 (パラメータ推定): 最尤推定

パラメータ推定: 最尤推定

- 最尤推定
  - 文の集合を観測したとき、その文の集合がそこに出現したのは、その文の集合が最も確率が高かったから、と考えるやり方
  - コインの例：表がでる確率θが未知のコインがある。100回投げたところ、62回表がでた。すると、その確率は$\theta^{62}(1-\theta)^{38}$となる。この確率はθ=0.62で最大となるので、θは0.62であったのだろう、と考えるのが最尤推定の考え方である。

最尤推定

- 最尤推定
  - 観測値 $x_1,\ldots,x_n$ が与えられた時、それぞれが独立に出現したと考えると、その確率はパラメータθの関数になる

$$p(x_1,\ldots,x_n)=\prod_{i=1}^{n}p(x_i;\theta)=l(\theta)$$

  - この$l(\theta)$を尤度(likelihood)もしくは尤度関数(likelihood function)と呼ぶ
  - 尤度関数を最大化するθを求める

$$\tilde{\theta}=\arg\max_{\theta} l(\theta)$$

最尤推定

- 最大を求めるために尤度関数の極値を求める

$$\frac{\partial}{\partial \theta} l(\theta) = 0$$

- コインの例を解析的に解いてみよう
  - $l(\theta) = \theta^{62}(1-\theta)^{38}$

最尤推定

- 対数尤度(log likelihood)を使うと計算が楽になる

$$\tilde{\theta} = \arg\max_{\theta} l(\theta) = \arg\max_{\theta} \log l(\theta)$$

- コインの例で解いてみよう
  - $\log l(\theta) = \log(\theta^{62}(1-\theta)^{38})$

最尤推定

- 正規分布の最尤推定
  - 正規分布N(μ, σ^2)から抽出された標本を$x_1,\ldots,x_n$とする
  - 尤度

$$l(\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right]$$

  - 対数尤度

$$\begin{aligned}\log l(\mu,\sigma^2) &= \sum_{i=1}^{n}\log\frac{1}{\sqrt{2\pi}\sigma}+\sum_{i=1}^{n}\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right] \\ &= -n\log(\sqrt{2\pi}\sigma)-\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}\end{aligned}$$

ラグランジュの未定乗数法

- ラグランジュの未定乗数法

$$\arg\max_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) \text{ただし} g_1(\boldsymbol{\theta}) = 0, \ldots, g_m(\boldsymbol{\theta}) = 0$$

$$\Rightarrow$$

$$L(\boldsymbol{\theta}) = f(\boldsymbol{\theta}) - \lambda_1 g_1(\boldsymbol{\theta}) - \ldots - \lambda_m g_m(\boldsymbol{\theta})$$

$$\frac{\partial L}{\partial \theta_1} = 0, \frac{\partial L}{\partial \theta_2} = 0, \ldots, \frac{\partial L}{\partial \theta_n} = 0$$

- $L(\boldsymbol{\theta})$ はラグランジュ関数と呼ばれる

まとめ

- HMM
  - 解析
    - ビタビアルゴリズム
  - 学習
    - 最尤推定
- 資料

  http://aiweb.cs.ehime-u.ac.jp/~ninomiya/ai2/