



人工知能特論II 第10回

二宮 崇

今日の講義の予定

- HMMの前向き後向きアルゴリズムによるパラメータ学習
- 生成モデルと識別モデル
- 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
 - Christopher M. Bishop “PATTERN RECOGNITION AND MACHINE LEARNING” Springer, 2006



HMMの教師無し学習

Unsupervised Learning of HMMs

- パラメータ推定

- 訓練データ (入力)

訓練データ: $x_1 x_2 \cdots x_n$

x_i : 文を表す記号列(単語列)。 $x_i = o_{i1} o_{i2} o_{i3} \cdots o_{iT_i}$ とする。

T_i : x_i の記号列長

- パラメータ (出力)

$$\begin{aligned} \pi, a, b &= \arg \max_{\pi, a, b} \prod_{i=1}^n p(x_i) \\ &= \arg \max_{\pi, a, b} \prod_{i=1}^n p(o_{i1} o_{i2} \cdots o_{iT_i}) \\ &= \arg \max_{\pi, a, b} \prod_{i=1}^n \sum_{q_1 \in Q, q_2 \in Q, \dots, q_{T_i} \in Q} p(q_1 o_{i1} q_2 o_{i2} \cdots q_{T_i} o_{iT_i}) \end{aligned}$$



HMMのEMアルゴリズム

- パラメータ更新式
 - π, a, b を π', a', b' に更新

$$\pi'_q = \frac{\sum_{i=1}^n \sum_{q_1 \in Q, \dots, q_{T_i} \in Q} p(q_1 \cdots q_{T_i} | o_1 \cdots o_{T_i}; \theta) [q_1 = q]}{n}$$

$$= \frac{\sum_{i=1}^n \sum_{q_2 \in Q, \dots, q_{T_i} \in Q} p(q q_2 \cdots q_{T_i} | o_1 \cdots o_{T_i}; \theta)}{n}$$

$$= \frac{\text{先頭が } q \text{ となる回数の期待値}}{n}$$

HMMのEMアルゴリズム

- パラメータ更新式

- π, a, b を π', a', b' に更新

$$a'_{q,r} = \frac{\sum_{i=1}^n \sum_{q_1 \in Q, \dots, q_{T_i} \in Q} p(q_1 \dots q_{T_i} | o_1 \dots o_{T_i}; \theta) \left(\sum_{t=2}^{T_i} [q = q_{t-1}][r = q_t] \right)}{\sum_{r' \in Q} \sum_{i=1}^n \sum_{q_1 \in Q, \dots, q_{T_i} \in Q} p(q_1 \dots q_{T_i} | o_1 \dots o_{T_i}; \theta) \left(\sum_{t=2}^{T_i} [q = q_{t-1}][r' = q_t] \right)}$$

$$= \frac{\sum_{i=1}^n \sum_{q_1 \in Q, \dots, q_{T_i} \in Q} p(q_1 \dots q_{T_i} | o_1 \dots o_{T_i}; \theta) C(q, r; q_1 \dots q_{T_i})}{\sum_{r' \in Q} \sum_{i=1}^n \sum_{q_1 \in Q, \dots, q_{T_i} \in Q} p(q_1 \dots q_{T_i} | o_1 \dots o_{T_i}; \theta) C(q, r'; q_1 \dots q_{T_i})}$$

$$= \frac{\text{状態}q\text{から状態}r\text{へ遷移する回数の期待値}}{\text{状態}q\text{から遷移する回数の期待値}}$$

$C(q, r; q_1 \dots q_T)$ $q_1 \dots q_T$ の中で q から r に遷移する回数

HMMのEMアルゴリズム

- パラメータ更新式

- π, a, b を π', a', b' に更新

$$b'_{q,o} = \frac{\sum_{i=1}^n \sum_{q_1 \in Q, \dots, q_{T_i} \in Q} p(q_1 \dots q_{T_i} | o_1 \dots o_{T_i}; \theta) \left(\sum_{t=1}^{T_i} [q = q_t][o = o_t] \right)}{\sum_{o' \in \Sigma} \sum_{i=1}^n \sum_{q_1 \in Q, \dots, q_{T_i} \in Q} p(q_1 \dots q_{T_i} | o_1 \dots o_{T_i}; \theta) \left(\sum_{t=1}^{T_i} [q = q_t][o' = o_t] \right)}$$

$$= \frac{\sum_{i=1}^n \sum_{q_1 \in Q, \dots, q_{T_i} \in Q} p(q_1 \dots q_{T_i} | o_1 \dots o_{T_i}; \theta) C(q, o; q_1 \dots q_{T_i}, o_1 \dots o_{T_i})}{\sum_{i=1}^n \sum_{q_1 \in Q, \dots, q_{T_i} \in Q} p(q_1 \dots q_{T_i} | o_1 \dots o_{T_i}; \theta) C(q; q_1 \dots q_{T_i})}$$

$$= \frac{\text{状態 } q \text{ に滞在し記号 } o \text{ を出力する回数の期待値}}{\text{状態 } q \text{ に滞在する回数の期待値}}$$

$C(q; q_1 \dots q_T)$ $q_1 \dots q_T$ の中で q が出現した回数

$C(q, o; q_1 \dots q_T, o_1 \dots o_T)$ $q_1 \dots q_T o_1 \dots o_T$ の中で q から o を出力する回数

HMMのためのEMアルゴリズム の問題点

- 期待値を計算するために全ての可能な隠れ状態を列挙すると文長に対し、指数爆発的計算量が必要

解決策：前向き後向きアルゴリズム。隠れ状態を列挙することなく、この期待値を計算する動的計画法。



前向きアルゴリズムと後ろ向きアルゴリズム



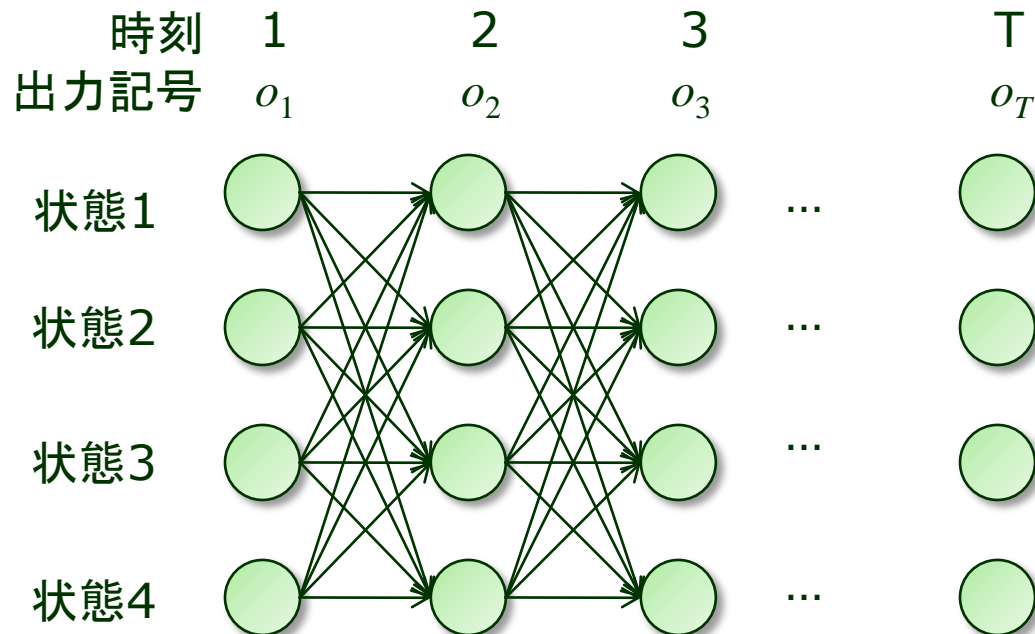
前向きアルゴリズム (1/3)

動的計画法

$\alpha(t, q)$: $o_1 \dots o_t$ を出力し、 o_t が出力される時(時刻 t) に状態 q である確率

$$\alpha(t, q) = \sum_{q_1 \in Q, \dots, q_{t-1} \in Q} \pi_{q_1} b_{q_1, o_1} \left(\prod_{u=2}^{t-1} a_{q_{u-1}, q_u} b_{q_u, o_u} \right) a_{q_{t-1}, q} b_{q, o_t}$$

全ての t, q に対し、 $\alpha(t, q)$ を求めれば良い



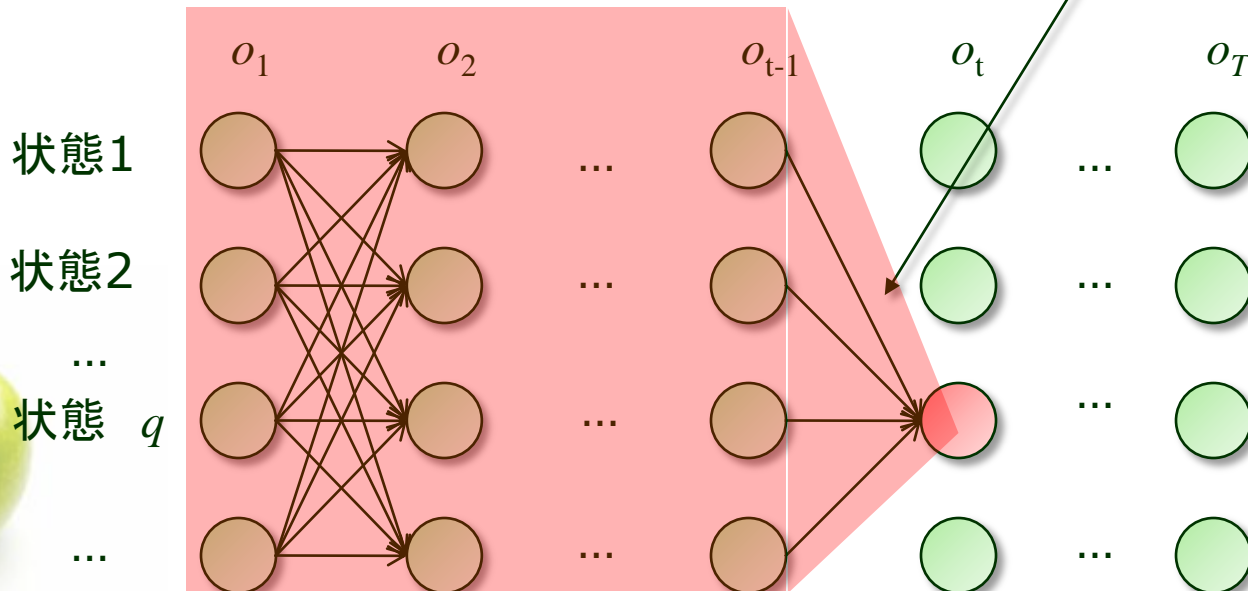
トレリス



前向きアルゴリズム (2/3)

$$\begin{aligned}
 \alpha(t, q) &= \sum_{q_1 \in Q, \dots, q_{t-1} \in Q} \pi_{q_1} b_{q_1, o_1} \left(\prod_{u=2}^{t-1} a_{q_{u-1}, q_u} b_{q_u, o_u} \right) a_{q_{t-1}, q} b_{q, o_t} \\
 &= \sum_{q_{t-1} \in Q} \left\{ \sum_{q_1 \in Q, \dots, q_{t-2} \in Q} \left\{ \pi_{q_1} b_{q_1, o_1} \left(\prod_{u=2}^{t-2} a_{q_{u-1}, q_u} b_{q_u, o_u} \right) a_{q_{t-2}, q_{t-1}} b_{q_{t-1}, o_{t-1}} \right\} a_{q_{t-1}, q} b_{q, o_t} \right\} \\
 &= \sum_{q_{t-1} \in Q} \alpha(t-1, q_{t-1}) a_{q_{t-1}, q} b_{q, o_t}
 \end{aligned}$$

全ての遷移確率の和



前向きアルゴリズム (3/3)

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

for $t = 2$ to T

for $q \in Q$

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



後ろ向きアルゴリズム (1/3)

- 前向きアルゴリズムを逆方向に適用
 - 文末から文頭に向かって実行
 - 向きが逆になることを除けば前向きアルゴリズムとまったく同じ
- $\beta(t, q)$: 時刻 t に状態 q から始める状態遷移によって $o_{t+1} \dots o_T$ まで出力する確率

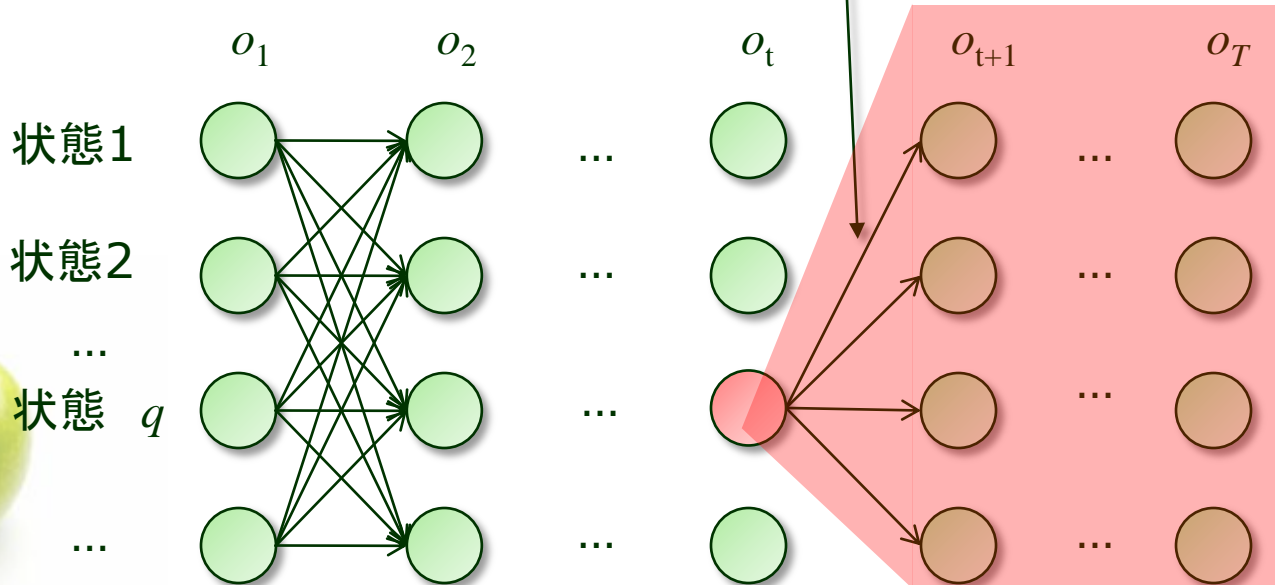
$$\beta(t, q) = \sum_{q_{t+1} \in Q, \dots, q_T \in Q} a_{q, q_{t+1}} b_{q_{t+1}, o_{t+1}} \prod_{u=t+2}^T a_{q_{u-1}, q_u} b_{q_u, o_u}$$



後ろ向きアルゴリズム (2/3)

$$\begin{aligned}
 \beta(t, q) &= \sum_{q_{t+1} \in Q, \dots, q_T \in Q} a_{q, q_{t+1}} b_{q_{t+1}, o_{t+1}} \prod_{u=t+2}^T a_{q_{u-1}, q_u} b_{q_u, o_u} \\
 &= \sum_{q_{t+1} \in Q} \left\{ a_{q, q_{t+1}} b_{q_{t+1}, o_{t+1}} \sum_{q_{t+2} \in Q, \dots, q_n \in Q} \left\{ a_{q_{t+1}, q_{t+2}} b_{q_{t+2}, o_{t+2}} \prod_{u=t+3}^T a_{q_{u-1}, q_u} b_{q_u, o_u} \right\} \right\} \\
 &= \sum_{q_{t+1} \in Q} a_{q, q_{t+1}} b_{q_{t+1}, o_{t+1}} \beta(t+1, q_{t+1})
 \end{aligned}$$

全ての遷移確率の和



後ろ向きアルゴリズム (3/3)

$\beta[T, q] := 1$ (for all q)

for $t := T-1$ to 1

for $q \in Q$

$$\beta[t, q] := \sum_{q' \in Q} \{a[q, q']b[q', o_{t+1}]\beta[t+1, q']\}$$



前向き後ろ向きアルゴリズムの 導出



前向き後ろ向きアルゴリズム: 生成確率

- 生成確率

$$p(o_1 \cdots o_{T_i}) = \sum_{q_1 \in Q, \dots, q_{T_i} \in Q} p(q_1 o_1 \cdots q_{T_i} o_{T_i}) = \sum_{q \in Q} \alpha(T_i, q)$$



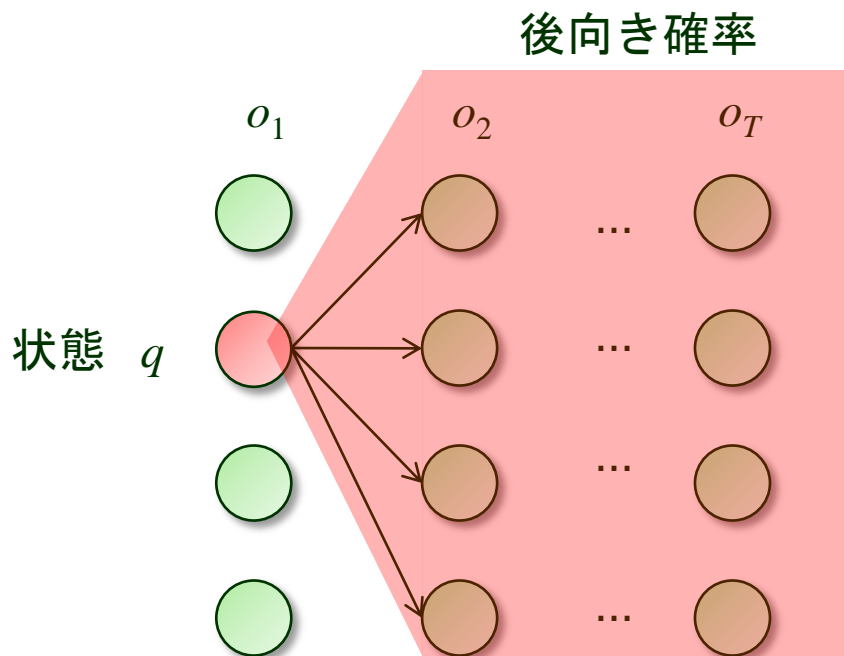
前向き後ろ向きアルゴリズム: π'

$$\begin{aligned}\pi'_q &= \frac{1}{n} \sum_{i=1}^n \sum_{q_1 \in Q, \dots, q_{T_i} \in Q} p(q_1 \cdots q_{T_i} | o_1 \cdots o_{T_i}; \theta) [q_1 = q] \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{p(o_1 \cdots o_{T_i}; \theta)} \sum_{q_2 \in Q, \dots, q_{T_i} \in Q} \pi_q b_{q, o_1} a_{q, q_2} b_{q_2, o_2} \prod_{t=3}^{T_i} a_{q_{t-1}, q_t} \prod_{t=3}^{T_i} b_{q_t, o_t} \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{p(o_1 \cdots o_{T_i}; \theta)} \pi_q b_{q, o_1} \beta(1, q) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{\alpha(1, q) \beta(1, q)}{\sum_{q' \in Q} \alpha(T_i, q')}\end{aligned}$$



前向き後ろ向きアルゴリズム: π'

$$\pi'_q = \frac{1}{n} \sum_{i=1}^n \frac{1}{p(o_1 \cdots o_{T_i}; \theta)} \pi_q b_{q, o_1} \beta(1, q) = \frac{1}{n} \sum_{i=1}^n \frac{\alpha(1, q) \beta(1, q)}{\sum_{q' \in Q} \alpha(T_i, q')}$$



前向き後ろ向きアルゴリズム: a'

$$\begin{aligned}
 a'_{q,r} \text{の分子部分} &= \sum_{i=1}^n \sum_{q_1 \in Q, \dots, q_{T_i} \in Q} p(q_1 \cdots q_{T_i} \mid o_1 \cdots o_{T_i}; \theta) \left(\sum_{t=2}^{T_i} [q = q_{t-1}][r = q_t] \right) \\
 &= \sum_{i=1}^n \frac{1}{p(o_1 \cdots o_{T_i}; \theta)} \sum_{q_1 \in Q, \dots, q_{T_i} \in Q} \pi_{q_1} \prod_{u=2}^{T_i} a_{q_{u-1}, q_u} \prod_{u=1}^{T_i} b_{q_u, o_u} \left(\sum_{t=2}^{T_i} [q = q_{t-1}][r = q_t] \right) \\
 &= \sum_{i=1}^n \frac{1}{p(o_1 \cdots o_{T_i}; \theta)} \sum_{t=2}^{T_i} \sum_{q_1 \in Q, \dots, q_{T_i} \in Q} \pi_{q_1} \prod_{u=2}^{T_i} a_{q_{u-1}, q_u} \prod_{u=1}^{T_i} b_{q_u, o_u} [q = q_{t-1}][r = q_t] \\
 &= \sum_{i=1}^n \frac{1}{p(o_1 \cdots o_{T_i}; \theta)} \sum_{t=2}^{T_i} \sum_{q_1 \in Q, \dots, q_{T_i} \in Q} \pi_{q_1} b_{q_1, o_1} \left(\prod_{u=2}^{t-2} a_{q_{u-1}, q_u} b_{q_u, o_u} \right) a_{q_{t-2}, q_{t-1}} b_{q_{t-1}, o_{t-1}} a_{q_{t-1}, q_t} b_{q_t, o_t} a_{q_t, q_{t+1}} b_{q_{t+1}, o_{t+1}} \left(\prod_{u=t+2}^{T_i} a_{q_{u-1}, q_u} b_{q_u, o_u} \right) [q = q_{t-1}][r = q_t] \\
 &= \sum_{i=1}^n \frac{1}{p(o_1 \cdots o_{T_i}; \theta)} \sum_{t=2}^{T_i} \sum_{q_1 \in Q, \dots, q_{t-2} \in Q} \sum_{q_{t+1} \in Q, \dots, q_{T_i} \in Q} \pi_{q_1} b_{q_1, o_1} \left(\prod_{u=2}^{t-2} a_{q_{u-1}, q_u} b_{q_u, o_u} \right) a_{q_{t-2}, q} b_{q, o_{t-1}} a_{q,r} b_{r, o_t} a_{r, q_{t+1}} b_{q_{t+1}, o_{t+1}} \left(\prod_{u=t+2}^{T_i} a_{q_{u-1}, q_u} b_{q_u, o_u} \right) \\
 &= \sum_{i=1}^n \frac{1}{p(o_1 \cdots o_{T_i}; \theta)} \sum_{t=2}^{T_i} \sum_{q_1 \in Q, \dots, q_{t-2} \in Q} \pi_{q_1} b_{q_1, o_1} \left(\prod_{u=2}^{t-2} a_{q_{u-1}, q_u} b_{q_u, o_u} \right) a_{q_{t-2}, q} b_{q, o_{t-1}} a_{q,r} b_{r, o_t} \sum_{q_{t+1} \in Q, \dots, q_{T_i} \in Q} a_{r, q_{t+1}} b_{q_{t+1}, o_{t+1}} \left(\prod_{u=t+2}^{T_i} a_{q_{u-1}, q_u} b_{q_u, o_u} \right) \\
 &= \sum_{i=1}^n \frac{1}{p(o_1 \cdots o_{T_i}; \theta)} \sum_{t=2}^{T_i} \sum_{q_1 \in Q, \dots, q_{t-2} \in Q} \pi_{q_1} b_{q_1, o_1} \left(\prod_{u=2}^{t-2} a_{q_{u-1}, q_u} b_{q_u, o_u} \right) a_{q_{t-2}, q} b_{q, o_{t-1}} a_{q,r} b_{r, o_t} \beta(t, r) \\
 &= \sum_{i=1}^n \frac{1}{p(o_1 \cdots o_{T_i}; \theta)} \sum_{t=2}^{T_i} a_{q,r} b_{r, o_t} \beta(t, r) \sum_{q_1 \in Q, \dots, q_{t-2} \in Q} \pi_{q_1} b_{q_1, o_1} \left(\prod_{u=2}^{t-2} a_{q_{u-1}, q_u} b_{q_u, o_u} \right) a_{q_{t-2}, q} b_{q, o_{t-1}} \\
 &= \sum_{i=1}^n \frac{1}{p(o_1 \cdots o_{T_i}; \theta)} \sum_{t=2}^{T_i} \alpha(t-1, q) a_{q,r} b_{r, o_t} \beta(t, r)
 \end{aligned}$$

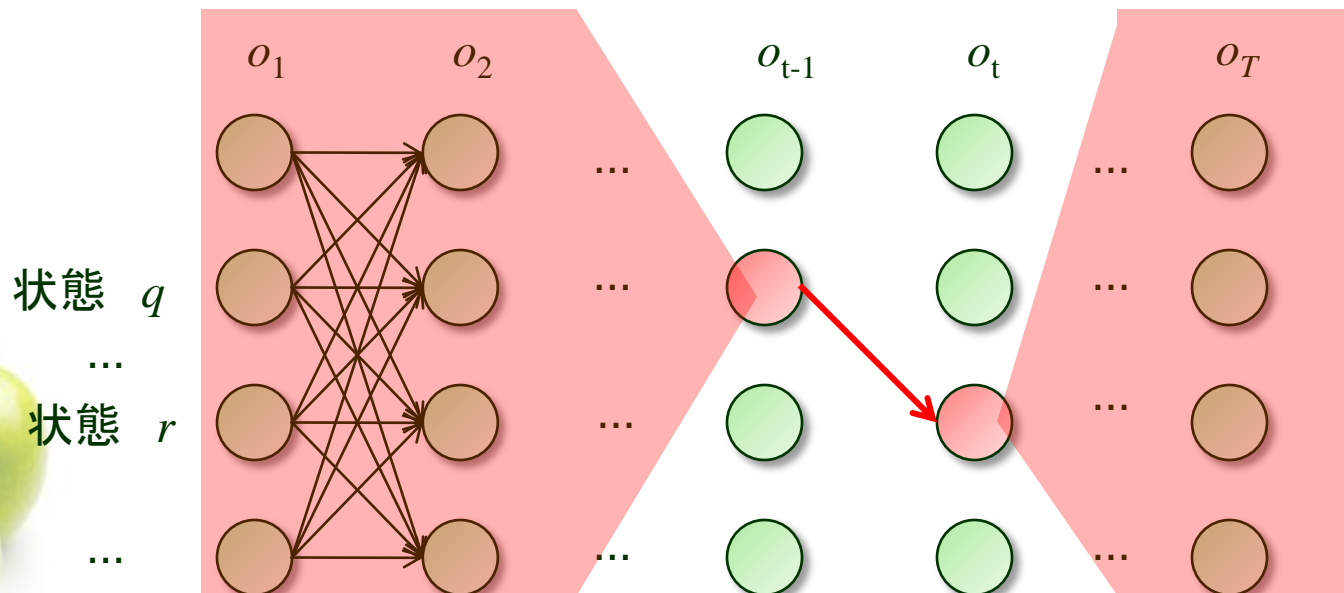
前向き後ろ向きアルゴリズム: a'

$\gamma(t, q, r)$: 記号列 $o_1 \dots o_T$ に対し、時刻 $t-1$ から時刻 t に状態 q から状態 r に遷移する確率

$$\gamma(t, q, r) = \frac{\alpha(t-1, q) a_{q,r} b_{r, o_t} \beta(t, r)}{\sum_{q' \in Q} \alpha(T, q')}$$

前向き確率

後ろ向き確率



前向き後ろ向きアルゴリズム: a'

$$\begin{aligned} a'_{q,r} &= \frac{\sum_{i=1}^n \sum_{q_1 \in Q, \dots, q_{T_i} \in Q} p(q_1 \cdots q_{T_i} \mid o_1 \cdots o_{T_i}; \theta) \left(\sum_{t=2}^{T_i} [q = q_{t-1}] [r = q_t] \right)}{\sum_{r' \in Q} \sum_{i=1}^n \sum_{q_1 \in Q, \dots, q_{T_i} \in Q} p(q_1 \cdots q_{T_i} \mid o_1 \cdots o_{T_i}; \theta) \left(\sum_{t=2}^{T_i} [q = q_{t-1}] [r' = q_t] \right)} \\ &= \frac{\sum_{i=1}^n \sum_{t=2}^{T_i} \gamma(t, q, r)}{\sum_{r' \in Q} \sum_{i=1}^n \sum_{t=2}^{T_i} \gamma(t, q, r')} \end{aligned}$$



前向き後ろ向きアルゴリズム: b'

$$\begin{aligned}
 b'_{q,o} \text{の分子部分} &= \sum_{i=1}^n \sum_{q_1 \in \mathcal{Q}, \dots, q_{T_i} \in \mathcal{Q}} p(q_1 \dots q_{T_i} | o_1 \dots o_{T_i}; \theta) \left(\sum_{t=1}^{T_i} [q = q_t][o = o_t] \right) \\
 &= \sum_{i=1}^n \frac{1}{p(o_1 \dots o_{T_i}; \theta)} \sum_{t=1}^{T_i} \sum_{q_1 \in \mathcal{Q}, \dots, q_{T_i} \in \mathcal{Q}} \pi_{q_1} \prod_{u=2}^{T_i} a_{q_{u-1}, q_u} \prod_{u=1}^{T_i} b_{q_u, o_u} [q = q_t][o = o_t] \\
 &= \sum_{i=1}^n \frac{1}{p(o_1 \dots o_{T_i}; \theta)} \sum_{t=1}^{T_i} \sum_{q_1 \in \mathcal{Q}, \dots, q_{T_i} \in \mathcal{Q}} \pi_{q_1} b_{q_1, o_1} \left(\prod_{u=2}^{t-1} a_{q_{u-1}, q_u} b_{q_u, o_u} \right) a_{q_{t-1}, q_t} b_{q_t, o_t} a_{q_t, q_{t+1}} b_{q_{t+1}, o_{t+1}} \left(\prod_{t=u+2}^{T_i} a_{q_{t-1}, q_t} b_{q_t, o_t} \right) [q = q_t][o = o_t] \\
 &= \sum_{i=1}^n \frac{1}{p(o_1 \dots o_{T_i}; \theta)} \sum_{t=1}^{T_i} \sum_{q_1 \in \mathcal{Q}, \dots, q_{t-1} \in \mathcal{Q}} \sum_{q_{t+1} \in \mathcal{Q}, \dots, q_{T_i} \in \mathcal{Q}} \pi_{q_1} b_{q_1, o_1} \left(\prod_{u=2}^{t-1} a_{q_{u-1}, q_u} b_{q_u, o_u} \right) a_{q_{t-1}, q_t} b_{q_t, o_t} a_{q_t, q_{t+1}} b_{q_{t+1}, o_{t+1}} \left(\prod_{t=u+2}^{T_i} a_{q_{t-1}, q_t} b_{q_t, o_t} \right) [o = o_t] \\
 &= \sum_{i=1}^n \frac{1}{p(o_1 \dots o_{T_i}; \theta)} \sum_{t=1}^{T_i} \sum_{q_1 \in \mathcal{Q}, \dots, q_{t-1} \in \mathcal{Q}} \pi_{q_1} b_{q_1, o_1} \left(\prod_{u=2}^{t-1} a_{q_{u-1}, q_u} b_{q_u, o_u} \right) a_{q_{t-1}, q_t} b_{q_t, o_t} [o = o_t] \sum_{q_{t+1} \in \mathcal{Q}, \dots, q_{T_i} \in \mathcal{Q}} a_{q_t, q_{t+1}} b_{q_{t+1}, o_{t+1}} \left(\prod_{t=u+2}^{T_i} a_{q_{t-1}, q_t} b_{q_t, o_t} \right) \\
 &= \sum_{i=1}^n \frac{1}{p(o_1 \dots o_{T_i}; \theta)} \sum_{t=1}^{T_i} \sum_{q_1 \in \mathcal{Q}, \dots, q_{t-1} \in \mathcal{Q}} \pi_{q_1} b_{q_1, o_1} \left(\prod_{u=2}^{t-1} a_{q_{u-1}, q_u} b_{q_u, o_u} \right) a_{q_{t-1}, q_t} b_{q_t, o_t} [o = o_t] \beta(t, q) \\
 &= \sum_{i=1}^n \frac{1}{p(o_1 \dots o_{T_i}; \theta)} \sum_{t=1}^{T_i} [o = o_t] \beta(t, q) \sum_{q_1 \in \mathcal{Q}, \dots, q_{t-1} \in \mathcal{Q}} \pi_{q_1} b_{q_1, o_1} \left(\prod_{u=2}^{t-1} a_{q_{u-1}, q_u} b_{q_u, o_u} \right) a_{q_{t-1}, q_t} b_{q_t, o_t} \\
 &= \sum_{i=1}^n \frac{1}{p(o_1 \dots o_{T_i}; \theta)} \sum_{t=1}^{T_i} [o = o_t] \alpha(t, q) \beta(t, q)
 \end{aligned}$$

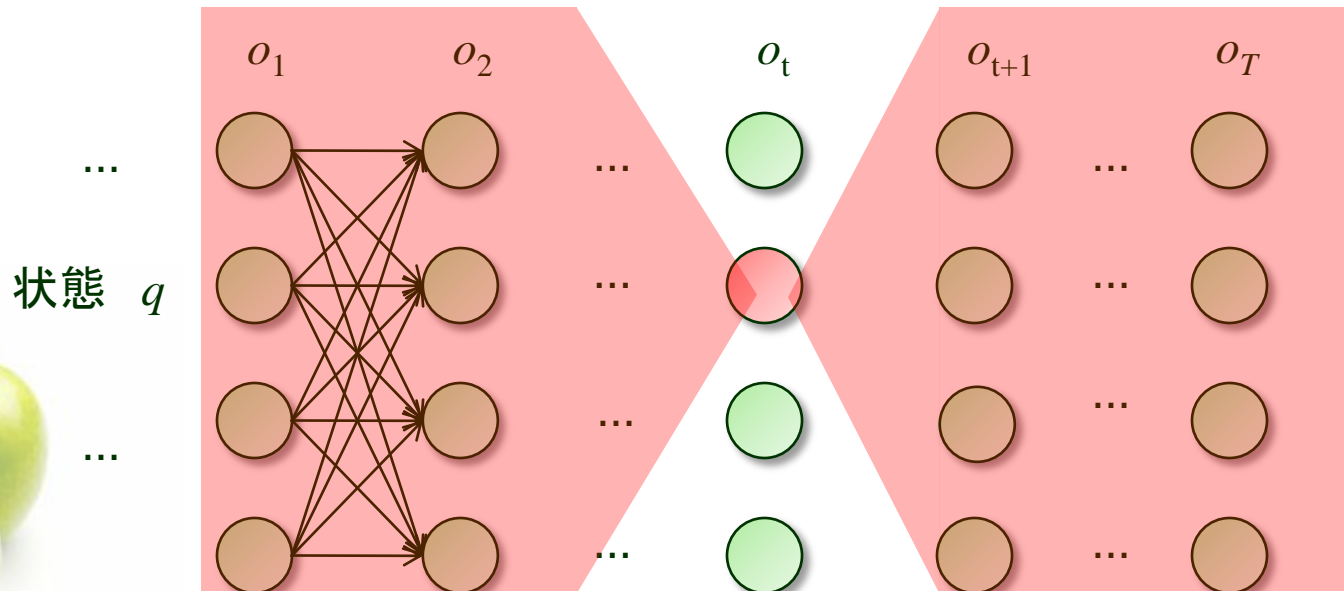
前向き後ろ向きアルゴリズム: b'

$\gamma(t, q)$: 記号列 $o_1 \dots o_T$ に対し、時刻 t に状態 q に滞在する確率

$$\gamma(t, q) = \frac{\alpha(t, q)\beta(t, q)}{\sum_{q' \in Q} \alpha(T, q')}$$

前向き確率

後向き確率



前向き後ろ向きアルゴリズム: b'

$$b'_{q,o} = \frac{\sum_{i=1}^n \sum_{q_1 \in Q, \dots, q_{T_i} \in Q} p(q_1 \dots q_{T_i} | o_1 \dots o_{T_i}; \theta) \left(\sum_{t=1}^{T_i} [q = q_t][o = o_t] \right)}{\sum_{o' \in \Sigma} \sum_{i=1}^n \sum_{q_1 \in Q, \dots, q_{T_i} \in Q} p(q_1 \dots q_{T_i} | o_1 \dots o_{T_i}; \theta) \left(\sum_{t=1}^{T_i} [q = q_t][o' = o_t] \right)}$$

$$= \frac{\sum_{i=1}^n \sum_{t=1}^{T_i} [o = o_t] \gamma(t, q)}{\sum_{o' \in \Sigma} \sum_{i=1}^n \sum_{t=1}^{T_i} [o' = o_t] \gamma(t, q)}$$

$$= \frac{\sum_{i=1}^n \sum_{t=1}^{T_i} [o = o_t] \gamma(t, q)}{\sum_{i=1}^n \sum_{t=1}^{T_i} \gamma(t, q)}$$

ちなみに、

$$\pi'_q = \frac{1}{n} \sum_{i=1}^n \frac{\alpha(1, q) \beta(1, q)}{\sum_{q' \in Q} \alpha(T_i, q')} = \frac{1}{n} \gamma(1, q)$$



前向き後向きアルゴリズム まとめ(1/2)

$\alpha(t, q)$: $o_1 \dots o_t$ を出力し、 o_t が出力される時(時刻 t)に状態 q である確率

$$\begin{aligned}\alpha(t, q) &= \sum_{q_1 \in Q, \dots, q_{t-1} \in Q} \pi_{q_1} b_{q_1, o_1} \left(\prod_{u=2}^{t-1} a_{q_{u-1}, q_u} b_{q_u, o_u} \right) a_{q_{t-1}, q} b_{q, o_t} \\ &= \sum_{q_{t-1} \in Q} \alpha(t-1, q_{t-1}) a_{q_{t-1}, q} b_{q, o_t}\end{aligned}$$

$\beta(t, q)$: 時刻 t に状態 q から始める状態遷移によって $o_{t+1} \dots o_T$ まで出力する確率

$$\begin{aligned}\beta(t, q) &= \sum_{q_{t+1} \in Q, \dots, q_T \in Q} a_{q, q_{t+1}} b_{q_{t+1}, o_{t+1}} \prod_{u=t+2}^T a_{q_{u-1}, q_u} b_{q_u, o_u} \\ &= \sum_{q_{t+1} \in Q} a_{q, q_{t+1}} b_{q_{t+1}, o_{t+1}} \beta(t+1, q_{t+1})\end{aligned}$$

$$\gamma(t, q) = \frac{\alpha(t, q) \beta(t, q)}{\sum_{q' \in Q} \alpha(T, q')}$$

$$\gamma(t, q, r) = \frac{\alpha(t-1, q) a_{q, r} b_{r, o_t} \beta(t, r)}{\sum_{q' \in Q} \alpha(T, q')}$$

前向き後向きアルゴリズム まとめ(2/2)

あるパラメータ π, a, b が与えられているとする

- α を求める(前向きアルゴリズム)
- β を求める(後向きアルゴリズム)
 - それぞれ動的計画法。計算量は $O(T)$ ※ビタビアルゴリズムも計算量は $O(T)$ 。ビタビアルゴリズムのおよそ3倍の計算量。
- 次のように π, a, b を更新する

$$\pi'_q = \frac{1}{n} (\text{先頭が} q \text{となる回数の期待値}) = \frac{1}{n} \sum_{i=1}^n \gamma(1, q)$$

$$a'_{q,r} = \frac{\text{状態} q \text{から状態} r \text{へ遷移する回数の期待値}}{\text{状態} q \text{から遷移する回数の期待値}} = \frac{\sum_{i=1}^n \sum_{t=2}^{T_i} \gamma(t, q, r)}{\sum_{r' \in Q} \sum_{i=1}^n \sum_{t=2}^{T_i} \gamma(t, q, r')}$$

$$b'_{q,o} = \frac{\text{状態} q \text{に滞在し記号} o \text{を出力する回数の期待値}}{\text{状態} q \text{に滞在する回数の期待値}} = \frac{\sum_{i=1}^n \sum_{t=1}^{T_i} [o = o_t] \gamma(t, q)}{\sum_{i=1}^n \sum_{t=1}^{T_i} \gamma(t, q)}$$

Generative Model and Discriminative Model

生成モデルと識別モデル



生成モデルから識別モデルへ

- 生成モデル

- HMM

- 状態遷移による生成
- 状態遷移の確率は一つ前の状態のみに依存
- 任意の文 s と状態列 q に対する同時確率 $p(s, q)$ を計算できる

- PCFG

- CFGによる生成のステップ（書換規則の適用 $A \rightarrow B C$ ）の確率を $p(B C | A)$ とする
- CFGによる生成の過程の確率を、独立な事象の(条件付き)確率の積とする
- 任意の文 s と構文木 t に対する同時確率 $p(s, t)$ を計算できる



生成モデルの改良の先に

- 構文木の分岐の外の情報を条件部に導入
 - PCFG: $p(BC | A, hw) [A_{hw} \rightarrow B_{hw} C]$ (hw: 主辞)
 - HMM: $p(q_j | q_{j-1}, q_{j-2})$
- 素性 (特徴) という考え方
条件部をどんどんリッチにして線形補間をとればよいのでは? → 構文木のノード x の確率
 $p(x) = p(x | x$ 周辺の状態)



生成モデルの問題点

- 生成モデルの問題点

- 独立性の仮定

- PCFGでは、各構文木ノードは独立な(条件付き) 試行により生成されていると仮定している
- 例えば、(S (NP John) (VP eats (NP apples)))という構文木は、以下の独立な (条件付き) 事象に分解されている
 - SからNP VPを生成
 - NPからJohnを生成
 - VPからeats NPを生成
 - NPからapplesを生成
- 結果、上記のNPの文脈(S→NP VP)から導出されたNPか (VP→eats NP)から導出されたNPかわからなくなっている→主語のNPなのか目的語のNPなのかわからない。



生成モデルの問題点

- 生成モデル

$$\tilde{t} = \arg \max_t p(t | s; \theta) = \arg \max_t p(s, t; \theta)$$

- しかし、

$$p(s, t; \theta) = p(t | s; \theta) p(s; \theta)$$

であるから $p(s; \theta)$ が計算できてしまう分、冗長なモデルになっている。間接的に分類問題を解いている。

- 独立性を仮定した事象に分解しないと同時確率を計算することができない



識別モデル

- 識別モデル
直接

$$\tilde{t} = \arg \max_t p(t | s; \theta)$$

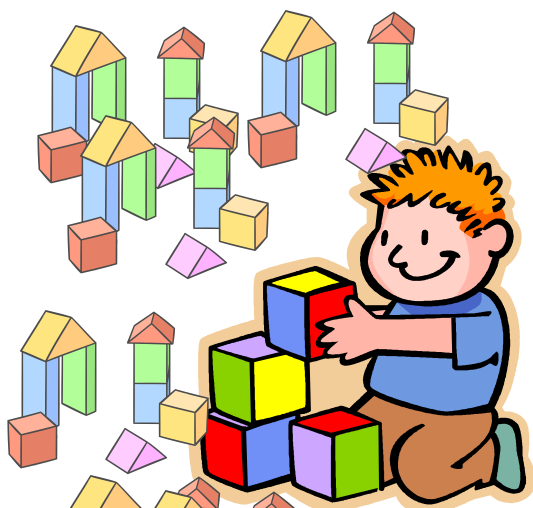
を解く

- 独立な事象を仮定しない
- 「条件部の確率」をモデルにいれない

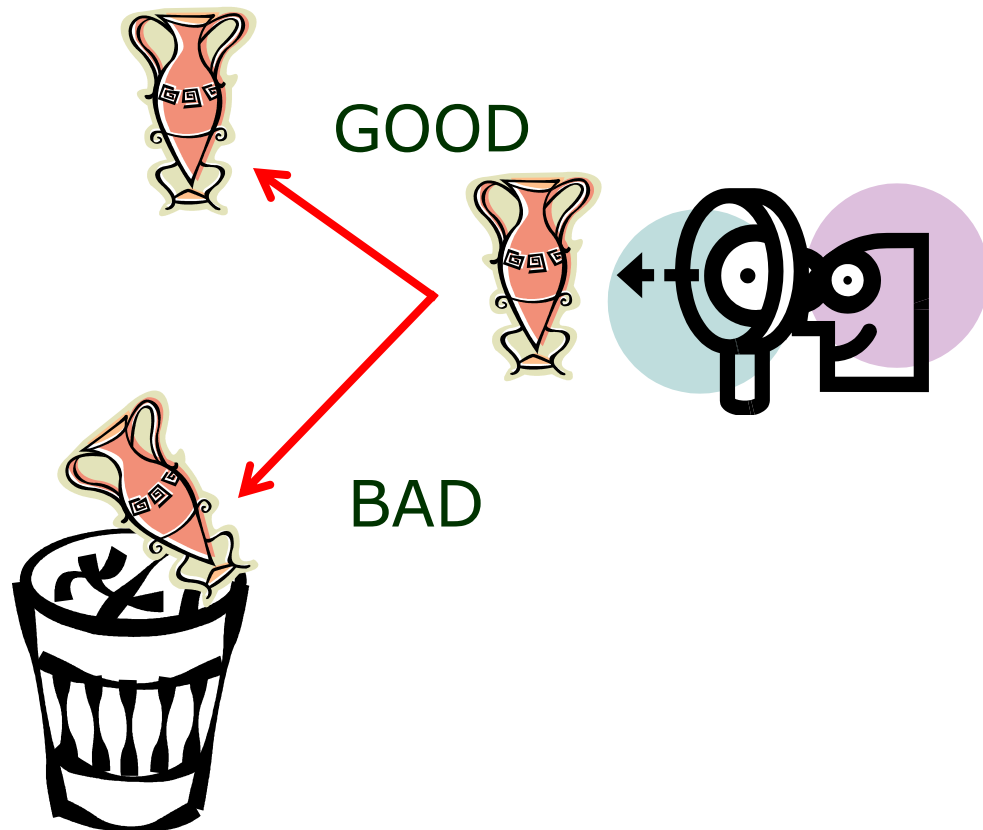


生成モデルと識別モデル (イメージ)

生成モデル



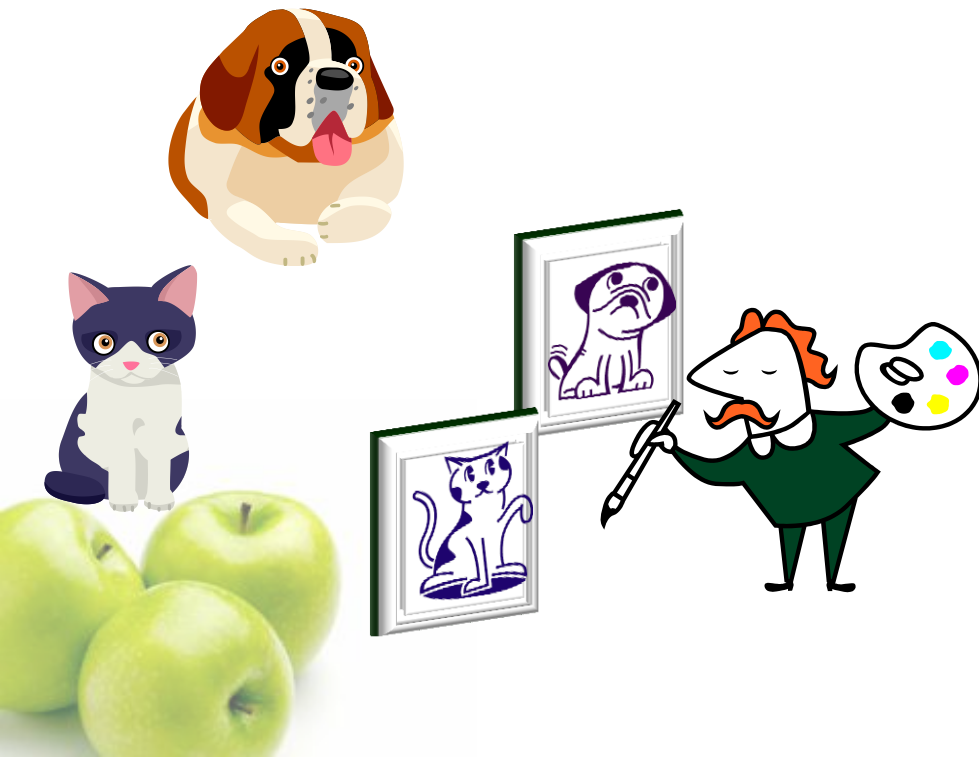
識別モデル



生成モデルと識別モデル (イメージ2)

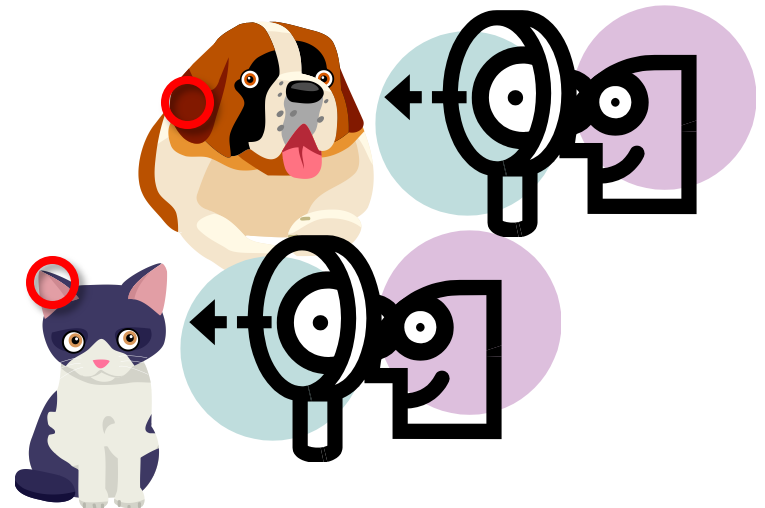
生成モデル

- 絵を描いて全体像の比較



識別モデル

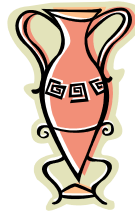
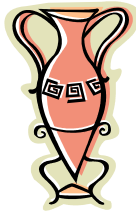
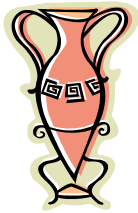
- それぞれの特徴を比較
 - 鼻の位置
 - 耳の形
 - 体の大きさ
 - 舌の表面



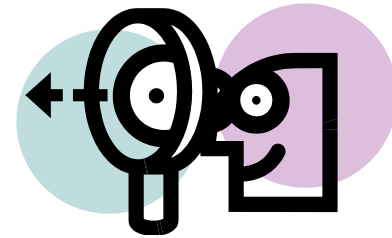
識別するための訓練

- 教師付学習

- 良い例と悪い例を与えて、どこに注目すればより良く識別出来るのか学習



good examples



bad examples

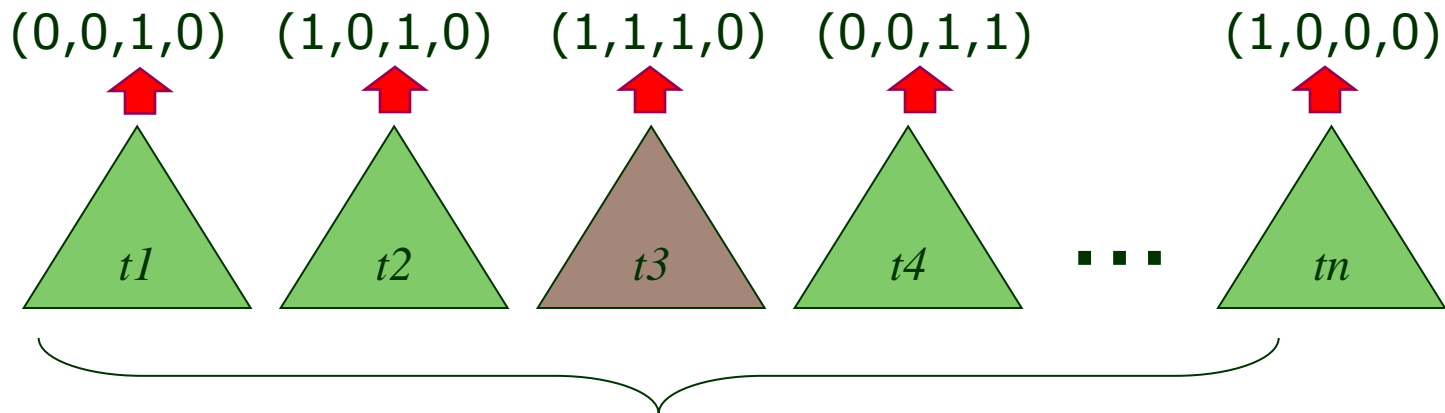


識別モデル

$$p(t | s)$$

$S = \text{"A blue eye girl with white hair and skin walked"}$

素性ベクトル
(特徴ベクトル)



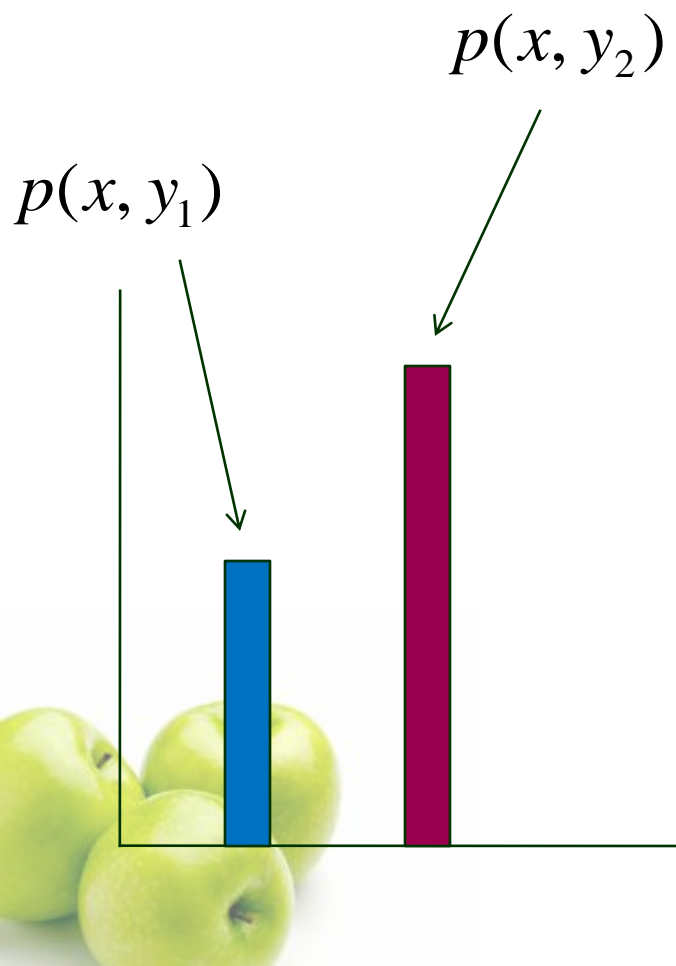
文法Gによりsから導出出来る全ての構文木集合

$p(t_3|s)$ は $t_1, t_2, t_3, \dots, t_n$ から t_3 を選択する確率

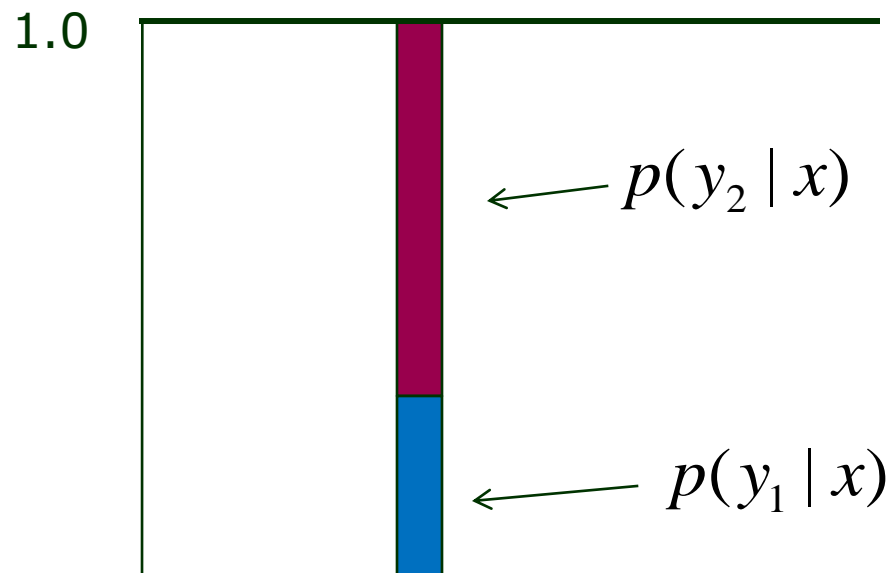


生成モデルと識別モデル

生成モデル



識別モデル



どちらが優秀なのか？

Tom Minka (2005) Discriminative models, not discriminative training, MSR-TR-2005-144, 1 p.

x を入力(文)、 y を出力(構文木)としたときのパラメータ推定

- 生成モデル

$$p(x, y; \theta) = p(x | y; \theta) p(y; \theta)$$

- 識別モデル

$$p(y | x; \theta)$$

$$\theta = \theta'$$

- 一般モデル

$$p(x, y; \theta, \theta') = p(y | x; \theta) p(x; \theta')$$



まとめ

- HMMの前向き後向きアルゴリズム
 - 前向き後向きアルゴリズム
 - 動的計画法
 - 計算量は $O(T)$ ※ビタビアルゴリズムも計算量は $O(T)$ 。ビタビアルゴリズムのおよそ3倍の計算量。
 - 隠れ状態を全て並べあげることなく、トレリス上で期待値を計算できる
 - ある状態からある状態への遷移回数の期待値
 - ある状態からある記号を出力する回数の期待値
 - 生成モデルと識別モデル
 - 今までの講義資料

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

