知的情報処理システム特論　第10回

二宮　崇

今日の講義の予定

- 系列ラベリングのためのHMM
  - 学習
    - 最尤推定
    - HMMの教師付き学習

- 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

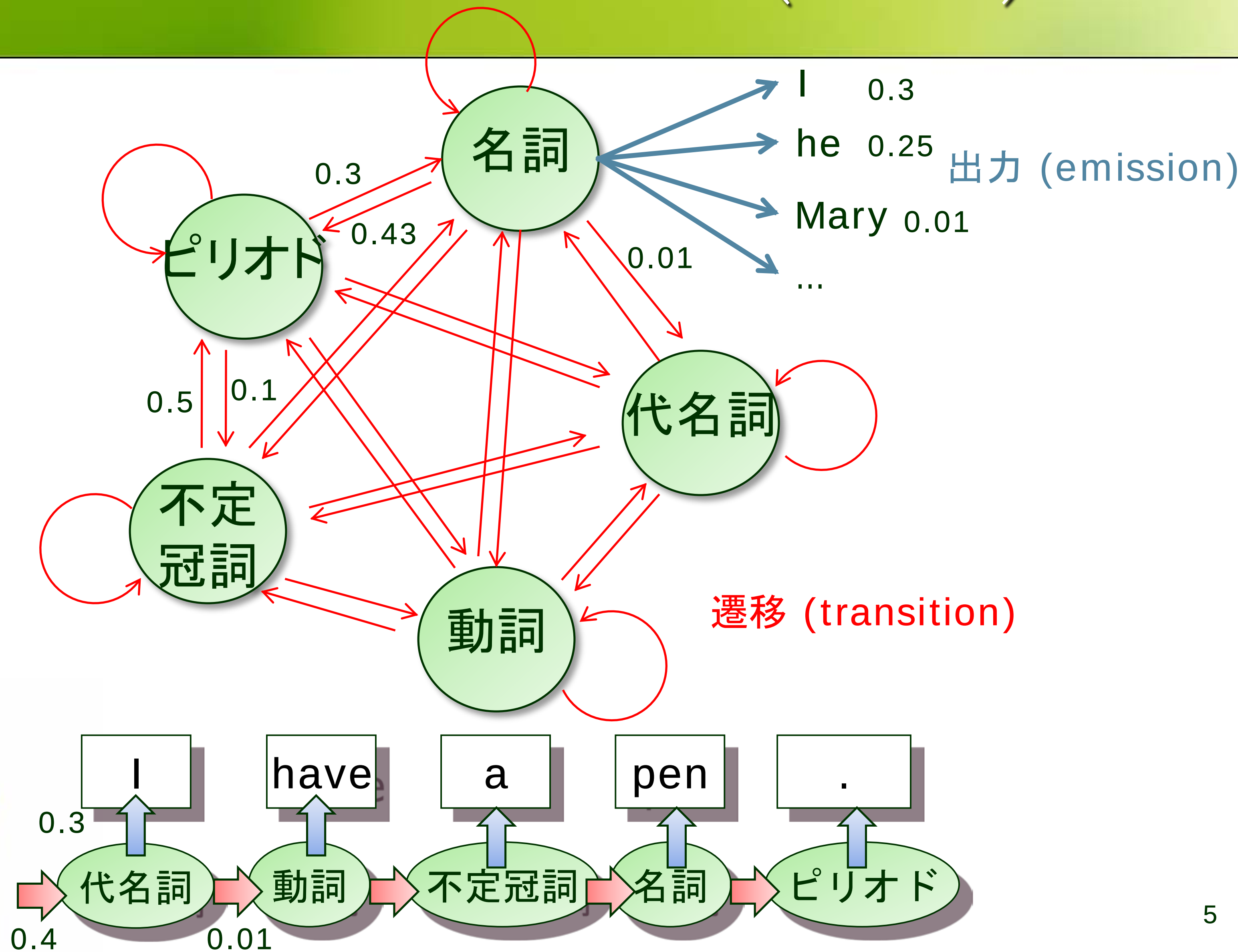
I | have | a | pen | .

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)

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

学習 (パラメータ推定): HMMの教師付学習

HMMの教師付学習
Supervised Learning of HMMs

- パラメータ推定
  - 訓練データ (入力)

訓練データ$: x_1 x_2 \cdots x_n, y_1 y_2 \cdots y_n$

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

y_i : x_iに対応する正解状態列(正解品詞列)。$y_i = q_{i1} q_{i2} q_{i3} \cdots q_{iT_i}$とする。

  - パラメータ (出力)

π_q … |Q|個の変数

$a_{q,r}$ … |Q|×|Q|個の変数

$b_{q,o}$ … |Q|×|Σ|個の変数

HMMの教師付学習 Supervised Learning of HMMs

- パラメータ推定

$$
\begin{aligned}
\pi, a, b &= \arg\max_{\pi,a,b} \prod_{i=1}^{n} p(x_i, y_i) \\
&= \arg\max_{\pi,a,b} \prod_{i=1}^{n} p(q_{i1} o_{i1} q_{i2} o_{i2} \cdots q_{iT_i} o_{iT_i}) \\
&= \arg\max_{\pi,a,b} \prod_{i=1}^{n} \pi_{q_{i1}} \prod_{t=2}^{T_i} a_{q_{i(t-1)}, q_{it}} \prod_{t=1}^{T_i} b_{q_{it}, o_{it}}
\end{aligned}
$$

HMMの教師付学習
Supervised Learning of HMMs

- パラメータ推定

$$l(\pi, a, b) = \prod_{i=1}^{n} \pi_{q_{i1}} \prod_{t=2}^{T_i} a_{q_{i(t-1)}, q_{it}} \prod_{t=1}^{T_i} b_{q_{it}, o_{it}}$$

$$\log l(\pi, a, b) = \sum_{i=1}^{n} \left(\log \pi_{q_{i1}} + \sum_{t=2}^{T_i} \log a_{q_{i(t-1)}, q_{it}} + \sum_{t=1}^{T_i} \log b_{q_{it}, o_{it}} \right)$$

制約付き最適化問題

$$\underset{\pi, a, b}{\arg\max} \log l(\pi, a, b)$$

$$\text{s.t.} \quad \sum_{q \in Q} \pi_q = 1$$

$$\sum_{r \in Q} a_{q,r} = 1 \quad (\text{for all } q)$$

$$\sum_{o \in \Sigma} b_{q,o} = 1 \quad (\text{for all } q)$$

HMMの教師付学習
Supervised Learning of HMMs

- ラグランジュの未定乗数法

ラグランジュ関数

$$L(\pi,a,b)$$

$$=\log l(\pi,a,b)-\rho\left(1-\sum_{q\in Q}\pi_q\right)-\sum_{q\in Q}\alpha_q\left(1-\sum_{r\in Q}a_{q,r}\right)-\sum_{q\in Q}\beta_q\left(1-\sum_{o\in\Sigma}b_{q,o}\right)$$

$$=\sum_{i=1}^{n}\left(\log\pi_{q_{i1}}+\sum_{t=2}^{T_i}\log a_{q_{i(t-1)},q_{it}}+\sum_{t=1}^{T_i}\log b_{q_{it},o_{it}}\right)-\rho\left(1-\sum_{q\in Q}\pi_q\right)-\sum_{q\in Q}\alpha_q\left(1-\sum_{r\in Q}a_{q,r}\right)-\sum_{q\in Q}\beta_q\left(1-\sum_{o\in\Sigma}b_{q,o}\right)$$

ラグランジュ乗数

ρ … 1個の変数

α_q … |Q|個の変数

β_q … |Q|個の変数

HMMの教師付学習
Supervised Learning of HMMs

- π_qを求める

$$\frac{\partial L}{\partial \pi_q} = \frac{C_1(q)}{\pi_q} + \rho = 0$$

ただし、$C_1(q) = \sum_{i=1}^{n} [q_{i1} = q]$

$$\pi_q = \frac{C_1(q)}{-\rho} \qquad \cdots(1)$$

ところで$\sum_{q \in Q} \pi_q = 1$であるから

$$\sum_{q \in Q} \pi_q = \frac{\sum_{q \in Q} C_1(q)}{-\rho} = \frac{n}{-\rho} = 1$$

よって、$\rho = -n$。 これを式(1)に代入して、

$$\pi_q = \frac{C_1(q)}{n}$$

アイバーソンの記法
(Iverson bracket)

$$[P] = \begin{cases} 1 & \text{if } P \text{ is true} \\ 0 & \text{otherwise} \end{cases}$$

$C_1(q)$は訓練データ中で、文の先頭がqになっている回数

HMMの教師付学習 Supervised Learning of HMMs

- $a_{q,r}$を求める

$$\frac{\partial L}{\partial a_{q,r}} = \frac{C(q,r)}{a_{q,r}} + \alpha_q = 0$$

ただし、$C(q,r) = \sum_{i=1}^{n} \sum_{t=2}^{T_i} [q_{i(t-1)} = q][q_{it} = r]$

$C(q,r)$は訓練データ中で、状態qから状態rに遷移した回数

$$a_{q,r} = \frac{C(q,r)}{-\alpha_q} \qquad \cdots(1)$$

ところで$\sum_{r \in Q} a_{q,r} = 1$であるから

$$\sum_{r' \in Q} a_{q,r'} = \frac{\sum_{r' \in Q} C(q,r')}{-\alpha_q} = 1$$

よって、$\alpha_q = -\sum_{r' \in Q} C(q,r')$。これを式(1)に代入して、

$$a_{q,r} = \frac{C(q,r)}{\sum_{r' \in Q} C(q,r')}$$

HMMの教師付学習
Supervised Learning of HMMs

- $b_{q,o}$を求める

$$\frac{\partial L}{\partial b_{q,o}} = \frac{C(q,o)}{b_{q,o}} + \beta_q = 0$$

ただし、$C(q,o) = \sum_{i=1}^{n} \sum_{t=1}^{T_i} [q_{it} = q][o_{it} = o]$

C(q,o)は訓練データ中で、状態qから記号oを出力した回数

$$\mathrm{b}_{q,o} = \frac{C(q,o)}{-\beta_q} \qquad \cdots(1)$$

ところで$\sum_{o \in \Sigma} b_{q,o} = 1$であるから

$$\sum_{o' \in \Sigma} b_{q,o'} = \frac{\sum_{o' \in \Sigma} C(q,o')}{-\beta_q} = 1$$

よって、$\beta_q = -\sum_{o' \in \Sigma} C(q,o')$。式(1)に代入して、

$$b_{q,o} = \frac{C(q,o)}{\sum_{o' \in \Sigma} C(q,o')} = \frac{C(q,o)}{C(q)}$$

ただし、$C(q) = \sum_{i=1}^{n} \sum_{t=1}^{T_i} [q_{it} = q]$

C(q)は訓練データ中で、状態qが出現した回数

HMMの教師付学習
Supervised Learning of HMMs

- パラメータ推定

$$\pi_q = \frac{C_1(q)}{n} = \frac{\text{先頭の状態が}q\text{になっている文の数}}{\text{全文数}}$$

$$a_{q,r} = \frac{C(q,r)}{\sum_{r'\in Q} C(q,r')} = \frac{q\text{から}r\text{に遷移した回数}}{q\text{から任意の}r\text{に遷移した回数}}$$

$$b_{q,o} = \frac{C(q,o)}{\sum_{o'\in \Sigma} C(q,o')} = \frac{C(q,o)}{C(q)} = \frac{q\text{から}o\text{を出力した回数}}{q\text{の出現回数}}$$

ただし、

$$C_1(q) = \sum_{i=1}^{n} [q_{i1} = q]$$

$$C(q,r) = \sum_{i=1}^{n} \sum_{t=2}^{T_i} [q_{i(t-1)} = q][q_{it} = r]$$

$$C(q,o) = \sum_{i=1}^{n} \sum_{t=1}^{T_i} [q_{it} = q][o_{it} = o]$$

$$C(q) = \sum_{i=1}^{n} \sum_{t=1}^{T_i} [q_{it} = q]$$

まとめ

- HMM
  - 学習
    - 最尤推定
    - 教師付き学習
- 資料

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