

言語処理のための機械学習入門

5章 系列ラベリング

5.1-5.3

08t4072I 全 太俊

5.1 準備

- 品詞タグ付け

”Nurture passes nature.”



”Nurture[名詞] passes[動詞] nature[名詞].”

- 系列

いくつかの要素が連なったもの

- 系列ラベリング

系列内のそれぞれの要素にラベルを付けること

系列ラベリングでは他のラベルに依存する場合を扱う

5.2 隠れマルコフモデル(HMM)

- HMM定義

「システムがパラメータ未知のマルコフ過程である」と仮定し、観測可能な情報からその未知のパラメータを推定するモデル

- 応用

音声認識、形態素解析

5.2.1 HMMの導入

- HMMでは、各状態はその直前の状態にのみ依存すると仮定する

$$\begin{aligned} P(x | y) &= P(x_k, y_k | x_{k-1}, y_{k-1}) P(x_{k-1}, y_{k-1} | x_{k-2}, y_{k-2}) \dots \\ & P(x_2, y_2 | x_1, y_1) P(x_1, y_1 | x_0, y_0) \\ &= \prod_i P(x_i, y_i | x_{i-1}, y_{i-1}) \\ &= \prod_i P(x_i | y_i) P(y_i | y_{i-1}) \end{aligned}$$

5.2.2 パラメータ推定

訓練データ

$$D = \{ (x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(|D|)}, y^{(|D|)}) \}$$

対数尤度

$$\begin{aligned} \log P(D) &= \sum_{(x^{(i)}, y^{(i)}) \in D} \log P(x^{(i)}, y^{(i)}) \\ &= \sum_{(x^{(i)}, y^{(i)}) \in D} \left(\sum_j \log P(x_j^{(i)} | y_j^{(j)}) + \sum_j \log P(y_j^{(i)} | y_{j-1}^{(i)}) \right) \\ &= \sum_{x, y} n((x, y), D) \log p_{x|y} + \sum_{y, y'} n((y', y), D) \log q_{y|y'} \end{aligned}$$

5.2.2 パラメータ推定

$n((x,y),D)$ はデータDにおいてxにラベルyが付いて回数
 $n((y',y),D)$ はデータDにおいてy'の次に出現した回数、

$p_{x|y} = P(x|y), q_{y|y'} = P(y|y')$ とする

ラグランジュ法を用いて最大化するパラメータをもとめる

$$P_{x|y} = \frac{n((x,y),D)}{\sum_x n((x,y),D)} \quad q_{y|y'} = \frac{n((y',y),D)}{\sum_y n((y',y),D)}$$

5.2.3 HMMの推論

$$y^{\max} = \underset{p}{\operatorname{argmax}} p(x, y)$$

可能なラベル列の数が膨大であり、すべて列挙することは現実的ではない

Viterbiアルゴリズム

$$\log p(x, y) = \sum_j \log P(x_j, y_j | x_{j-1}, y_{j-1}) \text{ を利用する}$$

できるだけ繰り返さないようにして計算量を抑えている

HMMのためViterbiアルゴリズム (前半)

入力:事例x

for j=2 |x|

for all y_i

$$t(j, y_j) = \max_{y_{j-1}} [\log P(x_j, y_j | x_{j-1}, y_{j-1}) + t(j-1, y_{j-1})]$$

$$s(j, y_j) = \arg \max_{y_{j-1}} [\log P(x_j, y_j | x_{j-1}, y_{j-1}) + t(j-1, y_{j-1})]$$

end for

end for

HMMのためのViterbiアルゴリズム (後半)

入力: $t(|x|, y), s$

$$y_{|x|}^{\max} = \arg \max_y t(|x|, y)$$

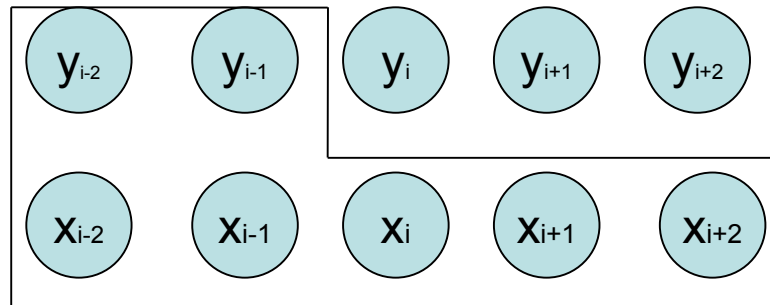
for $j = |x| - 1$ to 1

$$y_j^{\max} = s(j+1, y_{j+1}^{\max})$$

end for

5.3 通常の分類器の逐次適用

y_i を推定する際、 y_{i-1} 、 x_{i-1} 、 x_i 、 x_{i+1} などから素性を抽出する



分類器の逐次適用とHMMの違い

- HMM

x と y だけで一つの確立値が決まる

- 分類器の逐次適用

さまざまな素性を用いて x_i のベクトルを表現できる