

# 言語解析論

講師 竹内孔一

## 本日の内容

- 形態素解析のモデル
  - 確率の計算
  - Viterbi アルゴリズム

## 形態素解析器のモデル化

- HMMによる定式化

$$(\hat{W}, \hat{S}) = \arg \max_{W \subset L, S \subset W} P(W | S)P(S)$$

品詞列  $S = s_1, s_2, \dots, s_n$  全列を見るのは無理!!

単語列  $W = w_1, w_2, \dots, w_n$

マルコフ仮定 (Markov assumption)

1重マルコフモデル(1つ前だけ見よう)

$$P(S) = \prod_{i=1}^n P(s_i | s_{i-1})P(W | S) = \prod_{i=1}^n P(w_i | s_i)$$

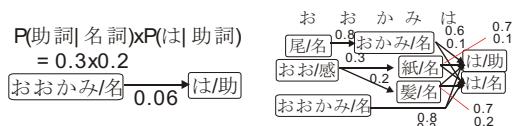
$$(\hat{W}, \hat{S}) = \arg \max_{W \subset L, S \subset W} \prod_{i=1}^n P(w_i | s_i)P(s_i | s_{i-1})$$

- 定式化

$$(\hat{W}, \hat{S}) = \arg \max_{W \subset L, S \subset W} \prod_{i=1}^n P(w_i | s_i)P(s_i | s_{i-1})$$

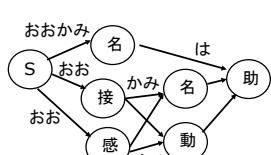
- 入力文字列に対して

- 辞書引きしたあらゆる組み合わせのパスを求める
- その中で確率値が最大のものを選択する



## 練習10

- 1重のHMMで形態素解析の結果を求めよ  
入力: 「おかげみは」

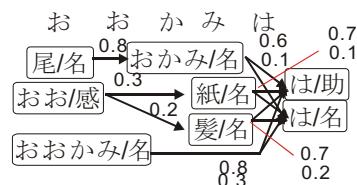


状態遷移確率: 単語の確率:  
S->名: 0.3 名-おかげ: 0.1  
S->接: 0.2 接-かみ: 0.2  
S->感: 0.1 感-お: 0.3  
名->助: 0.5 助-は: 0.5  
接->名: 0.3 助-は: 0.5  
接->動: 0.1 動-かみ: 0.1  
感->名: 0.3 動-かみ: 0.1  
感->動: 0.1 動-助: 0.2

## 形態素解析計算の難しさ

- パスの計算

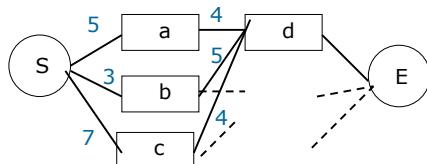
-(分岐の数)乗でパスの組み合わせが増える



## Viterbi アルゴリズム

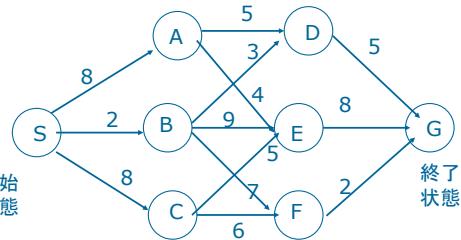
- 最適なパスを早く計算する方法
  - 前側から計算して最適な確率値のパスのみ残す
  - 最後まで計算してから最適パスを記録する

最適なパスを計算してみよう（コスト）



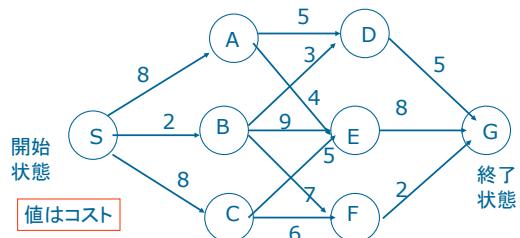
## 練習 11

- 最適なパスを求めよ



## N-best 解の探索

- 2番目、3番目に最適なパスを求めよ



## 探索問題

- 方法のヒント

- A\*探索を利用して2,3,4番目が求まる  
 $f(n)=g(n)+h(n)$   
 あるノードnまでのコスト  $g(n)$   
 ノードnから終点までの予測コスト  $h(n)$   
 (ただし実際のコストを超えないこと)  
 $f(n)$ はnを通ったときの全体の予測コスト

- Viterbi アルゴリズムをA探索に利用

## 確認問題

- 1番目、2番目、3番目に最適なパスを求めよ

