

9. 音声の認識：探索アルゴリズム

9.1 音響モデルと言語モデルのギャップを埋める

9.2 状態空間の探索

9.3 木構造化辞書で無駄を省く

9.4 ビームサーチで絞り込む

9.5 マルチパス探索で精度を上げる

9.1 音響モデルと言語モデルのギャップを埋める

- 音響モデルと言語モデルのギャップ
 - 音響モデル：音素単位（トライフォン）
 - 言語モデル：単語単位（文法 or N-グラム）
- 発音辞書
 - 音素列 V と単語 W との対応
 - ある表記に対して複数の読みがある場合は、確率を $P(V|W)$ と表現
- 事後確率の計算式の見直し

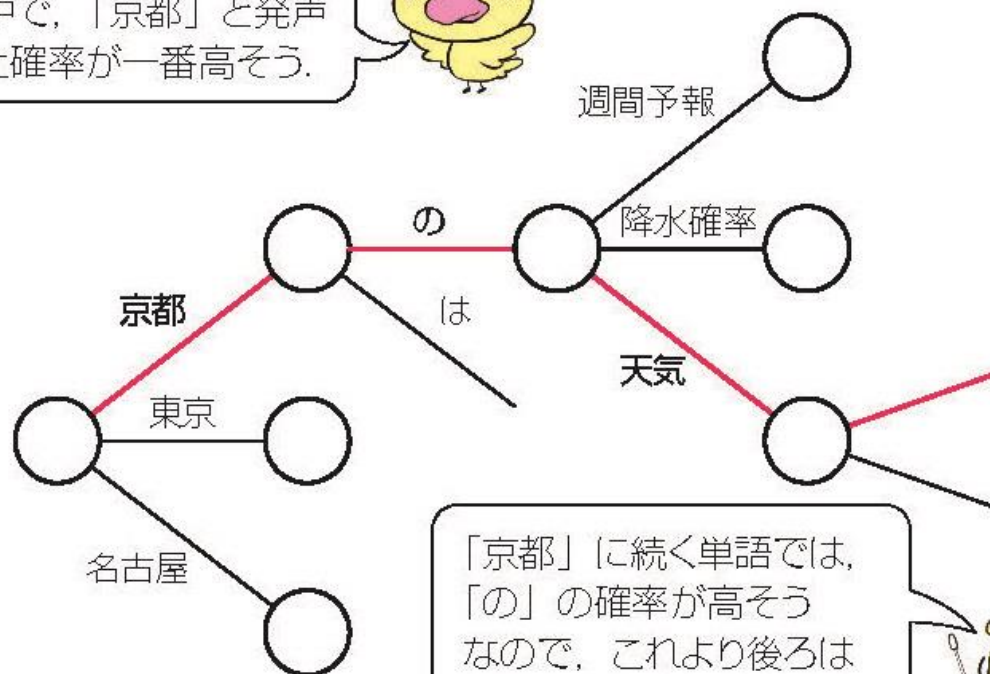
$$\hat{W} = \arg \max_W p(X|V)P(V|W)P(W)$$

9.2 状態空間の探索

- 探索の必要性
 - 大語彙連続音声認識において、すべての W について、 $p(X|V)P(V|W)P(W)$ の値を求めるのは不可能
 - 音声の先頭から、スコアの高い候補に絞って、接続可能な展開をおこなう探索処理が必要
- 探索の効率化・高精度化の工夫
 - 木構造化辞書
 - ビームサーチ
 - マルチパス探索

9.2 状態空間の探索

文法に従うと、最初の単語は地名です。
その中で、「京都」と発声された確率が一番高そう。



「京都」に続く単語では、「の」の確率が高そうなので、これより後ろは「京都 の」に続く単語に絞って探索します。



図 9.2 音声認識における探索のイメージ

9.3 木構造化辞書で無駄を省く

- 候補の展開数の削減
 - 単語単位で展開すると、候補は単語数個広がる
 - 単語間で先頭から共通する音素をまとめ、木構造にすると、展開数は音素数個

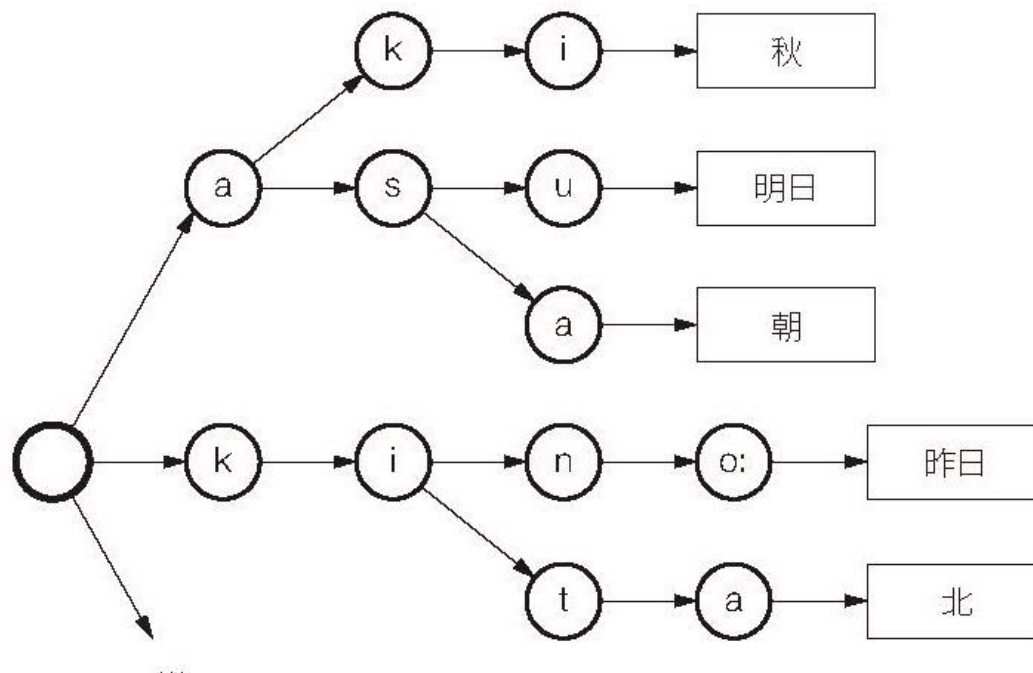


図 9.3

木構造化辞書

9.4 ビームサーチで絞り込む

- 探索の実時間化

- フレーム同期探索方式

- HMMのトレリス空間において、1フレームごとにスコアを計算

- ビームサーチ

- 1フレーム単位で、残す探索候補の数を一定幅に制限

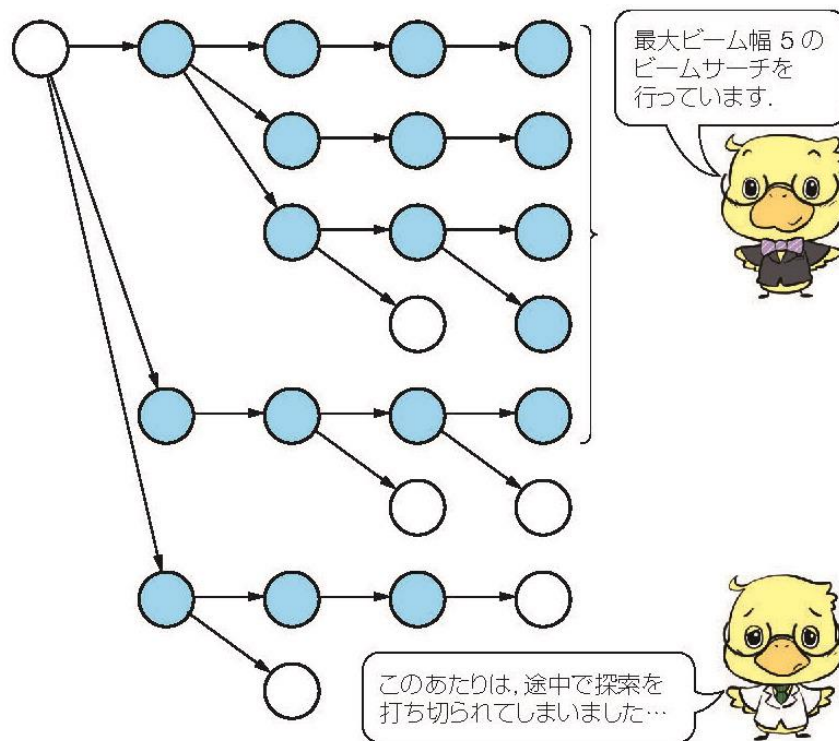


図 9.5 ビームサーチ

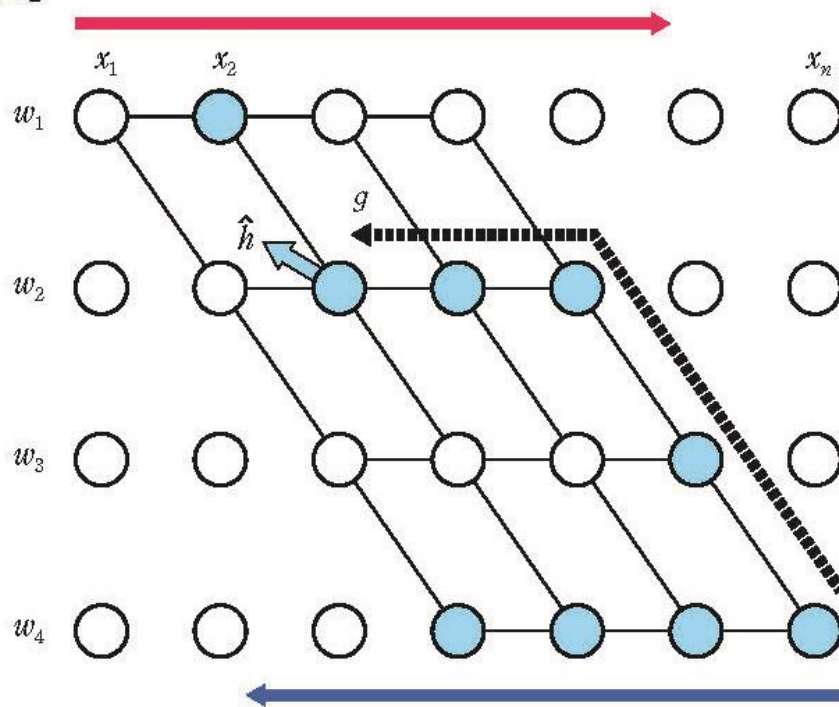
9.5 マルチパス探索で精度を上げる

- 探索の目的
 - 事後確率最大となる単語列（最適解）を求めたい
- （単純な）ビームサーチの問題点
 - 全体としては最適となる解が、途中のスコアの低さでビーム幅から外れてしまう可能性がある
- マルチパス探索によって最適化を求める
 - 未探索部分のスコアをヒューリスティックスにより見積もって、最適解をビーム幅の中にとどめる
 - ヒューリスティックスは、探索に先立って高速な処理によって求める

9.5 マルチパス探索で精度を上げる



1 回目の探索は前から行い,粗いモデルで計算した値を残します.



2 回目の探索は,逆向きに,
 $f = \hat{h} + g$ の値の高いものから展開します.



図 9.6 Julius の A* 探索