アルゴリズムとデータ構造III 5回目:10月30日

構文解析 (トップダウン)チャート法

授業資料 http://ir.cs.yamanashi.ac.jp/~ysuzuki/algorithm3/index.html

1

→ 授業の予定(中間試験まで)

1	10/02	スタック(後置記法で書かれた式の計算)
2	10/09	チューリング機械、文脈自由文法
3	10/16	構文解析 CYK法
		構文解析 CYK法
5		構文解析(チャート法), グラフ(ダイクストラ法)
6		グラフ(ダイクストラ法, DPマッチング)
7	11/13	グラフ(DPマッチング, ビームサーチ)
8	11/20	グラフ(ビームサーチ, A*アルゴリズム)
9	11/27	中間試験

_

授業の予定(中間試験以降)

[文本》]				
10	12/04	全文検索アルゴリズム (simple search, KMP)		
11	12/11	全文検索アルゴリズム (BM, Aho-Corasick)		
12	12/18	全文検索アルゴリズム (Aho-Corasick), データ 圧縮		
13	01/08	暗号(黄金虫, 踊る人形)		
		符号化(モールス信号, Zipfの法則, ハフマン符号)テキスト圧縮		
14	01/15	テキスト圧縮 (zip), 音声圧縮 (ADPCM, MP3, CELP), 画像圧縮(JPEG)		
15	01/29	期末試験		



本日のメニュー

- ■構文解析
- (トップダウン)チャート法
- 動的計画法
 - ダイクストラ法



チャート法(構文解析)

- トップダウンチャート法
 - Sから出発
 - 目的の単語列を導出 → 解析終了
- ボトムアップチャート法
 - 単語列から出発
 - Sを導出 → 解析終了

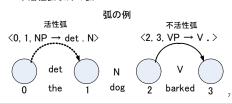
チャート法で使用する用語 1/3

- 節点(ノード)
 - 単語と単語の間に存在する仮想的な点
- 弧(アーク)
 - 節点間を結び、文の部分的な構造を表す
 - $\langle i,j,C \rightarrow \alpha.\beta \rangle$
 - iは弧の始点, jは弧の終点
 - ■.は解析が終了している位置
 - 節点iからjまで解析するとa
 - βまで解析できるとC

6

チャート法で使用する用語 2/3

- 不活性弧
 - 右辺の最後に•がある弧
- 活性弧
 - 不活性弧以外の弧

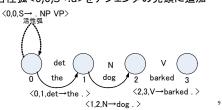


チャート法で使用する用語 3/3

- チャート
 - ノード、弧の集合
- アジェンダ
 - チャートに追加するべき弧のリスト

トップダウンチャート法のアルゴリズム(1/2)

- 辞書規則の適用
 - 入力文の各単語W_kについて、不活性弧<k,k+1, A→w,..>をアジェンダに追加
- 1 活性弧<0,0,S→.a>をアジェンダの先頭に追加



トップダウンチャート法のアルゴリズム(2/2)

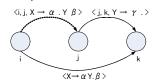
- アジェンダが空になるまで以下の操作を繰り返す
 - 弧の選択
 - アジェンダから弧を1個選びチャートに追加(選んだ弧=arc)
 - 弧の結合

 - arcが活性弧<i,j,X→α.Yβ>のとき,
 arcの右にある不活性弧<j,k,Y→γ.>を探し、結合する(次ページ)
 - arcが不活性弧<i,j,Y→γ.>のとき,
 arcの左にある活性弧<k,i,X→α.Yβ>を探し, 結合する
 - 結合してできた新しい弧<i,k,X→αY.β>をアジェンダに追加
 - 新しい弧の提案

 - arcが活性弧</jj,X→a.Yβ>のとき Yを左辺とする規則Y→y(辞書規則を除く)があれば、新しい活性弧 <j,j,Y→y>を作ってアジェンダに追加

トップダウンチャート法のアルゴリズム

- 弧の結合
 - 例えば
 - <i, j, X \rightarrow a. Y β > + <j, k, Y \rightarrow γ .>
 - \rightarrow <i, k, X \rightarrow aY. β >



■ 不活性弧<0,n,S→a.>が生成できれば解析成功

(トップダウン)チャート法を用い



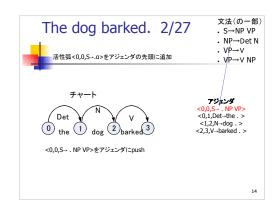
- た構文解析例 (例文)
- The dog barked.
- 文法

■ 解析文

- S→NP VP
- NP→Det N
- VP→V
- VP→V NP
- Det → the
- N → dog V → barked

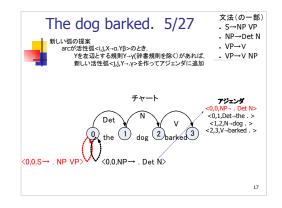
12

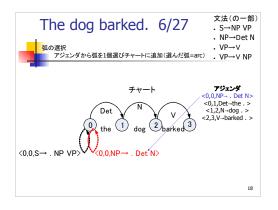


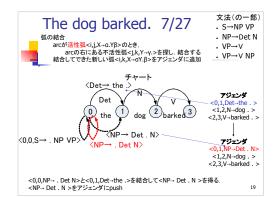


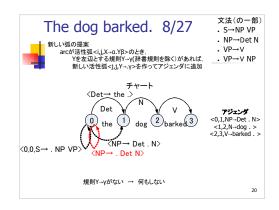


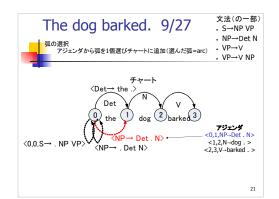


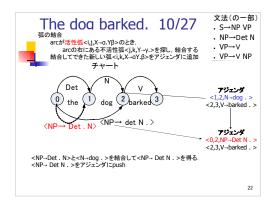








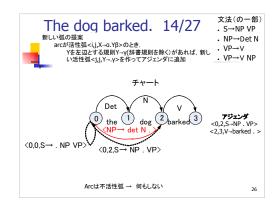






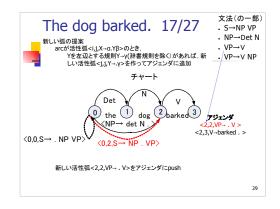


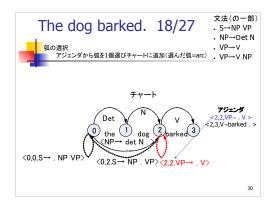


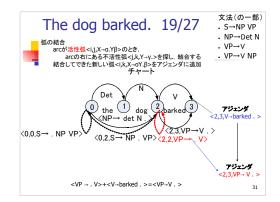






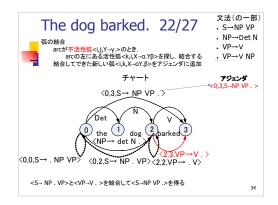


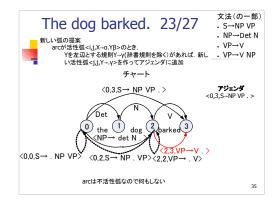


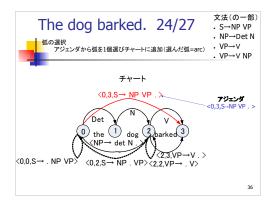


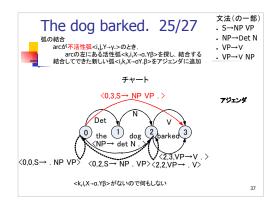


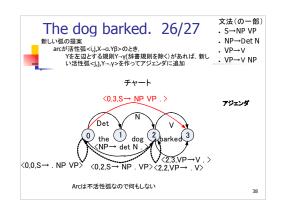




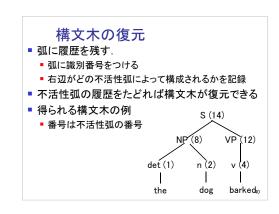












チャート法の特徴

- 計算量は O(n³)
- 任意の文脈自由文法が扱える
- A→BCDも, A→bCもOK
- 4種類の方式
 - トップダウンとボトムアップ
 - 縦型探索と横型探索
- 文法の予測能力が使える
- 無駄な弧を生成しないので効率が良い(トップダウンチャート法)
- 広く使われている

41

縦型探索と横型探索

- 縦型探索
 - 1つの解の候補の解析を優先的に進める
- 文が文法によって生成できるかだけを調べるときに便利
- 横型探索
 - 全ての解の候補の解析を並列に進める
 - * 至しいがいた…・ ビームサーチが使える
- チャート法では両方とも可能
- アジェンダをスタック(LIFO)にしたときは縦型探索
- アジェンダをキュー(FIFO)にしたときは横型探索

42

文法の予測能力 ■ 無駄な弧は生成されない ■ 文法によってDetの後にはVが現れないことが 予想されている 文法 . S→NP VP <NP→Det . N> NP→det N VP→V VP→V NP <∀P.→∀.→ '→dog VP→V INT Det → the N → dog <N→dog . Det dog . V → dog 0 <NP→. Det N>



- ■動的計画法(Dynamic Programming)
 - ■解くのに時間のかかる問題を、複数の部分問題に 分割することで効率的に解くアルゴリズム
 - 個
 - ダイクストラ法
 - DPマッチング
 - CYK法

44



ダイクストラ法

<S→ . NP VP>

- 動的計画法を最短経路問題に適用
- 最適経路中の部分経路もまた最適経路になっている



身近な最短経路問題

■ 道路の経路探索(カーナビなど)

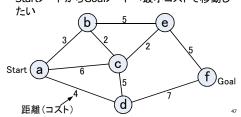
46

45

43

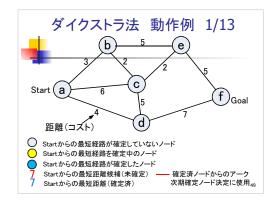
ダイクストラ法(最短経路問題用 アルゴリズム)

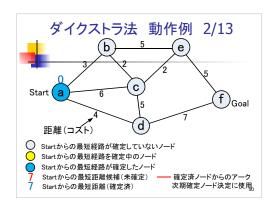
StartノードからGoalノードへ最小コストで移動し

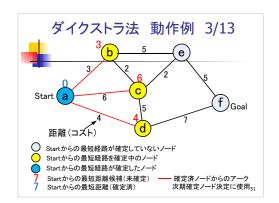


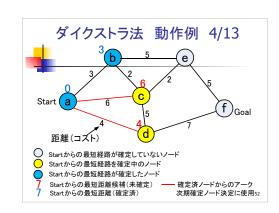


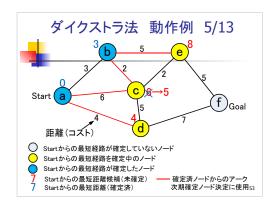
距離(コスト)

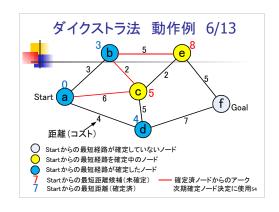


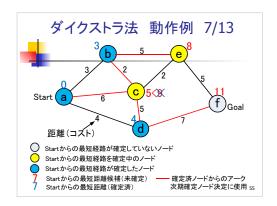


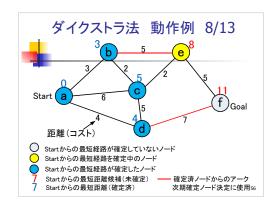


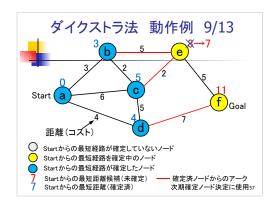


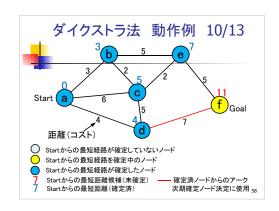


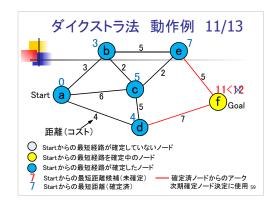


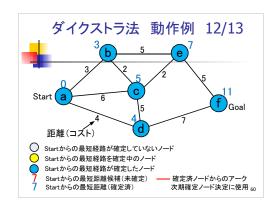


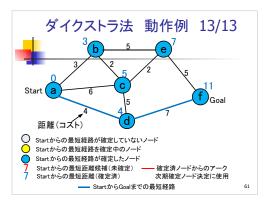












ダイクストラ法 アルゴリズム



- 初期化:スタートノードの値(最小コスト候補)を0,他の ノードの値を無限大に設定
- 未確定ノードが無くなるまで以下のループを繰り返す。
- 確定中ノードのうち、最小の値を持つノードを見つけ、確定ノ ードとする.
- 確定ノードからのエッジに対して「確定ノードまでのコスト+エッジのコスト」を計算し、そのノードの現在値よりも小さけれ

62



ダイクストラ法の特徴

- 最短経路の見つけ方
 - ゴールノードから「どこから来たのか」調べ、さかの
- マイナスのコストを持つエッジは扱えない.
- 特定のノードからの最短距離およびその経路 が全てのノードに対して求まる.

63

DPマッチング



(例:文字列の照合)

- 2つの文字列がどのくらい似ているかを調べる.
 - Yamanashi は kamonohashiとtakahashi
- 音声認識にも使える
 - 音声を文字列に変換した後、登録単語と比較
 - (現在主流の)HMM(Hidden Markov Model)に拡張
- DNAの比較にも使える
 - A(アデニン), G(グアニン), C(シトシン), T(チミン) の並び方の比較
 - ACTGAGCATTとCTGGACTACGの比較