自然言語処理2014

-平成26年11月17日(No7)-

東京工科大学
コンピュータサイエンス学部
亀田弘之

復習（頭のリフレッシュ）

- 前回のパワポを読み返してみよう。
  （前回の残りも一緒に話をします。）

今日の内容

1. 前回の復習
2. 構文解析プログラムの動作を再度理解する。
   （自分の言葉で誰にでも説明できることを目指す。）
3. 日本語処理のための技術的注意点
   - 文字コード
   - 形態素解析
4. ここまでのアプローチの限界
   （ここが大切。 一緒に議論しましょう！）

学習目標

1. 構文解析の動きを説明できる.
2. ユニフィケーションに十分なれる.
3. 技術の良し悪しを判断できる.

構文解析プログラムの動作を再度理解する

[Step 1] 分析対象英文の決定

- 処理対象:

John ran fast.

(参考)
- 名詞(noun), 動詞(verb), 副詞(adverb)
- 動詞句(verbal phrase)
- 文(sentence)

[Step 2] 文法の設定

例

文法G=<Vn, Vt, σ, P>
ここで、

- **非終端記号**の集合:
  Vt={s, n, vp, adv }
- **終端記号**の集合:
  Vn={ john, ran, fast }
- **開始記号**σ=s
- **書き換え規則**群P:
  右記の通り。

{s → n,vp.
vp → v,adv.

統語規則

n → john.
v → ran.
adv → fast.

単語辞書

}

[Step 3] Prolog形式への変換

練習

[形式文法の書式]

s → n,vp.
vp → v,adv.

n → john.
v → ran.
adv → fast.

[Prologの書式]
（書き換えてみよう）

[Step 3] Prolog形式への変換

解答例

[形式文法の書式]

s → n,vp.
vp → v,adv.

n → john.
v → ran.
adv → fast.

[Prologの書式]

```
s(A,C,s(_n,_vp)) :-
  n(A,B,_n),vp(B,C,_vp).
vp(A,C,vp(_v,_adv)):-
  v(A,B,_v),adv(B,C,_adv).

n([john|T],T,n(john)).
v([ran|T],T,v(ran)).
adv([fast|T],T,adv(fast)).
```

動作させてみよう

1. Prologを起動する
2. 構文解析のプログラムを読み込む
3. プログラムを実行する
   （自分でできますか？）

(注) Prologのプログラムの実行方法はいろいろあります.
自分のやり方で結構です.

構文解析の動作を理解しよう

- 構文解析プログラム＝Porlogプログラム
- Prologを理解する必要がある。
- Prologを理解するポイントは、ユニフィケーション(unification)という概念の理解にある。

- 今日の学習目標の1つがこれ。

（以下、板書で詳細に説明します。）

1. ユニフィケーションとは (What is unification?)
2. 実例による理解 (Can you figure out how Prolog works?)
3. 動作のトレース (let's illustrate a trace of processing as a picture!)

頑張って理解してください.
人工知能の勉強にもなります.

（確認）構文解析プログラムの作成手順

1. 開発環境の整備
   - ① 各種ソフトウェアのインストール
   - ② 文字コードの決定　など
2. プログラム作成手順
   - ① 言語データの収集
   - ② 言語分析　<= 対象言語の知識が必要
   - ③ 形式文法の設定　<= 形式言語の知識が必要
   - ④ Prolog形式への書き換え　<= Prologの知識が必要
   - ⑤ NLPプログラムの実行
   - ⑥ 構文木を出力するプログラム (draw_term/1)
   - ⑦ 動作することの確認
   - ⑧ 妥当性の検証　など
   - ⑨ 公開へ向けてのドキュメント等の整備

限界は？

みんなで議論しましょう！（10分）

解決方法は？

参考情報

自然言語処理関係の学会・国際会議

- ACL (Association for Computational Linguistics)
- COLING
- PACLING
- 言語処理学会
- 電子情報通信学会
  - 思考と言語研究会
  - 言語理解とコミュニケーション研究会
- 情報処理学会
  - 自然言語処理研究会
- 計量言語学会
- その他