

言語解析論

講師 竹内孔一

本日の内容

- 言語の意味処理について
 - 言語の構造(言語学からの知見)
 - アプリとして質問応答・含意認識

2

言語処理で行いたいこと(1/2)

- 大量の文書から
 - 必要な文書の取り出し (検索)
 - 必要な情報の取り出し(質問応答)
 - 「(ある範囲で)オープンキャンパスの開催日は?」
 - 「カレンダーで直近のスケジュールを出すには?」
 - なにが起きているか? (テキストマイニング)
 - あるプリンタには同様の苦情がいくつかある
- 会話から
 - ロボットによる会話の理解
 - 相手の命令の理解
 - 積み木の操作を言語で行う(SHRDLU)
 - 相手の気持ちの理解

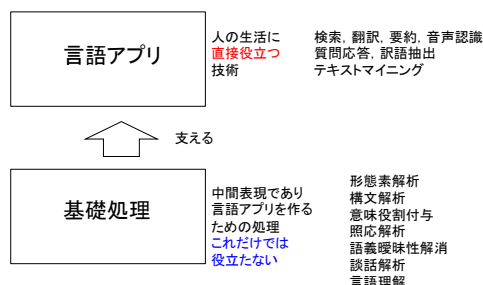
3

言語処理で行いたいこと(2/2)

- 言葉の運用に関する補助
 - 対訳を探す (翻訳支援)
 - swine flu の日本語訳は? フランス語は?
- 文書要約
- 機械翻訳などいろいろ. .

4

言語応用(アプリ)と基礎の違い



やりたいことに対する必要な技術

- 文・文書の構造に対するモデル化
 - 文の構造をよく調べて, コンピュータで処理しやすい形が何かを明らかにする
 - 同様の意味のものはまとめたい
 - 「彼はその劇の主役だ」「その劇は, 彼が主役だ」
- 意味に対するモデル化
 - 知識をどう補完するか?
 - 言語的知識 (WordNet) と 分野依存知識
 - 分野依存知識が無いとわからない
 - (例) 日本語でも自分の知らない専門文書を読んだとき理解できない => 知識を補って理解している

6

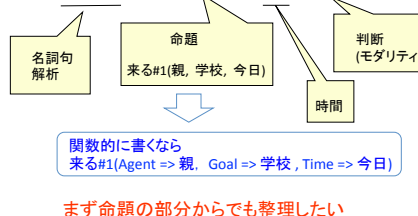
アプローチ

- 文の構造に対するモデル化
 - SRL (semantic role label) 意味役割解析
 - 語義曖昧性解消
 - モダリティ解析
 - 名詞述語文, 名詞句解析...
- 文書の構造に対するモデル化
 - 照応解析, 省略解析
 - RST (rhetorical structure theory) 修辞構造理論
- 意味に対するモデル化
 - 言語的知識
 - 語同士の類似関係辞書=>シソーラス (WordNet)
 - 分野依存知識
 - 現状モデル化はよくわかっていないがWikiは有用
 - LinkedOpenData など (知識処理, 形式意味論の分野)

文の構造化

命題部分 と モダリティ

(例) 彼の親は今日学校に来る だろう



文書の構造化

照応・省略・RST

(例)

「太郎は人工知能学会に入会したい

(a)

参加費は2千円らしい

(b)

高いので

(c)

今は止めておく」

(d)

照応解析

(b) 「何の」参加費 は 人工知能学会の参加費
RST(修辞構造解析)

(a) は(b)に対する目的

(b) は(c)に対する原因

(c) は(d)に対する原因

意味に対するモデル化(言語的知識)

• 類義語をまとめる

- 言語資源として構築されて提供 (ある種の辞書)
- (語・フレーズ間の意味的關係)

言語資源

- FrameNet (英語他) 日本語FN 慶応大 小原先生
 - 意味概念をベースに動詞, 名詞, 副詞, 形容詞を関係づけ
- WordNet (英語他) 岡山大
 - 類語関係を記述 動詞項構造シソーラス
- 日本語語彙体系 (日本語) (NTT)
 - 日本語の名詞, 動詞に関する意味関係を木構造で記述
- EDR (日本語) (商用)
 - 日本語の名詞, 動詞, 形容詞, 副詞に関する巨大な概念辞書, 英語の翻訳辞書, 意味概念付与テキストデータ

国語辞典や翻訳辞書も言語資源

アプローチ

- 命題部分の解析
 - 意味役割解析
 - 語義曖昧性解消

述語の項構造

• 項構造とは何か?

- 述語の取る概念的なタイプ

(例)「雇う」

- 誰が(動作主体), 誰を(対象), なんの職で(職務?),
- どこに(Location), いつ(Time), 期間(Time-Span)

「社長が太郎を雇った」

「社長が会計士を雇った」

「社長が太郎を会計士として雇った」

雇う(Agent => 社長, Theme => 太郎, Role => 会計士)

=> コンピュータで扱いやすい形式に近づいた

意味役割のタイプ

- 様々な提案
 - 格文法(Fillmore)
 - 動作主格, 対象格, 目的格, 時間格, 場所格, 道具格. .
 - VerbNet とPropBank (Palmer)
 - arg0, 1, 2 と番号 + 付加詞 (23種類)
 - 語義によって意味役割は違うというポリシー
 [Agent Nora] brought [Theme the book]
 - FrameNet (Fillmore)
 - 1000種類以上の意味役割(更新され続けている)
 - [Item Colgate's stock] rose [Difference \$3.64] [Final_value to \$49.94]
 [Sleeper They] [Copula were] asleep [Duration for hours]

例題

- 意味役割のタイプを考えてみよう
 (教科書p72の表5.2から)
 「彼は鍵でドアを開けた」
 「彼は足でドアを開けた」

語義の曖昧性(1/2)

- 前項の意味役割は語義に依存して異なる
 - 語義ごとに異なる表現が可能(類義語)
 - (例)「握る」
 - おにぎりを握る/作る 握る#1
 - 手すりを握る/つかむ 握る#2
 - 秘密を握る/知る/つかむ 握る#3
 - . . .
- 「握る」「つかむ」などが類義語集合 => WordNetや動詞項構造シソーラスなどでまとめられている

語義の曖昧性(2/2)

- 各語義での異なる意味役割
 - (例)「走る」
 - 「日本刀がさやから走る」
 - 走る(Theme => 日本刀, Source => さや)
 - 「選手がコースを東に走る」
 - 走る(Agent => 選手, Path => コース, Direction => 東)
 - 「悪事に走る」
 - 走る(Theme (動作) => 悪事)

例題

- 下記の語義は「生徒が校庭を走る」の意味と近い異なるか? どう違うか答えよ
- 「鉄道が南北に走る」

質問応答や含意認識における
 基礎技術
 (応用処理を意識して今までの技術
 を見直してみよう)

必要とする要素技術

(1) 語, 句, 文単位での類似表現(言い換え)

- 「造語」⇒「作った言葉」

(2) 特定の概念 (e.g., 「理由」「方法」)

- 車の購入理由の調査ならば
 - ・「ooが理由でX車を買いました」

(3) 分野依存での知識(推論)

- 購入の調査ならば
 - ・「この車に決めました」⇒「この車を購入した」
- 人の異動の調査
 - ・「Sam is succeeded by Ken」
KenがSamのあとを引き継いだ → Kenが就任

言語資源
(人手の
辞書)
または
自動獲得

19

練習問題

- 下記の2つの文は含意関係にあるが正確には推論が必要となる。どこか？

<t1>伊坂幸太郎は直木賞候補になった2003年の『重力ビエロ』で一般読者に広く認知されるようになった。</t1>
<t2>『重力ビエロ』は伊坂幸太郎による小説で直木賞候補作品だった。</t2>

NTCIR-10 RITE2 タスク 含意認識開発データより

(1) 語, 句, 文単位での類似表現

• 語の類似

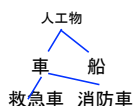
- シソーラス(階層構造)
- 類語辞書

• 述語の項構造での類似

- 組み合わせでの類似
- 動詞: 「彼を雇う」⇔「彼を雇用する」
- 名詞: 「彼はこの本の著者」⇔「彼はこの本を書いた」

• 多義性の解消

- 上記の「語」「項構造」それぞれで単語は同じでも意味が違う⇒どちらの意味か識別が必要
- 「学校まで走る」(移動) vs 「稲妻が走る」(知覚)



(2) 特定の概念 (「理由」「方法」)

• 意味役割

- 述語と係り関係にある語(「項」と呼ぶ)との関係
- 言語学では項構造(argument structure)と呼ぶ

[彼は]動作主 [このWebサイトの評価で]理由
[この車を]対象 購入しました

- 「動作主」や「対象」などが意味役割でいくつあるかはいろいろな議論がある

- 言語学から網羅的なのは「現代日語文法2 第3部格と構文 第4部ヴォイス(くろしお出版)」

(3) 分野依存での知識(推論)

• 手法は確立してない

- 統計的にある分野で共起する表現や文の構造を手掛かりにする

事例

「[彼は]動作主 [この作品で]手段 [有名に]着点(状態) なった」

- 上記の文から「彼」が「この作品」の著者であることがわかる

明示的で無い項構造から推定する研究の一例
A. Stern, I. Dagan: Recognizing Implied Predicate-Argument Relationships in Textual Inference, ACL 2014.

意味役割の体系は<http://pth.cl.cs.okayama-u.ac.jp>

具体的な処理を例に

参考文献

磯崎秀樹, 東中竜一郎, 永田昌明, 加藤恒昭, (監修: 奥村 学):
質問応答システム コロナ社 (2009).
金山博・武田浩一: Watson: クイズ番組に挑戦する質問応答システム
情報処理 Vol.52 No.7 (2011).

質問応答・含意認識

- 質問応答
 - Factoid 型と non factoid 型
 - 前者: ある物の属性(長さや名前など)
 - 後者: 手段や原因など
 - 例)「瀬戸大橋は全長いくら?」
- 含意認識
 - 質問応答などの基礎技術
 - 「『重力ピエロ』の作者は誰?」
- 実処理システム
 - 人間のクイズ番組で勝つ IBM Watson
 - 言語資源, 構文解析, 照応解析, 統計的学習モデル

質問応答の例

(質問) 明石大橋は全長で何メートル?
 (知識源) 明石海峡大橋は全長3911メートルである

項構造解析: 「[明石海峡大橋] 主体 [全長] 対象
 [3911メートル] 補語相当 である

構造化: [明石大橋
 Form: (全長 3911メートル)]

質問文の構造化: [明石海峡大橋
 Form: (全長 X)]

解候補の生成 (「明石海峡大橋」⇔「明石大橋」)
 解の正しらしさの評価 X=3911メートル?

発展練習

- 下記の知識源から質問に答える手法を考えてみよう
- (1) (知識源)「夏目漱石は「坊ちゃん」の著者である」
 (質問) 小説「坊ちゃん」を書いた人は誰
- (2) (知識源) 大学のオープンキャンパスは
 10:30～16:00まで開催しています
 (質問) 何時に行けば良い?

まとめ

- 言語の意味処理(質問応答を例に)
 - 類義語の情報が必要
 - 項構造による組み合わせでの類義情報が必要
 - 多義性解消が必要
 - 分野依存の推論的知識が必要
 - 解候補の選択には統計的学習モデルが有効