

談話関係認識への連想情報の応用

杉浦 純 井之上 直也 乾 健太郎 (東北大学)

談話関係認識

談話関係とは

- 談話片と談話片の間にある**意味的關係**
- 因果関係, 対比関係, 例示関係, etc.

談話関係

因果

S1: The projects already under construction will **increase** Las Vegas's supply of **hotel rooms** by 11,795, or nearly 20%, to 75,500.
 S2: By a rule of thumb of 1.5 new jobs for each new hotel room, Clark County will **have** nearly 18,000 **new jobs**. (WSJ_0994)

連想情報

部屋が増える → 従業員が必要 → 新しい雇用生まれる

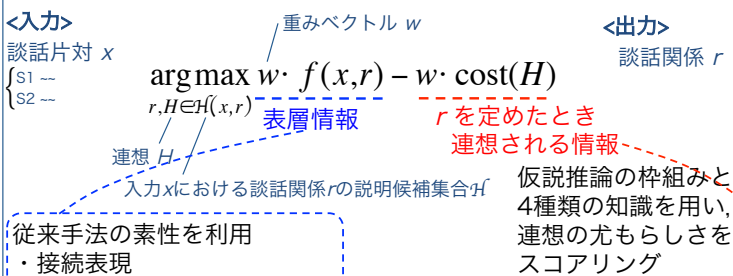
従来手法の問題点

- 接続表現を含まない談話関係 (Implicitな談話関係) の認識精度が低い (F値 およそ0.4 : [Lin+ 10])
- 主に談話片間に存在する単語対や単語品詞情報, 文の統語構造, 接続表現 (butやbecauseなど) といった**表層の情報のみを手がかり**として解析を試みている
- **外部知識が使えていない**

外部知識を活用し, 談話から連想される情報を手がかりに談話関係を認識するモデルの提案

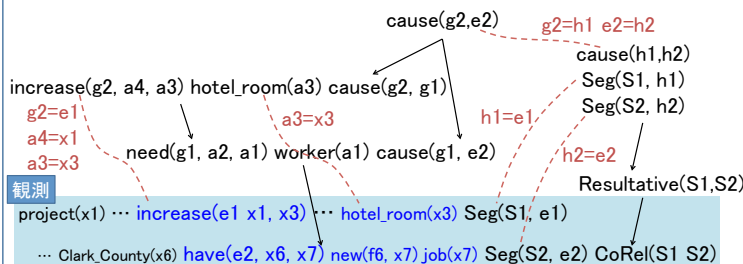
連想情報を手がかりとした談話関係認識モデル

表層情報と連想情報を同時に考慮できる線形モデル



従来手法の素性を利用

- ・ 接続表現
- ・ 単語品詞情報
- ・ 文の統語構造
- ・ 談話片間に存在する単語対 ← 本研究で実装



仮説推論

観測 O を含意する最良の仮説 (説明) H^* を, 背景知識 B を用いて求める論理的推論

- ただし, 仮説候補 H は
 - $H \cup B \models O$ (観測を含意する仮説)
 - $H \cup B \not\models \perp$ (無矛盾な仮説)
- 情報の想起が仮説生成の枠組みとして自然に定式化できる
- Interpretation as Abduction [Hobbs+ 93]
- **最小コストの説明を求める**

談話関係成立条件

$\text{cause}(f1, f2) \wedge \text{fact}(f1) \wedge \text{valNeg}(f2) \wedge \text{Seg}(f1, s1) \wedge \text{Seg}(f2, s2) \Rightarrow \text{Adversative}(s1, s2)$

関係間の関係

$\text{cause}(f1, f2) \wedge \text{cause}(f2, f3) \Rightarrow \text{cause}(f1, f3)$
 $\text{cause}(f1, f2) \wedge \text{presuppose}(f2, f3) \Rightarrow \text{cause}(f1, f3)$

共参照

$\text{hotel_room}(a1) \Rightarrow \text{room}(a1)$

事象間の関係

$\text{need}(f1, a, b) \wedge \text{worker}(b) \wedge \text{cause}(f1, f2) \Rightarrow \text{have}(f2, c, d) \wedge \text{new}(f, d) \wedge \text{job}(d)$
 $\text{sell}(f1, a, b) \wedge \text{cause}(f1, f2) \Rightarrow \text{give}(f2, a, b)$

[杉浦+ 12]を拡張

大規模語彙知識

WordNet3.0 [11.5万axioms]
 FrameNet1.5 [1.2万axioms]

実験

設定

- ベースライン: **表層的な情報のみ**を素性として学習
- データセット: Penn Discourse Treebank [Prasad+ 08] より約**500談話片対** (前処理として照応解析 (StanfordCoreNLP*1を用いた) の結果を反映)
- 重み w はマージン最大化オンライン学習アルゴリズム [Crammer+ 06]を用いてPDTBの500談話片対より学習
- 仮説推論エンジンには [Inoue and Inui 11] を利用

➢ 結果 ※1 <http://nlp.stanford.edu/software/corenlp.shtml>

手法	談話関係	Recall	Precision	F値	Accuracy
表層のみ	順接	54.0	50.4	52.1	48.8
	逆接	43.2	46.8	44.9	
+ 推論	順接	59.5	48.7	53.6	46.7
	逆接	33.1	43.3	37.5	

分析

- 談話片間のイベント間の関係がほぼ連想されていない
- より多様で大規模な知識の獲得が必要 (例えば因果関係)
- 推論時間を爆発させずに深い推論をする工夫が必要

まとめ

- 連想情報を考慮できる談話関係認識モデルを提案した
- 実在のデータを対象に, 表層的な情報のみを手がかりとして用いた場合と比較実験を行った
- 知識不足や推論の深さ不足といった課題が残った

今後の課題・展望

- 連想された情報・連想されなかった原因の分析
- 知識の拡充
 ConceptNet5*2, YAGO[Hoffart+ 11]など他の語彙資源の活用
- factualityやsentiment polarityの自動認識
 既存の解析器を利用できるように
- 複数談話関係の結合推論 (i.e. text全体)
 談話関係同士の相互作用を考慮

※2 <http://conceptnet5.media.mit.edu/>