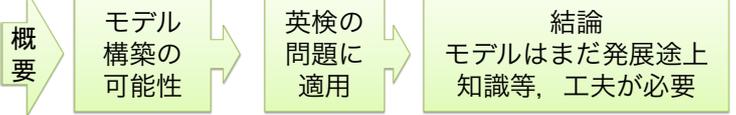


# 大規模語彙知識を用いた仮説推論による文章理解モデルの構築に向けて

杉浦 純 井之上 直也 乾 健太郎 (東北大学)

## 研究の目的

談話に明示的に書かれていない意図や因果関係などの文脈情報を統一的に推論して**談話を理解したい**



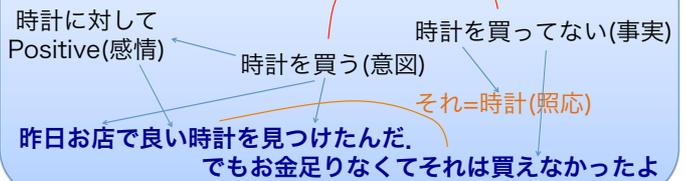
結論  
モデルはまだ発展途上知識等、工夫が必要

- 「談話理解」とは (80年代)  
仮説推論によって談話に対する**最良の説明**を求めること



- 重みつきアブダクション [Hobbs et al. '93]  
コストを用いて仮説を評価することにより最良の説明を求める枠組み  
最近RTEのタスクにも適用されている [Ovchinnikova+'11]

## 仮説推論による談話理解例



## 当時の問題点

- 頑健な推論に必要な**世界知識が足りなかった**  
→近年になって大量の語彙知識が**利用可能**に  
→ WordNet, FrameNet, 感情極性辞書など
- 現実的な推論時間で**実装できるモデルがなかった**  
→高速化された枠組みが**提案**されている  
→ Inoue and Inui '11

## 頑健な談話理解モデル構築の可能性

本当にできるの？

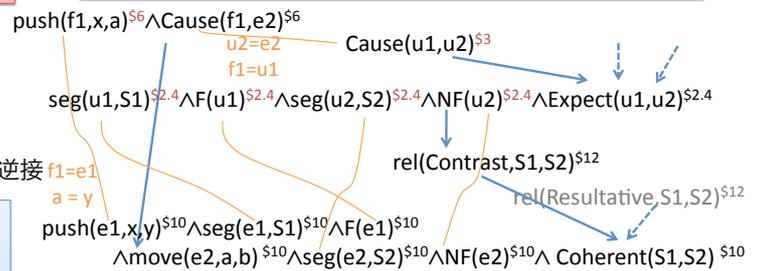
## 談話理解モデル (仮説推論) による談話関係認識

- 談話理解モデルを直接評価することは**難しい**  
モデルのひとつの成果として談話関係認識を用いる
- 手法: 大規模知識のほか、**談話関係認識に関するメタな知識**を人手で与えて説明を試みる
- 談話関係: **順接と逆接の2種類**
- 方針: 起こりそうなことが起こったら順接, 起きなかったら逆接

### メタな知識

Resultative(S1,S2) → Coherent(S1,S2) ← Contrast(S1,S2)  
 $\text{seg}(e1,S1) \wedge \text{seg}(e2,S2) \wedge F(e1) \wedge NF(e2) \wedge \text{Expect}(e1,e2) \rightarrow \text{Contrast}(S1,S2)$   
 $\text{Cause}(e1,e2) \rightarrow \text{Expect}(e1,e2)$   
 $\text{Happens-Before}(e1,e2) \rightarrow \text{Expect}(e1,e2)$   
 $\text{Inheritance}(e3,e1) \wedge \text{Inheritance}(e3,e2) \rightarrow \text{Expect}(e1,e2)$   
 $\text{Positive}(S1) \wedge \text{Positive}(S2) \rightarrow \text{Resultative}(S1,S2)$   
 $\text{Negative}(S1) \wedge \text{Negative}(S2) \rightarrow \text{Resultative}(S1,S2)$   
 F=Factual, NF=Non-Factual

知識:  $\text{push}(g1,x,y) \wedge \text{Cause}(g1,g2) \rightarrow \text{move}(g2,y,a)$



観測: [S1] I pushed the door. [S2] But, it didn't move.

## 実験

- 対象データセット  
英検 2級より1~3文程度の、談話関係を持つ文章30問
- 結果  
談話関係を認識できた説明は**少ない**
- 考察

推論エンジンの速さ不足

知識不足

改善の余地

実問題に大規模知識 → 現実的時間での推論の深さ  $d \leq 2$  程度  
 $d \leq 2$  では**知識があっても使いこなせない**  
 例:  $\text{Expect}(\text{study}, \text{understand})$  の関係は  $d=4$   
 動詞の知識 > 名詞・形容詞・副詞の知識  
 実問題は動詞だけでは解けない  
 談話関係とは? (メタ知識のパターン)

観測を述語2つのみに →  $d=4$  程度まで推論可  
 無駄な推論の除去,  $d$  を消費しないような知識を書き方

## 大規模語彙知識

- WordNet3.0 [Fellbaum'98] (約30万 axioms)
- FrameNet1.5 [Baker+'98] (約2万 axioms)
- Semantic Orientations of Words [Takamura+'05] (約9万 axioms)
- VerbOcean [Chklovski '04] (約6千 axioms)

## 自然文

It rained all day. [So.] we had to postpone our soccer game.

構文解析  
述語項構造解析

Enju [Tsuji Lab]

論理式へ

推論エンジン [Inoue & Inui '11]

談話理解モデルの今後  
深く推論を行えるよう工夫すれば説明が談話関係を認識する可能性がある

