

# 分散表現を用いた因果関係のモデル化

佐々木翔大 高瀬翔 井之上直也 岡崎直観 乾健太郎  
東北大学

目的: 大規模なコーパスから因果関係知識を獲得する(COPAで評価)

COPA [Roemmeleら, 11]: ある文に対し原因あるいは結果を選択するタスク

Steam rose off of the pot of water 原因 × The water boiled  
I put my hand to my chest 結果 ○ I felt my heartbeat  
× My heart rate sped up

## 従来手法

因果関係を表すマーカ(Causal cues)を用いてコーパスから因果関係にある単語集合を抽出、原因と結果の単語ペアをカウント [Zhiyiら, 16]

Causal cues (53種類)

A lead to B If A, then B  
A cause B B because A etc...

単語共起頻度カウント

The storm caused tremendous flooding  
(storm, tremendous), (storm, flooding)

$$CS(i,j) = \left( \frac{\frac{f(i,j)}{N}}{\left( \sum_{w \in W} \frac{f(i,w)}{M} \right)^\alpha * \sum_{w \in W} \frac{f(w,j)}{M}} \right)^\lambda * \left( \frac{\frac{f(i,j)}{N}}{\sum_{w \in W} \frac{f(i,w)}{M} * \left( \sum_{w \in W} \frac{f(w,j)}{M} \right)^\alpha} \right)^{1-\lambda}$$

## 問題点

単語間の類似性を扱えない 例: hurricaneとstormが似た事象であると認識することができない

## 提案手法

分散表現を用いることで、単語間の類似性を扱う(似た事象を扱えるようにする)  
原因の単語ベクトルから結果の単語ベクトルを予測できるよう学習

ベクトルの学習

The storm caused tremendous flooding  
 $v_a$   $v_{b1}$   $v_{b2}$

Objective:  $J = -\sum \log P(b|a)$

条件付き確率をnegative sampling [Mikolov, 13]で計算

$$\log P(b|a) = \log \sigma(v_a * v_b) + \sum E_{b' \sim P(b)} [\log \sigma(-v_a * v_{b'})]$$

## 実験(COPA)

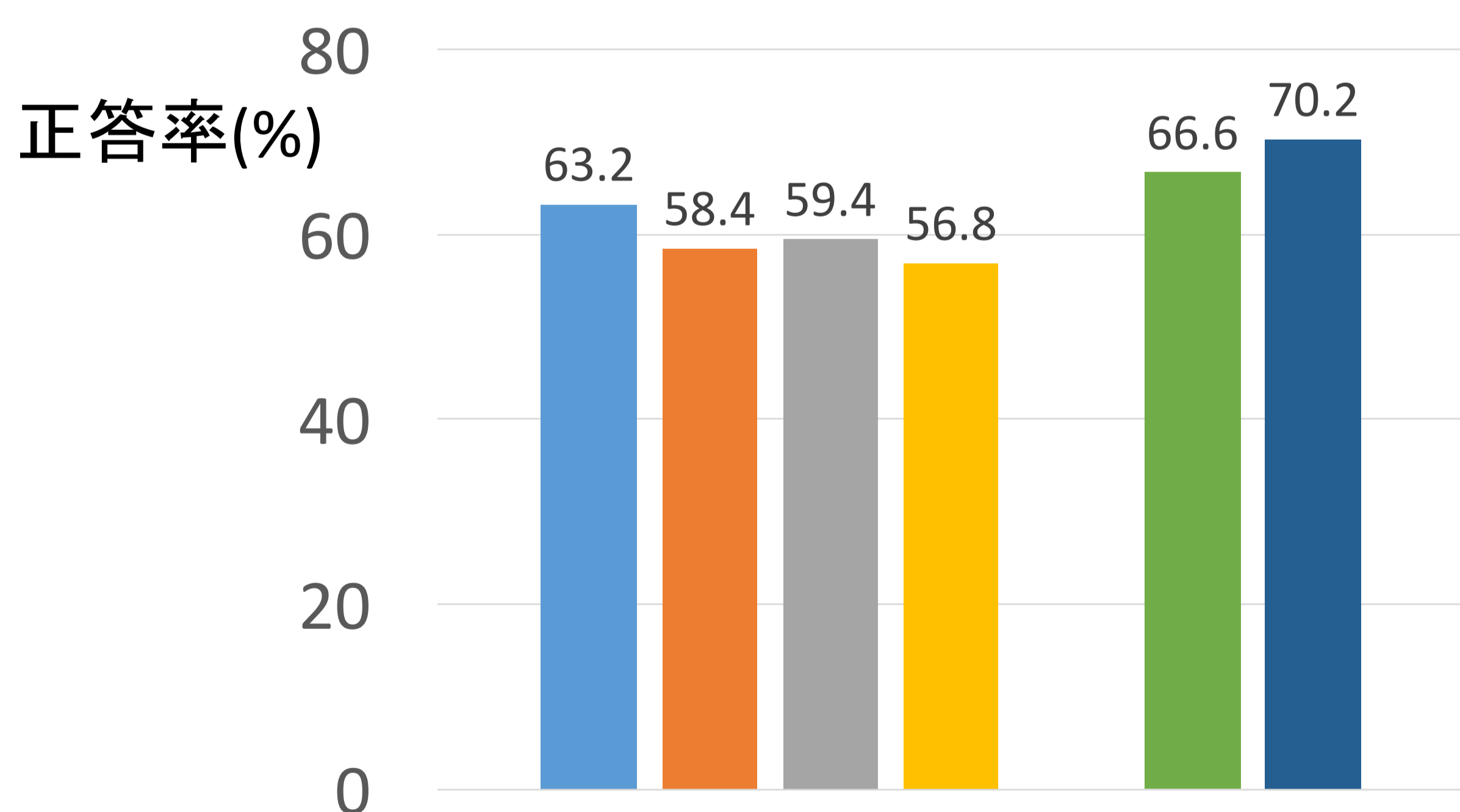
I put my hand to my chest 結果 I felt my heartbeat  
を選択 My heart rate sped up

従来手法: 各単語のCSが大きい候補を選択

$CS(\text{put}, \text{felt}) + CS(\text{put}, \text{heartbeat}) + \dots$  VS.  $CS(\text{put}, \text{heart}) + CS(\text{put}, \text{rate}) + \dots$

提案手法: 原因, 結果ベクトル間の内積が高い候補を選択

$\text{dot}(\text{put}, \text{felt}) + \text{dot}(\text{put}, \text{heartbeat}) + \dots$  VS.  $\text{dot}(\text{put}, \text{heart}) + \text{dot}(\text{put}, \text{rate}) + \dots$



- 従来手法(再実装)
- 提案手法 w/ 前処理(従来手法と条件が同じ)
- 提案手法 w/o 前処理
- Word2Vec w/o 前処理
- 従来手法(本発表とほぼ同じコーパス量 [Zhiyiら, 16])
- 従来手法 [Zhiyiら, 16]

## 定性分析

	ある原因に対する結果の内積の上位	word2vecでのコサイン類似度上位
	flooding(洪水)	tornado
storm	mudslide(地滑り)	hurricane/tropical
	outage(停電)	hurricane
	doubleheader	tourney(トーナメント)
rain + game	unplayable	thunder
	softball	raining
	evaporation(蒸発)	120 f°
water + heat	scald(やけどさせる)	140of
	evaporative	120 f°

原因(または結果)に対して結果(または原因)として妥当そうな単語との内積が高い  
=原因に対して結果を予測する分散表現が得られている

## まとめ

分散表現を用いて因果関係をモデル化する手法を提案した  
提案手法によって因果関係がモデル化できることを確認した

## 今後の課題

単語集合から事象へのエンコード手法を探求する  
3つ以上の事象間の因果関係(A→B→C)を扱えるようにする