

ナレッジグラフ推論チャレンジ 本部門応募フォーム

グラフ畳み込みネットワークを用いた推理小説の犯人推定

東京都市大学大学院 総合理工学研究科情報専攻 | 産業技術総合研究所:

勝島修平

東京都市大学大学院 総合理工学研究科情報専攻: 穴田一

産業技術総合研究所: 江上周作

産業技術総合研究所: 福田賢一郎

目次

- 問題の概要
- アプローチの特徴と結論
- 既存研究
- 提案手法
- 結果
- 課題
- まとめ

問題の概要

人工知能の発展に伴った説明性を持つAIの開発の必要性

ナレッジグラフ推論チャレンジの開催

小説の内容を構造化した大規模ナレッジグラフを利用し、
推理小説の犯人を説明付きで推定



アプローチの特徴と結論

特徴

- グラフ畳み込みネットワークによる小説の学習
- Layer-wise relevance propagationによるノード分析

結論

小説に必要な知識を追加した場合、犯人を推定可能

既存研究の問題点

黒川ら TransEによる埋め込み手法

場面	主語	述語	対象	場所	起点	終点	時間	何	理由
1	Helen	come				House of Holmes			
2	Helen	beScared							
...									
6	Helen	obtain					Within 2 months	money	

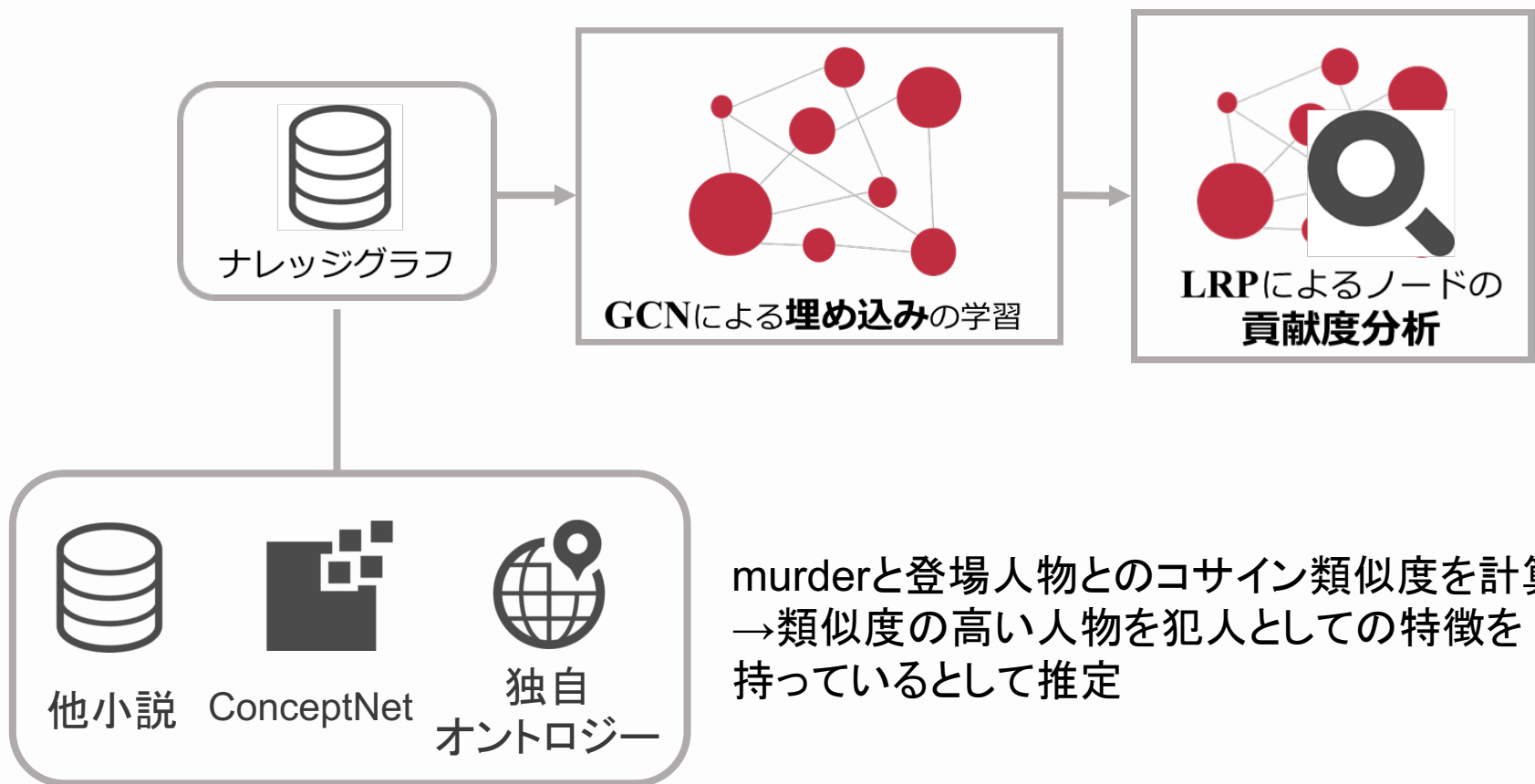
↓ SVO形式に分解

場面	主語	述語	目的語
1	Helen	come	house of Holmes
2	Helen	beScared	
...			
6	Helen	obtain	within 2 months
6	Helen	obtain	money

小説上の場面で起こる単語の同時性を考慮できてない

提案手法

提案手法のフロー



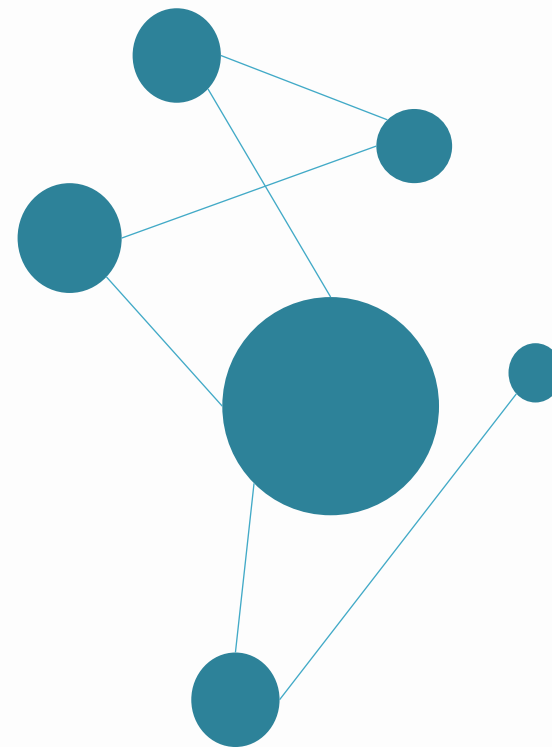
提案手法 graph convolutional network

概要

- ノード同士の隣接関係：隣接行列 \mathbf{A}
- ノードの特徴ベクトル：特徴行列 \mathbf{H}

$$\mathbf{H}^{(l+1)} = \sigma\left(\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{H}^{(l)}\mathbf{W}^{(l)}\right)$$

σ : 活性化関数 $\mathbf{W}^{(l)}$: 重み



損失関数 Reconstruction error

テストデータとなる隣接行列との誤差が小さくなるように学習

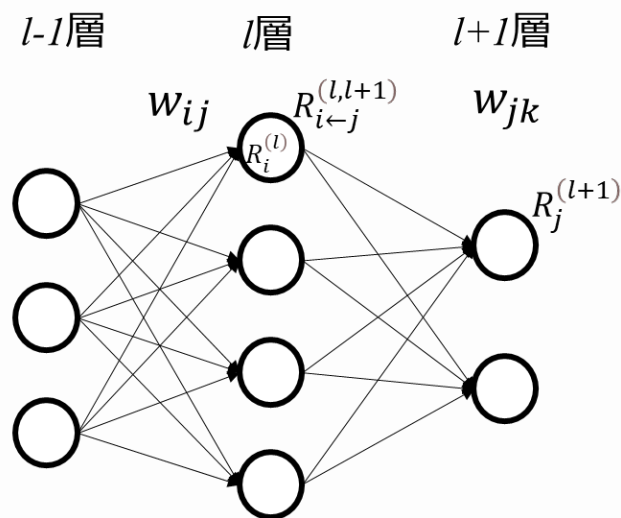
$$\mathbf{A}' = \text{Sigmoid}(\mathbf{H}\mathbf{H}^T)$$

$$L = \|\mathbf{y} - \mathbf{A}'\|_2^2 \quad \mathbf{y} = \tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}$$

提案手法 layer-wise relevance propagation

概要

- 深層学習における説明手法
- 層のユニットごとの関係性を逆伝播、入力データの出カデータへの関係性を計算



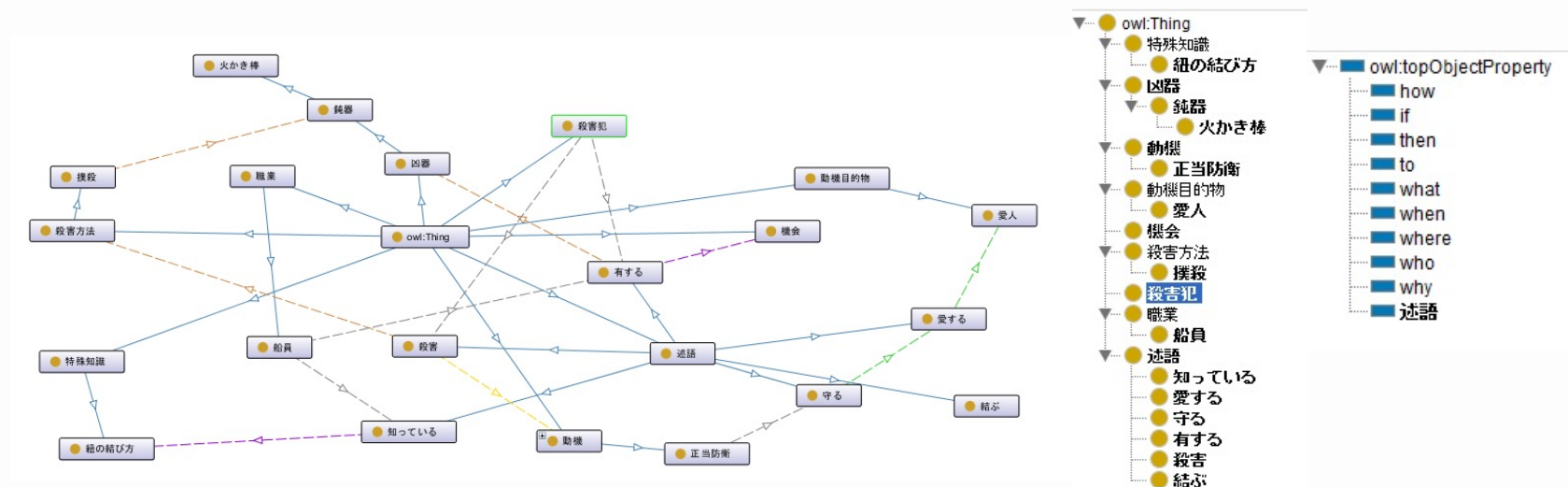
$$R_{i \leftarrow j}^{(l, l+1)} = \frac{z_{ij}}{z_j} R_j^{(l+1)}$$

$$R_i^{(l)} = \sum_j R_{i \leftarrow j}^{(l, l+1)}$$

- $R_i^{(l)}$ は層 l のユニット i の出力値に対する貢献度,
- $R_{i \leftarrow j}^{(l, l+1)}$ は層 $l+1$ のユニット j から層 l ユニット i へ逆伝播する貢献度,
- z_j は層 $l+1$ ユニットの出力値,
- z_{ij} は層 l のユニット i から層 $l+1$ ユニットへ順伝播する値

提案手法 独自オントロジー(僧坊莊園用)

独自オントロジー (鶉飼さん作成 犯罪方法オントロジー参考)



GCNの形式に合わせて、述語もノードとして僧坊莊園用のネットワークを構築

- 殺害犯は動機・凶器・機会を有する
- 正当防衛で愛する人 (lady_brackenstall) 守る → 殺害動機
- 船員は紐の結び方 (特殊知識) を知っている → 殺害機会

実験

実験設定

変更前	変更後
追加知識であるConceptNetに対して施した処理	
冠詞	無
複数形	単数形
2語以上の空白	_ (アンダーバー)
antonym	含めない
大文字	小文字
小説データに対して施した処理	
場面番号	場面番号_小説名
大文字	小文字

Antonymは反意語を示すため、学習には含めず
_小説名はConceptNet上の数字と差別化するため

結果

犯人推定

犯人Roylottの順位			
まだらの紐	+ConceptNet[murder]	+ConceptNet[murder] +ConceptNet[kill]	+ConceptNet[murder] +ConceptNet[kill] +ConceptNet[snake]
10%欠損	2	2	1
25%欠損	2	2	2

ConceptNetのデータを段階的に追加

ConceptNetのmurder, 実際の事件のデータを加えた場合は犯人推定一位

25%欠損で犯人推定が出来ていない

→snakeの知識が十分でない

結果

犯人推定

犯人Roylottの順位			
まだらの紐	+ConceptNet[murder] +ConceptNet[kill] +悪魔の足	+ConceptNet[murder] +ConceptNet[kill] +僧坊荘園	
10%欠損	2	2	
25%欠損	2	2	

犯人Jack Crockerの順位			
僧坊荘園	+独自オントロジー		
10%		1	

他小説のデータを加えただけでは、犯人推定はできず、構築した独自オントロジーを加えた場合は順位が一位

結果

LRP

LRPによる貢献度のRoyslottから見た重要ノード上位5つ

1	130_speckledband
2	Dog_whip_of_roylott
3	22_speckledband
4	23_speckledband
5	42_speckledband

Dog_whip_of_roylottは、実際の小説のまだらの紐にて殺害に用いられる犯行手段

→グラフ構造から関係性を学習

結果

LRP

LRPによる貢献度のjack crockerから見た重要ノード5つ	
1	thought_of_jack_crocker
2	308_abbey_grange
3	283_abbey_grange
4	332_abbey_grange
5	Sincerity_of_jack_crocker

犯人推定方法オントロジーにおける愛人lady brackenstall をbrackenstallから守る(正当防衛による)殺害

→thout_of_jack_crockerを近く学習

- ConceptNet上の語義の曖昧性
似たような意味、品詞の違う単語を考慮できていない。
WSD(Word Sense Disambiguation)により密なグラフを作成
- 嘘の考慮
ナレッジグラフの情報をそのまま学習しているため、
嘘の考慮が出来ない
- LRP
隣接しているノードの貢献度しか示すことが出来ない

推論チャレンジ(評価項目)

1. ホームズの推理を再現

- 小説と同じ知識で推論

2. 説明の納得性

- グラフが学習した内容を数的に可視化可能

3. 技術的工夫

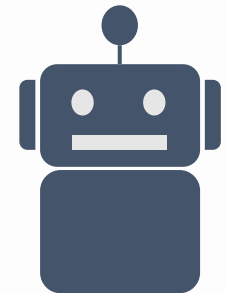
- 推論方法: GCN+LRP(機械学習)
- 拡張知識: 僧坊荘園用 独自オントロジー
- ConceptNetとナレッジグラフとの整合性処理

4. ポイント

- GCNは推論チャレンジのどのデータに適用可能
- 必要最低限の知識があれば推論
- 独自オントロジー 職業専門知識の導入

まとめ

- GCNとLRPを組み合わせることで、必要な知識があった場合犯人推定を行うことができた
- 追加知識にConceptNetと独自オントロジーを定義
- 追加知識の準備や定義方法に関して課題が残る
- より説明性の高いシステムの構築を目指す



ACKNOWLEDGEMENTS

This study is based on results obtained from projects, JPNP20006 and JPNP180013, commissioned by the New Energy and Industrial Technology Development Organization (NEDO).