

ナレッジグラフ推論チャレンジ ツール部門用応募フォーム

# グラフニューラルネットワークによる犯人推定

～まだらの紐と未知エンティティ問題～

穴田研究室4年

勝島修平

g1623019@tcu.ac.jp

# 目次(推論・推理過程の説明)

---

- 問題の概要
- アプローチの特徴と結論
- 既存研究の問題点
- 問題の定式化
- グラフニューラルネットワークについて
- ナレッジグラフデータのトリプルへの変換
- 結果
- 課題
- まとめ
- 実行環境・資料共有について



## 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

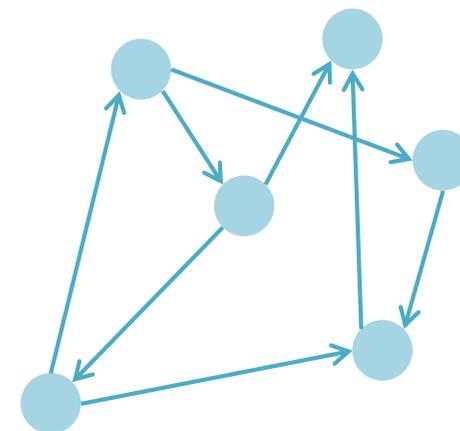
### ナレッジグラフ推論チャレンジの開催

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すること”

昨年度のチャレンジでは、

“知識グラフ補完＋述語論理式の充足可能性問題”

として解いた野村総合研究所が最優秀賞を獲得



# アプローチの特徴と結論

---

- 特徴

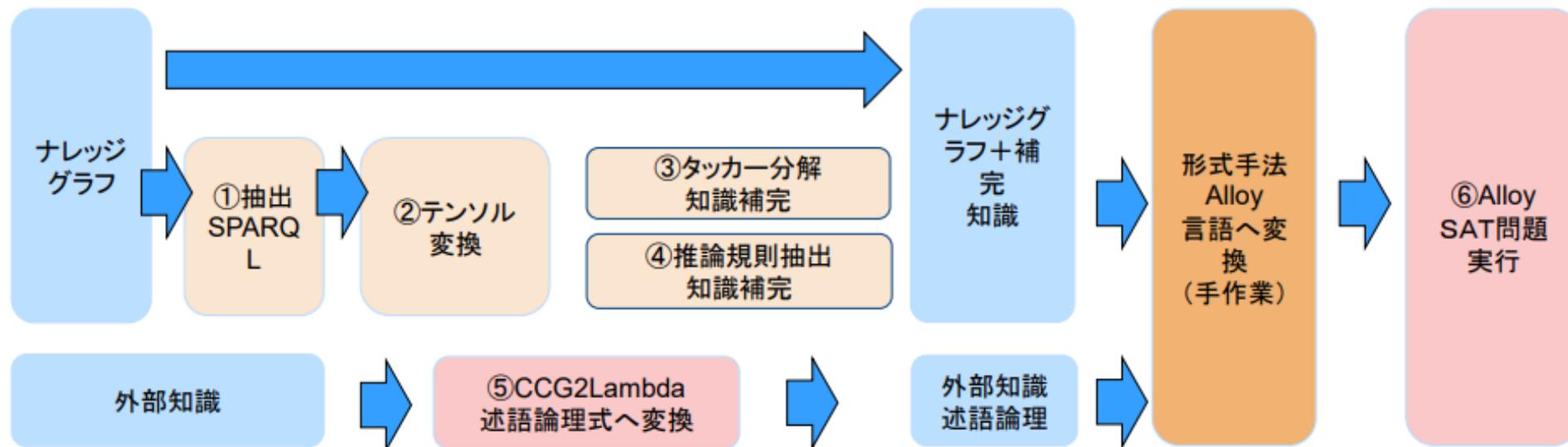
- ① 今回の問題を、文章のナレッジグラフ上に捜査方法が欠落していると考え、それを補完するために通常の知識グラフ補完ではなく、“**未知エンティティ問題**”とみなした
- ① 未知エンティティ問題の解法に“**グラフニューラルネットワーク**”を利用

- 結論

小説上の登場人物と、犯人であるロイロットとのトリプルの補完を確認(犯人推定のみ)

# 既存研究の問題点

## 野村総合研究所の推定フロー



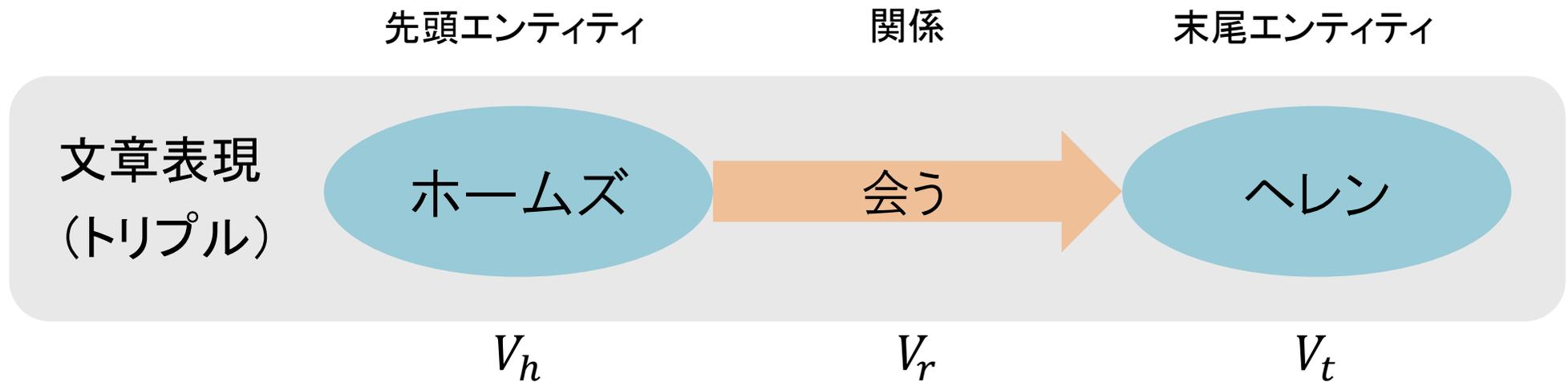
昨年度野村総合研究所の応募フォームから抜粋

日本語文からSAT問題に移行する際の自然言語処理の反復による、元の解釈とのずれ  
→できるだけ元の知識グラフの表現は変えないまま犯人推定を行いたい

ところが、通常の知識グラフ補完では、説明性のための“ルール”や“常識”の追加ができない

# 知識グラフ補完

TransE; エンティティや関係を連続値のベクトルで表現

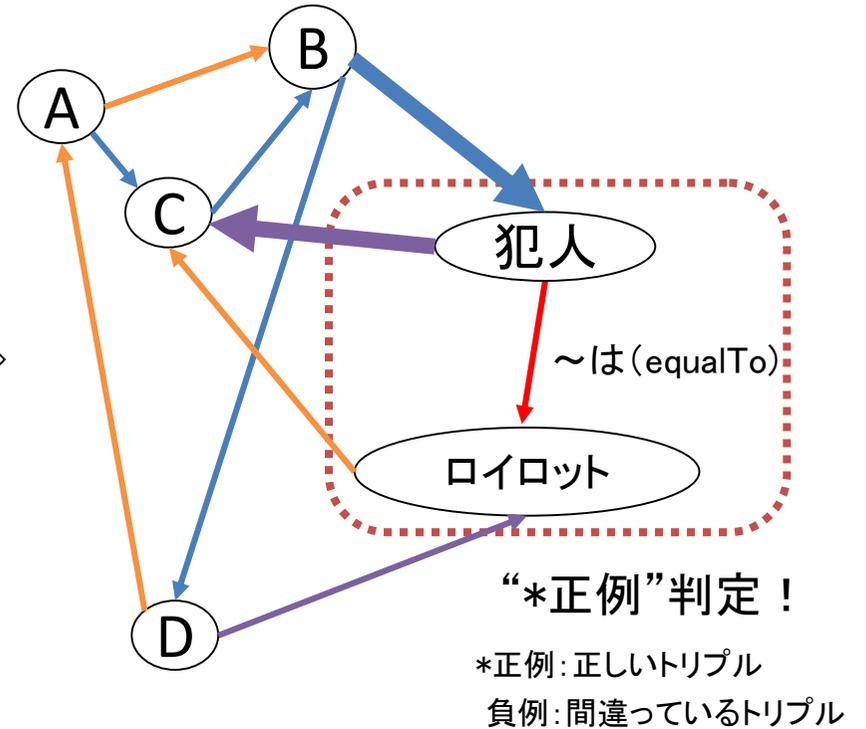
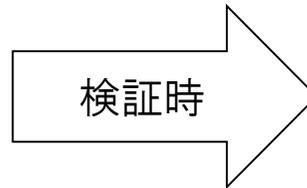
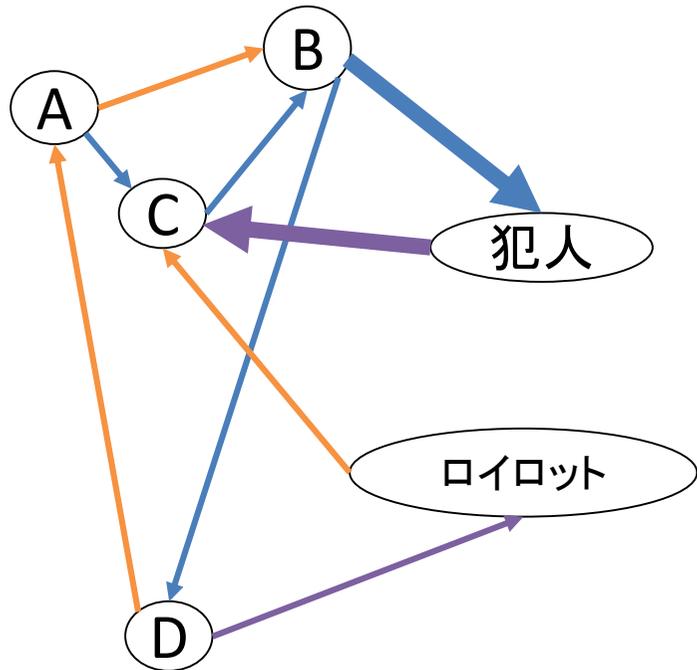


トリプルの先頭エンティティと関係、末尾エンティティの表現ベクトルがそれぞれ $V_h, V_r, V_t$ の時、

$V_h + V_r = V_t$  が成り立つようにベクトルの学習を行う

\*今回のナレッジグラフの加工についてはスライドp15参照

# 問題の定式化(通常の知識グラフ補完)

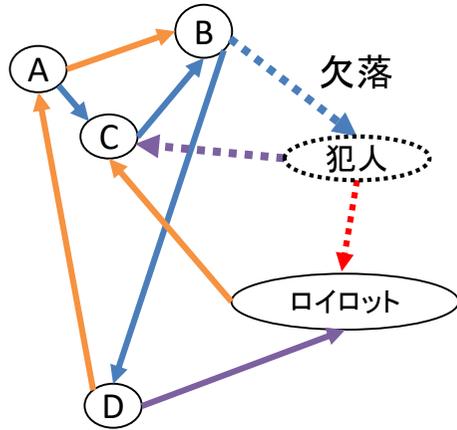


- 既存のナレッジグラフ上のエンティティのベクトルを訓練

- 学習したベクトルを利用し、赤線の関係を含めたオレンジ点線枠のトリプルの正誤を判断

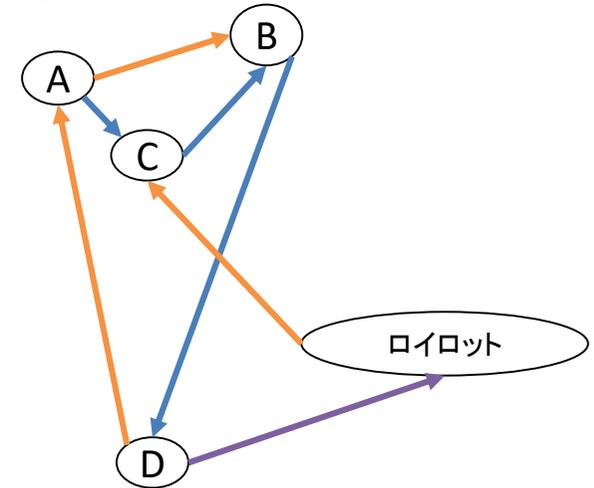
“犯人エンティティと小説上の犯人であるロイロットとのトリプルの補完”を目指す

# 問題の定式化(未知エンティティ問題)

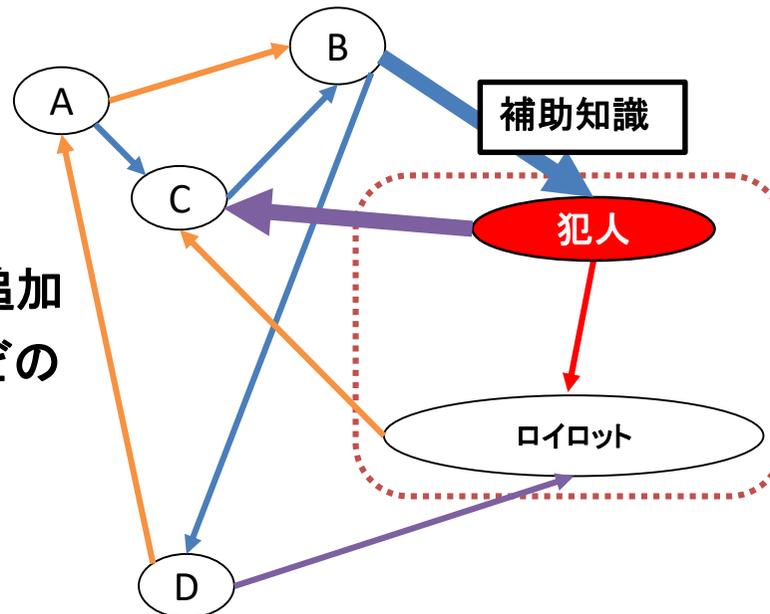


犯人を推定するための知識  
“常識”や“ルール”が欠落している  
=エンティティが不完全  
=**未知エンティティ問題**とみなす

ナレッジグラフを訓練



- 犯人エンティティを後から追加するため、そのベクトルをどのように獲得するかが課題



未知エンティティと補助知識を追加

# グラフニューラルネットワーク(GNN)

---

- 概要

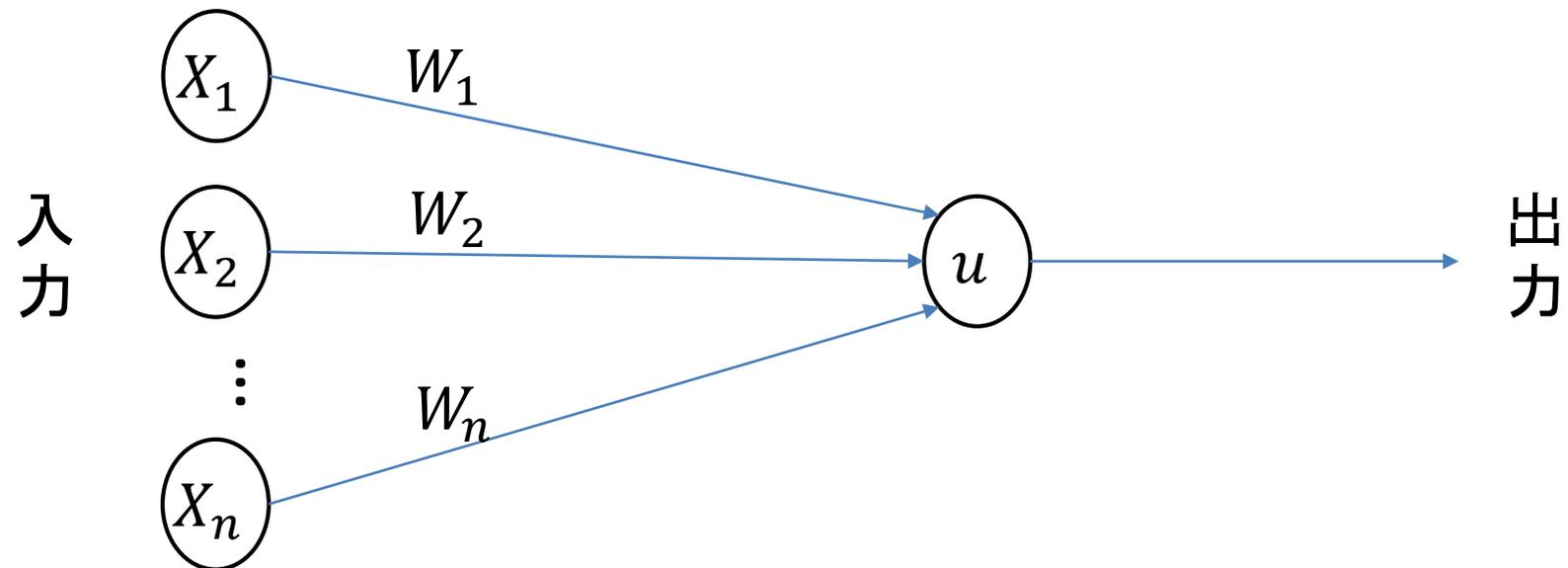
小説ナレッジグラフ上にないエンティティのベクトルを獲得するため、濱口らのグラフニューラルネットワークを導入

- 特徴

- ① グラフ構造上に定義されたニューラルネットワークの一種
- ② 伝播モデルと出力モデルを組み合わせることで、ベクトルの学習とトリプルの正誤判断を行う

# グラフニューラルネットワーク(GNN)

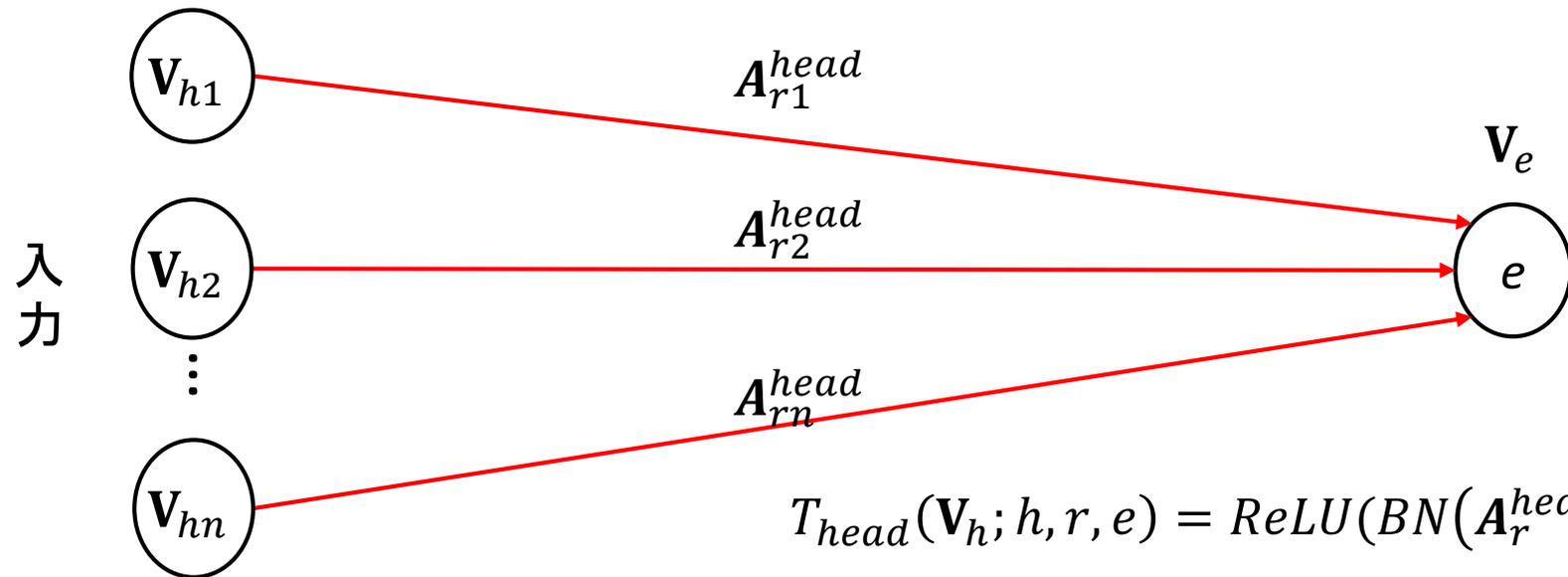
## ニューラルネットワーク



- ニューロンの信号を入力とし、それに重みをかけ、出力を行う

# グラフニューラルネットワーク(GNN)

## グラフニューラルネットワーク



$$T_{head}(\mathbf{V}_h; h, r, e) = \text{ReLU}(\text{BN}(\mathbf{A}_r^{head} \mathbf{V}_h))$$

$\mathbf{V}_h$ : 先頭エンティティのベクトル、 $T_{head}$ : 遷移関数、  
 $\text{BN}$ : batch normalization、 $\text{ReLU}$ : ランプ関数

- グラフニューラルネットワークでは、あるエンティティ( $e$ )の近傍エンティティのベクトルが入力、そしてそれをつなぐ関係が重みとなる

# グラフニューラルネットワーク(GNN)

## 伝播モデル

- 概要

伝播モデルを用いることにより、近傍のベクトルから未知エンティティのベクトルを獲得する

- 遷移関数

エンティティ $e$ とその近傍 $h, t$ との関係 $r$ を反映するような近傍のベクトル $\mathbf{V}_h, \mathbf{V}_t$ を変換する

$$T_{head}(\mathbf{V}_h; h, r, e) = \text{ReLU}(\text{BN}(\mathbf{A}_r^{head} \mathbf{V}_h))$$

$$T_{tail}(\mathbf{V}_t; e, r, t) = \text{ReLU}(\text{BN}(\mathbf{A}_r^{tail} \mathbf{V}_t))$$

- プーリング関数

先頭・末尾の近傍集合から共通の側面を抜き出す

$$S_{head}(e) = \{T_{head}(\mathbf{V}_h; h, r, e) | (h, r, e) \in N_h(e)\}$$

$$S_{tail}(e) = \{T_{tail}(\mathbf{V}_t; e, r, t) | (e, r, t) \in N_t(e)\}$$

$$\mathbf{V}_e = P(S_{head}(e) \cup S_{tail}(e))$$

$N_h(e), N_t(e)$ はそれぞれ先頭近傍、末尾近傍  
 $P$ はプーリング関数(要素ごとの平均)

# グラフニューラルネットワーク(GNN)

## 出力モデル

- 概要

伝播モデルで計算されたベクトルを用いてトリプルの正誤と最小化すべき関数を定義

- スコア関数

あるトリプル $(h, r, t)$ がどれくらい妥当かを評価(正しい知識を表すトリプルに対しては小さい値)

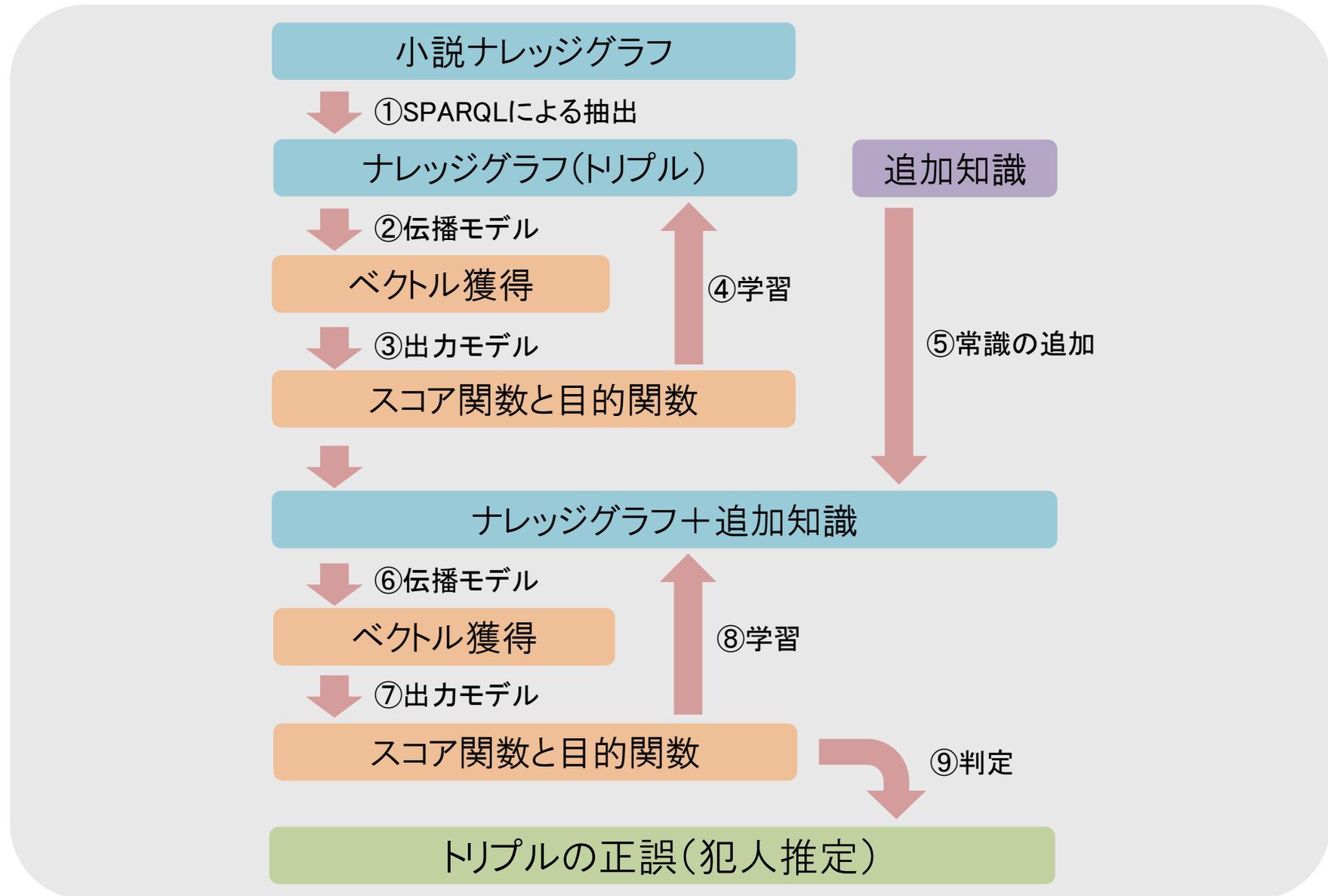
$$f(h, r, t) = \|\mathbf{V}_h + \mathbf{V}_r - \mathbf{V}_t\|$$

- 目的関数

スコア関数に基づいて最小化すべき関数を定義: $(h'_i, r_i, t'_i)$ は負例(間違っているトリプル)を表す

$$\zeta = \sum_i f(h_i, r_i, t_i) - [\tau - f(h'_i, r_i, t'_i)]_+$$

# 犯人推定までのプロセス



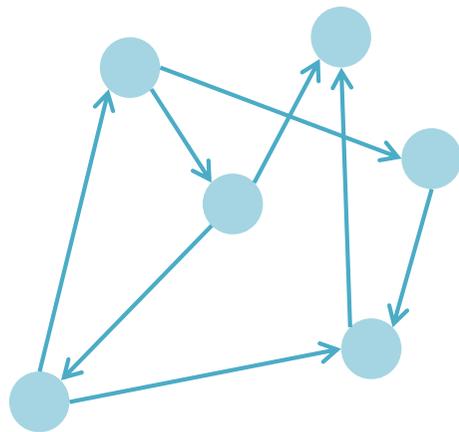
# ナレッジグラフデータのトリプルへの変換

- 目的

グラフニューラルネットワークを用いて解くためにナレッジグラフの変換を行う

- 概要

昨年度の野村総合研究所のテンソル変換を参考にし、文章のオントロジーデータからSPARQLによってRDFデータを抽出し、トリプルの表現に変換



id	主語	述語	目的語
1	Helen	come	house_of_Holmes
4	Helen	notHave	money
5	Helen	getMarried	within_2_months
6	Helen	obtain	money
6	Helen	obtain	within_2_months
7	Helen	pay	money
7	Helen	pay	reward_of_request
8	Helen	live	mansion_of_Roylott
9	Roylott	equalTo	father-in-law
10	Roylott	getMarried	mother_of_Helen
11	sister	include	Julia

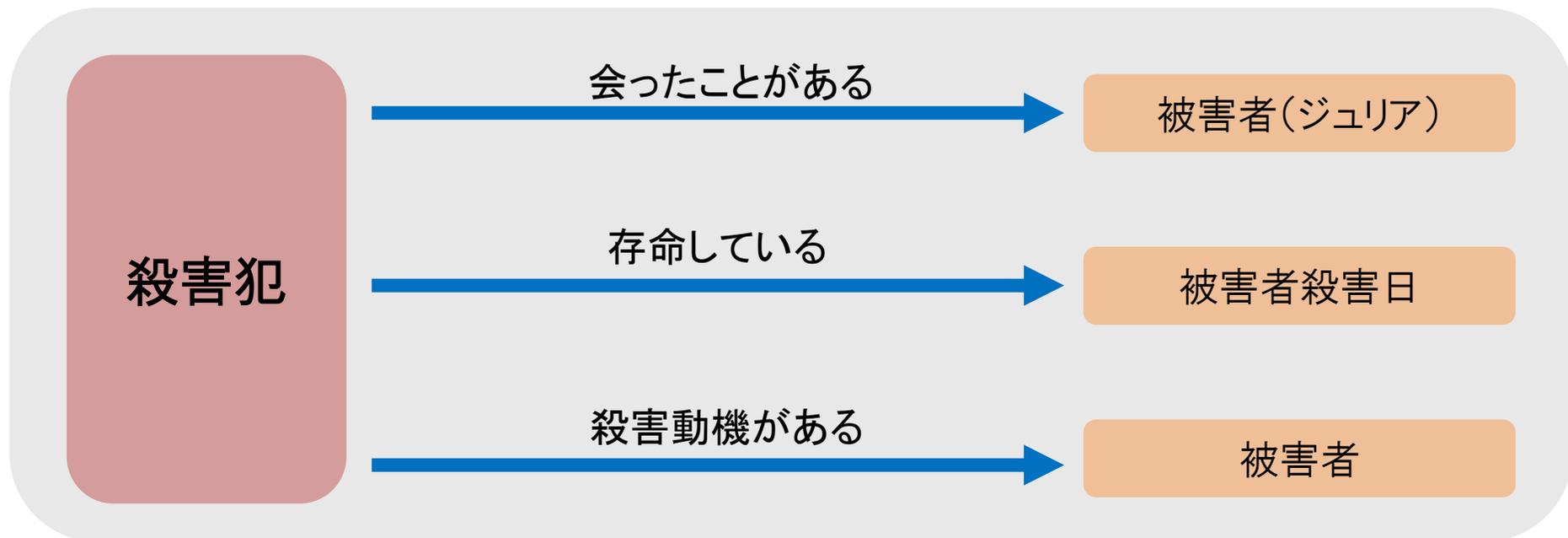
⋮

## 結果(仕様)

### ナレッジグラフの各要素数

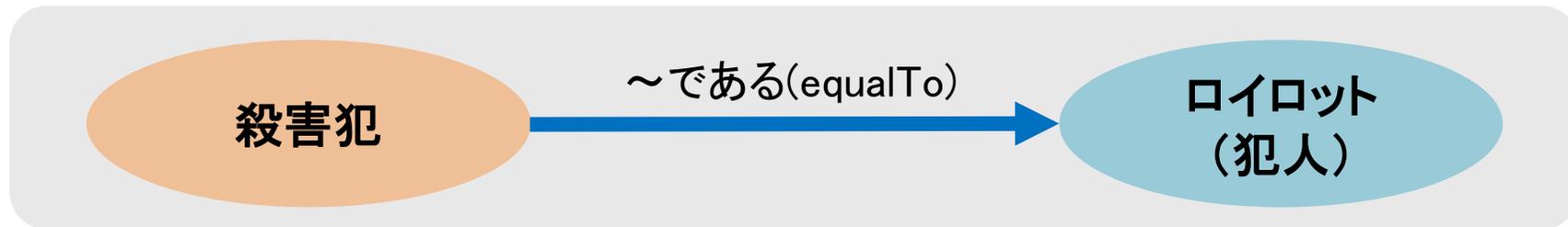
主語	述語	目的語	トリプル数
45	27	129	258

### 主な追加したエンティティと補助知識



# 結果

## “トリプルの補完”を確認

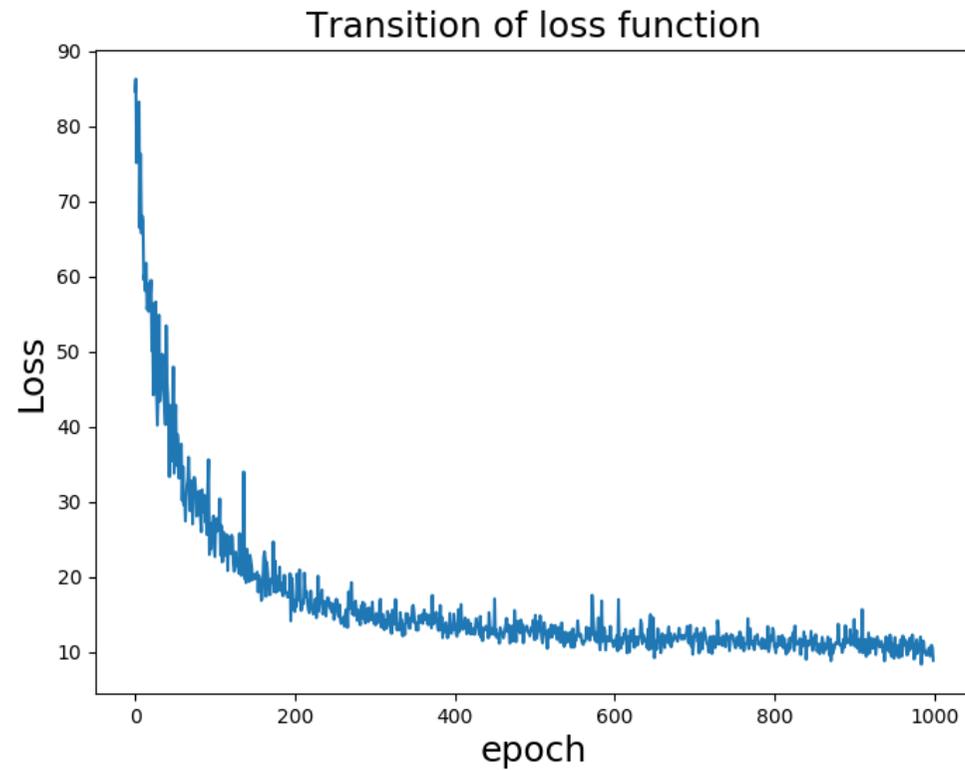


## 既存手法と提案手法の表現力の違い

	既存手法 (Tucker分解のみ)	提案手法
エンティティ数	18	151
総関係数	146	27
新知識の追加	不可	可能
犯人推定	不可	可能

Tucker分解のみでは捜査手法となる知識の追加が出来ないため犯人推定まではできないが提案手法では追加するエンティティとの関係性を定義することで犯人推定が行える

## 結果(目的関数の推移)



- 0には収束していないが、  
目的関数で負例のトリプルを考慮しているためだと考えられる

# 課題

---

- 追加知識の精査

今回は犯人推定を行えることを示すために主観に基づき追加知識を設定したが、説明性のために、文章からトリプルを作成したり、テキストからベクトルを求める方法を使って犯人推定を行うことが望まれる

- 嘘の考慮

小説のナレッジグラフの情報をそのまま取り込み学習を行っていくため、このモデルでは嘘の考慮ができない

- 手作業の削減

文章のオントロジーデータからトリプルを作成する際に手作業が多く入るためデータ整形に関するシステム化の向上を目指したい

## まとめ

---

- 本問題をナレッジグラフにおける“未知エンティティ問題”として解くことで小説上の犯人であるロイロットの推定を行うことができた
- エキスパートシステム寄りの方法ではなく、あくまで機械学習での推定を行った
- ただし、説明性の点では不十分なままとなってしまった
- 他の説明性のあるモデル(決定木など)での解釈が可能でないか、今後も模索したい

## 実行環境・資料共有について

---

- Windows10
- プロセッサ: Intel® Core™ i7-9700CPU
- Python3.7.3



- 参考文献

濱口拓男他: “未知エンティティを伴う知識ベース補完: グラフニューラルネットワークを用いたアプローチ”, 人工知能学会論文誌, 33巻2号F(2018年)

- 応募フォーム: 共有可
- 変換したナレッジグラフデータ: 共有可