

一般常識のナレッジグラフへの結合

～ 動詞の多義性解消の検討 ～

荒川博 黄瀬輝 佐々木裕之 辻寿嗣 大坪紹二
パナソニック株式会社 コネクティッドソリューションズ社 イノベーションセンター
arakawa.hrs@jp.panasonic.com

要旨

- 小説「まだらのひも」として提供されたナレッジグラフにおいて、英語特有の動詞の多義性により、SPARQLクエリ結果に不適切な結果が混入し、期待する推論結果を得られない場合がある。
- そこで、ナレッジグラフ中の多義性解消を目的に、word2vec、BERT の自然言語処理をベースに、対象文章と WordNet で取得した動詞の定義の文章との類似度を算出し、その類似度により多義性を判断する多義性自動判定手法を考案した。
- また、この多義性自動判定手法については、word2vec と BERT の各実験から、自動判定率の高かった BERT の結果を用いて、小説全体に適用した際の評価結果についても報告を行う。

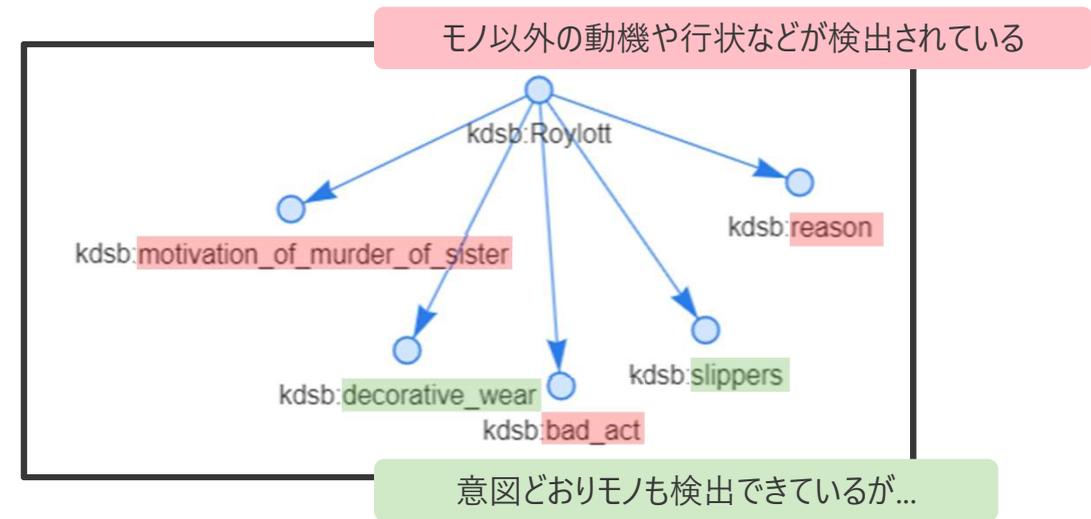
目次

1. 課題設定
2. 単語の定義を WordNet により特定
- 3-1. word2vec による動詞の多義性解消
- 3-2. BERT による動詞の多義性解消
4. 考察・課題

● 動機・疑問点

- 容疑者である Roylott が持っているモノ(凶器の候補)を検索しようとして
左下のクエリを発行すると、右下のグラフの通り意図しない項目も検索結果に含まれた

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX kgc: <http://kgc.knowledge-graph.jp/ontology/kgc.owl#>
SELECT ?sid ?s ?p ?o
FROM <http://kgc.knowledge-graph.jp/data/SpeckledBand>
WHERE {
  ?sid kgc:hasPredicate <http://kgc.knowledge-graph.jp/data/predicate/have> ;
  kgc:subject <http://kgc.knowledge-graph.jp/data/SpeckledBand/Roylott>;
  kgc:subject ?s;
  kgc:hasPredicate ?p;
  kgc:what ?o.
}
```



● 同じ単語 “have” であっても、別の意味で用いられる場合がある

- もちろん、前後の文脈等で単語の意味は特定できるものの...
- 以下の観点で“have-1”:モノを持つ、“have-2”:動機がある に分ける必要があると考える
 - 推論精度向上 (例えば、link prediction 等での潜在関係の抽出時、異なる意味のノードは峻別されるべき)
 - 推論結果をグラフ等で 可視化する場合の 説明容易性

1. 課題設定：「まだらのひも」主語:Royslott, 述語:have, 検索結果

```
<http://kgc.knowledge-graph.jp/data/SpeckledBand/91>
rdf:type      kgc:Situation;
kgc:source    "ロイロット博士の日頃の行いが悪い。"@ja ;
kgc:source    " Royslott's daily work is bad."@en ;
kgc:hasPredicate  <http://kgc.knowledge-graph.jp/data/predicate/have> ;
kgc:subject   <http://kgc.knowledge-graph.jp/data/SpeckledBand/Royslott> ;
kgc:what      <http://kgc.knowledge-graph.jp/data/SpeckledBand/bad_act>.
```

```
<http://kgc.knowledge-graph.jp/data/SpeckledBand/149>
rdf:type      kgc:Thought;
kgc:source    "ホームズは、ロイロット博士が娘を殺す動機があると証明した。"@ja ;
kgc:source    "Holmes proved that Royslott had a motive to kill her daughter."@en ;
kgc:infoSource  <http://kgc.knowledge-graph.jp/data/SpeckledBand/Holmes> ;
kgc:hasPredicate  <http://kgc.knowledge-graph.jp/data/predicate/have> ;
kgc:subject   <http://kgc.knowledge-graph.jp/data/SpeckledBand/Royslott> ;
kgc:what      <http://kgc.knowledge-graph.jp/data/SpeckledBand/motivation_of_murder_of_sister>.
```

```
<http://kgc.knowledge-graph.jp/data/SpeckledBand/177>
rdf:type      kgc:Situation;
kgc:source    "ロイロット博士は口実を作った。"@ja ;
kgc:source    " Royslott made an excuse."@en ;
kgc:hasPredicate  <http://kgc.knowledge-graph.jp/data/predicate/have> ;
kgc:subject   <http://kgc.knowledge-graph.jp/data/SpeckledBand/Royslott> ;
kgc:what      <http://kgc.knowledge-graph.jp/data/SpeckledBand/reason> ;
kgc:then      <http://kgc.knowledge-graph.jp/data/SpeckledBand/178>.
```

```
<http://kgc.knowledge-graph.jp/data/SpeckledBand/348>
rdf:type      kgc:Situation;
kgc:source    "ロイロット博士が、化粧着を着ていた"@ja ;
kgc:source    "Royslott was wearing a make-up cloth."@en ;
kgc:hasPredicate  <http://kgc.knowledge-graph.jp/data/predicate/have> ;
kgc:subject   <http://kgc.knowledge-graph.jp/data/SpeckledBand/Royslott> ;
kgc:what      <http://kgc.knowledge-graph.jp/data/SpeckledBand/decorative_wear>.
```

```
<http://kgc.knowledge-graph.jp/data/SpeckledBand/349>
rdf:type      kgc:Situation;
kgc:source    "ロイロット博士はスリッパをつけていた"@ja ;
kgc:source    "Royslott was wearing slippers."@en ;
kgc:hasPredicate  <http://kgc.knowledge-graph.jp/data/predicate/have> ;
kgc:subject   <http://kgc.knowledge-graph.jp/data/SpeckledBand/Royslott> ;
kgc:what      <http://kgc.knowledge-graph.jp/data/SpeckledBand/slippers>.
```

モノでない
動機(motive)の場合

意図どおりモノが
検出できている場合

ただし
kgc:source での動詞は
wear

1. 課題設定：他の多義語の例 (Roma と find)

課題：推論精度向上等のため、ナレッジグラフの単語定義の明確化が必要

● 以下の用語(ロマ)の意味は？

“Roma”とは小説の時代ではどのような人達？

```
<http://kgc.knowledge-graph.jp/data/SpeckledBand/25>
  rdf:type      kgc:Situation;
  kgc:source    "ロイロットはロマと付き合う"@ja ;
  kgc:source    "Royslott goes out with Roma"@en ;
  kgc:hasPredicate <http://kgc.knowledge-graph.jp/data/predicate/equalTo> ;
  kgc:subject   <http://kgc.knowledge-graph.jp/data/SpeckledBand/Roma> ;
```

● “find” という動詞でも、文脈で意味が異なる

```
<http://kgc.knowledge-graph.jp/data/SpeckledBand/253>
  rdf:type      kgc:Situation;
  kgc:source    "ホームズは、小さな犬鞭を見つけた。"@ja ;
  kgc:source    "Holmes found a small dog whip."@en ;
  kgc:hasPredicate <http://kgc.knowledge-graph.jp/data/predicate/find>
```

上の文章は「物を見つける」だが、下の文章は「現象を明らかにする」

```
<http://kgc.knowledge-graph.jp/data/SpeckledBand/267>
  rdf:type      kgc:Thought;
  kgc:source    "ホームズとワトソンは、ヘレンを悩ました音の正体を突き止める"@ja ;
  kgc:source    "Holmes and Watson will find the cause of the troubled sound."@en ;
  kgc:infoSource <http://kgc.knowledge-graph.jp/data/SpeckledBand/Holmes> ;
  kgc:hasPredicate <http://kgc.knowledge-graph.jp/data/predicate/find> ;
```

2. 単語の定義を WordNet により特定

課題である ナレッジグラフの単語定義明確化のため、英単語の概念辞書である WordNet を用いる。以下の通り、テキスト処理により英単語定義を入手可能のため。

- 公開された N-Triples の検索により
英単語の WordNet Identifier を
取得可能

```
$ grep find#find-[0-9]*-v wordnet.nt | grep LexicalSense | awk '{print $1}'  
<http://wordnet-rdf.princeton.edu/rdf/lemma/find#find-00251577-v>  
<http://wordnet-rdf.princeton.edu/rdf/lemma/find#find-00523831-v>  
<http://wordnet-rdf.princeton.edu/rdf/lemma/find#find-00717144-v>  
<http://wordnet-rdf.princeton.edu/rdf/lemma/find#find-00723361-v>  
<http://wordnet-rdf.princeton.edu/rdf/lemma/find#find-00920843-v>  
<http://wordnet-rdf.princeton.edu/rdf/lemma/find#find-00974031-v>  
<http://wordnet-rdf.princeton.edu/rdf/lemma/find#find-01641524-v>  
<http://wordnet-rdf.princeton.edu/rdf/lemma/find#find-02025505-v>  
<http://wordnet-rdf.princeton.edu/rdf/lemma/find#find-02133338-v>  
<http://wordnet-rdf.princeton.edu/rdf/lemma/find#find-02158938-v>  
<http://wordnet-rdf.princeton.edu/rdf/lemma/find#find-02202295-v>  
<http://wordnet-rdf.princeton.edu/rdf/lemma/find#find-02217057-v>  
<http://wordnet-rdf.princeton.edu/rdf/lemma/find#find-02218118-v>  
<http://wordnet-rdf.princeton.edu/rdf/lemma/find#find-02252785-v>  
<http://wordnet-rdf.princeton.edu/rdf/lemma/find#find-02253273-v>  
<http://wordnet-rdf.princeton.edu/rdf/lemma/find#find-02290474-v>
```

- WordNet Identifier を含む URL により当該定義を取得可能

```
$ curl http://wordnet-rdf.princeton.edu/json/id/00920843-v  
[{"definition": "establish after a calculation, investigation, experiment, survey, or study; ¥'find the product of two numbers¥'", "examples": [], "lemmas": [{"lemma": "ascertain", "language": "en", "forms": [], "sense_key": "ascertain%2:32:01::", "subcats": ["They %s up the hill", "They %s down the river", "The coins %s ", "They %s the countryside"], "importance": 3}, {"lemma": "determine", "language": "en", "forms": [], "sense_key": "determine%2:32:00::", "subcats": ["They %s the car down the avenue", "They %s the river", "The coins %s ", "The banks %s the check", "They %s up the hill", "They %s the countryside"], "importance": 6}, {"lemma": "find", "language": "en", "forms": ["found"], "sense_key": "find%2:32:00::", "subcats": ["The balls %s ", "They %s the car down the avenue", "The coins %s ", "The men %s the horses across the field", "The ropes %s", "The banks %s the check", "The horses %s across the field", "They %s up the hill", "His fields %s mine at this point", "They %s the countryside", "They %s the coin ", "The children %s to the playground"], "importance": 5} ...  
"foreign": {"jpn": ["突きとめる", "突き止める", "突止める", "見いだす", "見い出す", "見つけだす", "見つける", "見つけ出す", "見付ける", "見付け出す", "見出す", "見出だす", "解明+する"] ...
```

2. 単語の定義を WordNet により特定

WordNet RDF の定義を検索可能な python script を作成

- **Roma : 2個の定義があり、2番目の 定義が正解**

```
$ python ./scripts/search-rel_type.py Roma -1
```

```
-----  
1 08824749-n capital and largest city of Italy; on the Tiber; seat of the Roman Catholic Church; formerly the capital of the Roman Republic and the Roman Empire  
-----
```

```
2 10173755-n a member of a people with dark skin and hair who speak Romany and who traditionally live by seasonal work and fortunetelling; they are believed to have originated in northern India but now are living on all continents (but mostly in Europe, North Africa, and North America)
```

- **find : 18個の定義があり、前ページの場面ID(253,267) だと、それぞれ 17番目と 6番目の 定義が正解**

```
$ python ./scripts/search-rel_type.py find -1 | grep [0-9]*-v
```

```
...
```

```
-----  
6 00920843-v establish after a calculation, investigation, experiment, survey, or study; "find the product of two numbers"; "The physicist who found the elusive particle won the Nobel Prize "  
-----
```

```
7 00974031-v decide on and make a declaration about; "find someone guilty"
```

```
...
```

```
-----  
17 02290474-v come upon after searching; find the location of something that was missed or lost; "Did you find your glasses?"; "I cannot find my gloves!"
```

→ 上記 script は検索のみ。何らかの方法で、複数の定義から最適なものを選択したい。

3-1. word2vec による動詞の多義性解消：処理概要

動詞について、複数定義からの最適定義を自動選択する方式を検討
→ 定義に表れる例文の 主語と目的語の word2vec 埋め込みベクトルの
コサイン類似度最大となるものを最適定義として選択

WordNet と word2vec埋め込みベクトルを用いた 自動選択方式の概要

kill
問い合わせ

```
$ python search-rel_type.py kill -1;
-----
1 00218460-n the destruction of an enemy plane or ship or tank or missile; "the pilot reported two kills during the mission"
-----
2 00219879-n the act of terminating a life
...
10 01326622-v cause to die; put to death, usually intentionally or knowingly; "This man killed several people when he tried to rob a bank"; "The farmer killed a pig for the holidays "
...
18 02754370-v be fatal; "cigarettes kill"; "drunken driving kills"
```

10番目の定義を
選択できる仕組みを検討する

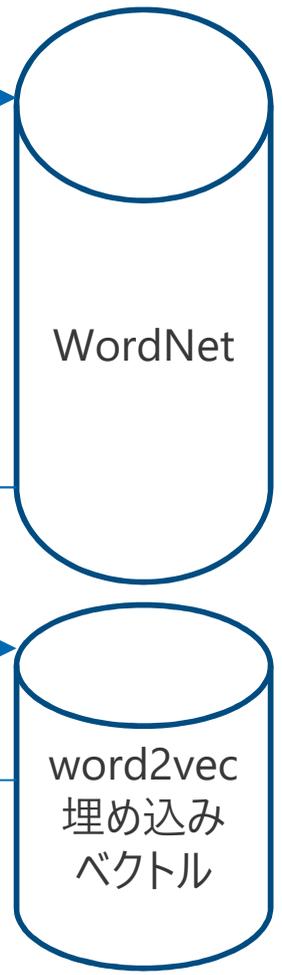
kill の定義毎、例文の主語・目的語の
word2vec 埋め込みベクトルを抽出して
平均ベクトルvecS, vecO算出

主語の平均ベクトルと目的語の平均ベクトルとを、
動詞としての kill (15個) 全てについて作成

$$\begin{aligned} \text{vecS}["\text{kill}"][10] &= \{ \text{vec}(\text{"man"}) + \text{vec}(\text{"farmer"}) \} / 2 \\ \text{vecO}["\text{kill}"][10] &= \{ \text{vec}(\text{"people"}) + \text{vec}(\text{"pig"}) \} / 2 \end{aligned}$$

● 入力ナレッジグラフ(ttl) での kill が出現する 場面ID での “主語” と “目的語” について、以下で
コサイン類似度算出

$$\text{Cost}(\text{"主語"}, \text{"目的語"}, \text{vecS}["\text{kill}"][i], \text{vecO}["\text{kill}"][i]) = \cos(\text{vec}(\text{"主語"}, \text{vecS}["\text{kill}"][i])) + \cos(\text{vec}(\text{"目的語"}, \text{vecO}["\text{kill}"][i]))$$



3-1. word2vec による動詞の多義性解消：特徴量ベクトル

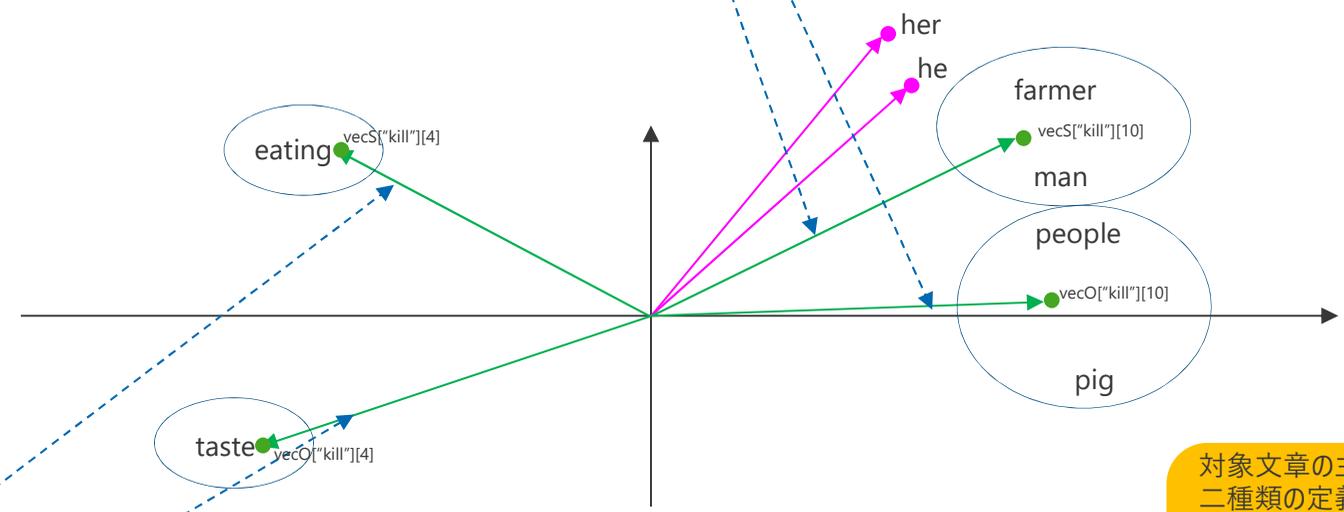
最適定義の自動選択の仕組み：対象文章が “He kills her” の場合

```
$ pythonw36 search-rel_type.py kill 10  
-----  
10 01326622-v cause to die; put to death, usually intentionally or knowingly; "This man killed several people when he tried to rob a bank"; "The farmer killed a pig for the holidays"
```



$$\begin{aligned} \text{vecS}["kill"][10] &= \{ \text{vec}(\text{"man"}) + \text{vec}(\text{"farmer"}) \} / 2 \\ \text{vecO}["kill"][10] &= \{ \text{vec}(\text{"people"}) + \text{vec}(\text{"pig"}) \} / 2 \end{aligned}$$

二種類の定義
01326622-v
00355831-v
のいずれが、対象文章に近いか、
を決定したい



$$\begin{aligned} \text{vecS}["kill"][4] &= \{ \text{vec}(\text{"eating"}) \} / 1 \\ \text{vecO}["kill"][4] &= \{ \text{vec}(\text{"taste"}) \} / 1 \end{aligned}$$

対象文章の主語 “he” の特徴量ベクトル $\text{vec}(\text{"he"})$ と二種類の定義での主語の特徴量ベクトル
01326622-v : $\text{vecS}["kill"][10]$
00355831-v : $\text{vecS}["kill"][4]$
との コサイン類似度は、前者の方が大きい。
 $\text{vecO}["kill"]$ についても同様のため、
前者の定義 (01326622-v) を最適定義として選択する

```
$ pythonw36 search-rel_type.py kill 4  
-----  
4 00355831-v destroy a vitally essential quality of or in; "Eating artichokes kills the taste of all other foods"  
  hypernym , 01622948-v, destroy  
  similar  , 00356019-v, kill
```

● 文例と 期待する正解：Speckled band での “find”

小説の一文	Wordnet での 対応すべき定義(正解)
<253> <find> "Holmes found a small dog whip." "ホームズは、小さな犬鞭を見つけた。"	02290474-v come upon after searching ; "find the location of something that was missed or lost"; "Did you find your glasses?"; "I cannot find my gloves!"
<267> <find> "Holmes and Watson will find the cause of the troubled sound." "ホームズとワトソンは、ヘレンを悩ました音の正体を突き止める"	00920843-v establish after a calculation, investigation, experiment, survey, or study ; "find the product of two numbers"; "The physicist who found the elusive particle won the Nobel Prize"

● word2vec を用いた場合の結果：全てNG

小説の一文	自動判定により抽出された定義(処理結果)
<253> <find> "Holmes found a small dog whip." "ホームズは、小さな犬鞭を見つけた。"	02202295-v, "perceive oneself to be in a certain condition or place "; ""I found myself in a difficult situation""; ""When he woke up, he found himself in a hospital room""
<267> <find> "Holmes and Watson will find the cause of the troubled sound." "ホームズとワトソンは、ヘレンを悩ました音の正体を突き止める"	02253273-v, " come upon, as if by accident ; meet with; ""We find this idea in Plato""; ""I happened upon the most wonderful bakery not very far from here""; ""She chanced upon an interesting book in the bookstore the other day""

● 考察

- <267> で 0253273-v が選ばれたのは、目的語 “idea” のためか
- 他の動詞 でも評価(次ページへ)

● 文例と 期待する正解：Speckled band での “put”

小説の一文	Wordnet での 対応すべき定義(正解)
<300> <put> "Watson put gently the handgun in the corner of the desk." "ワトソンは拳銃 (リヴォルヴァ) 机の隅に (そっと) 置いた"@	01496967-v put into a certain place or abstract location; "Put your things here"; "Set the tray down"; "Set the dogs on the scent of the missing children"; "Place emphasis on a certain point"
<391> <put> "Roylott put the snake in the safe." "博士は蛇を金庫に入れた"	01496967-v put into a certain place or abstract location; "Put your things here"; "Set the tray down"; "Set the dogs on the scent of the missing children"; "Place emphasis on a certain point"

● word2vec を用いた場合の結果：全てNG

小説の一文	自動判定により抽出された定義(処理結果)
<300> <put> "Watson put gently the handgun in the corner of the desk." "ワトソンは拳銃 (リヴォルヴァ) 机の隅に (そっと) 置いた"@	02115982-v,"cause (someone) to undergo something; ""He put her to the torture""
<391> <put> "Roylott put the snake in the safe." "博士は蛇を金庫に入れた"	00983308-v,"formulate in a particular style or language; ""I wouldn't put it that way""; ""She cast her request in very polite language""

● 考察

- NGの原因は、“put” の例文での 目的語 が “things” で抽象的だったため(?)
- 主語と目的語だけでは、制限が十分と思われる
→ [BERT](#)による、場面IDの文章と WordNet定義文毎との類似度による選択を試す(次ページ)

3-2. BERT による動詞の多義性解消

● 文例と 期待する正解 : Speckled band での “find”

小説の一文	Wordnet での 対応すべき定義(正解)
<253> <find> "Holmes found a small dog whip." "ホームズは、小さな犬鞭を見つけた。"	02290474-v come upon after searching; "find the location of something that was missed or lost"; "Did you find your glasses?"; "I cannot find my gloves!"
<267> <find> "Holmes and Watson will find the cause of the troubled sound." "ホームズとワトソンは、ヘレンを悩ました音の正体を突き止める"	00920843-v establish after a calculation, investigation, experiment, survey, or study; "find the product of two numbers"; "The physicist who found the elusive particle won the Nobel Prize"

● BERT を用いた場合の結果 : 全てOK

----- 253 find -----, -----, Holmes found a small dog whip. 0.902, 02290474-v, come upon after searching; find the location of something that was missed or lost; Did you find your glasses?; I cannot find my gloves! 0.900, 02218118-v, get something or somebody for a specific purpose; I found this gadget that will serve as a bottle opener; I got hold of these tools to fix our plumbing; The chairman got hold of a secretary on Friday night to type the urgent letter
----- 267 find -----, -----, Holmes and Watson will find the cause of the troubled sound. 0.904, 00920843-v, establish after a calculation, investigation, experiment, survey, or study; find the product of two numbers; The physicist who found the elusive particle won the Nobel Prize 0.903, 02290474-v, come upon after searching; find the location of something that was missed or lost; Did you find your glasses?; I cannot find my gloves!

● 考察

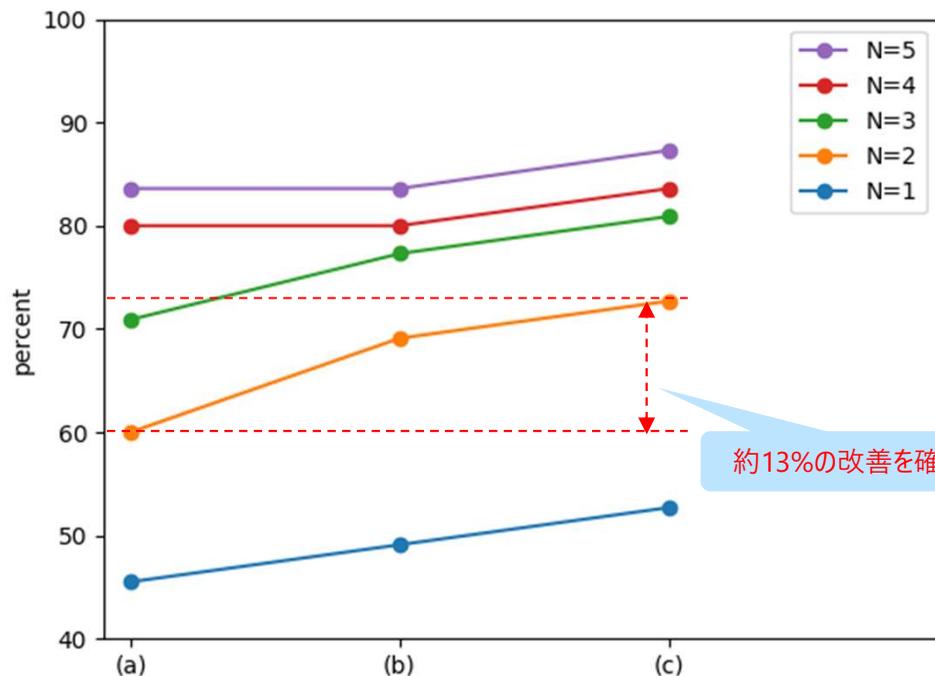
- 正解を選択。但し、第2候補との類似度差は微差。
- 「まだらのひも」の全文章について処理し、正誤を調べてみる(次ページ)

3-2. BERT による動詞の多義性解消 : 「まだらのひも」を処理してみる

下記のような kgc:source と kgc:hasPredicate の不整合 (lodge と live) を除いた 110文の真値を手動で作成し、性能評価を実施

```
<http://kgc.knowledge-graph.jp/data/SpeckledBand/26>
rdf:type      kgc:Situation;
kgc:source    "ロマはロイロットの土地で野宿している"@ja ;
kgc:source    "Roma is lodging on the land of Roylott"@en ;
kgc:hasPredicate <http://kgc.knowledge-graph.jp/data/predicate/live> ;
```

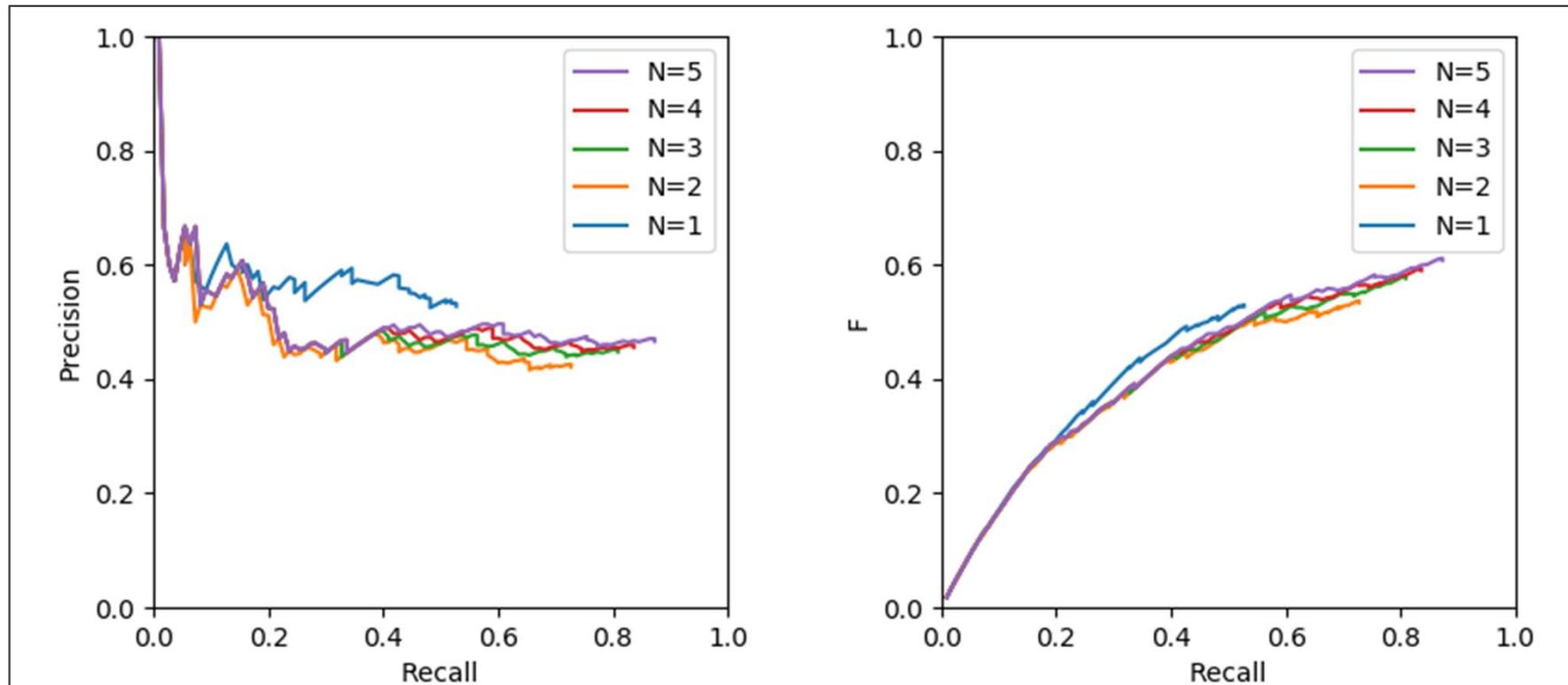
- 類似度で上位 N 個の中に正解が含まれる割合(%)を 3 つの設定 (a)~(c) で算出



	(a)	(b)	(c)
N=5	83.6	83.6	87.3
N=4	80.0	80.0	83.6
N=3	70.9	77.3	80.9
N=2	60.0	69.1	72.7
N=1	45.5	49.1	52.7

- (a) 固有名詞を一般名詞に変更
- (b) 固有名詞を変更しない
- (c) 自動判定で誤検出された名詞を除外

前ページで「類似度で上位 N 個の中に正解が含まれる割合」の意味を Precision/Recall および F値/Recall の観点で確認する



- F値が大きい場合が最良との基準だと、N=1～5の全てについて、Recall が最大の場合が最良となる。これは、
 - 類似度に対する閾値を、110文中での正解の類似度のうち最小の場合に設定することに相当する。また、
 - 前ページでの「類似度で上位N個の中に正確が含まれる割合」でもある

考察

- **一般常識のナレッジグラフへの結合に向け、動詞の多義性解消の方式を検討した**
 - 主語・目的語の word2vec埋め込みベクトル類似度による定義選択は機能しなかった
 - 用いた WordNet の例文での主語・目的語に問題があった、ため
 - BERT が文章全体の情報を用いるのに対し、word2vec による本方式は、主語と目的語しか用いないため
 - 文章の BERTベクトル類似度による定義選択の正解率(上位 2 候補内)は約73%
 - 不正解の場合の原因分析中

今後の課題

- **BERT文章ベクトルの改善による識別精度向上**
 - 定義の文章ベクトルの改善：定義を補う文章の追加(Web等の情報を追加?)
 - 対象の文章ベクトルの改善：場面ID の前後の文章を活用 & 元文章から動詞を抽出
- **元になる情報の検討**
 - WordNet 以外の Wikidata 等の利用可否など
- **日本語への直接適用**

- 対象とした作品

- まだらのひも (Speckled Band)

- 使用した知識

- WordNet RDF : <http://wordnet-rdf.princeton.edu/about>
 - N-Triples : <http://wordnet-rdf.princeton.edu/static/wordnet.nt.gz>
 - kill: 01326622-v の場合 → <http://wordnet-rdf.princeton.edu/json/id/01326622-v>

```
[{"definition": "cause to die; put to death, usually intentionally or knowingly; ¥\"This man killed several people when he tried to rob a bank¥\"; ¥\"The farmer killed a pig for the holidays¥\"", "examples": [], "lemmas": [{"lemma": "kill", "language": "en", "forms": [], "sense_key": "kill%2:35:00::", "subcats": ["They %s the car down the avenue", "The coins %s ", "The cars %s down the avenue", "They %s the coin ", "The children %s to the playground"], "importance": 7}], "id": "01326622-v", "ili": "i28306", "pos": "v", "subject": "verb. contact", "relations": [{"src_word": "kill", "trg_word": "killable"}, ...
```

- 使用したツール

- word2vec : [GloVe](#) の 学習済み埋め込みベクトル
 - glove.6B.100d.txt in <http://nlp.stanford.edu/data/glove.6B.zip>
- [bert-as-service](#) : 学習済みモデル
 - https://storage.googleapis.com/bert_models/2018_10_18/uncased_L-24_H-1024_A-16.zip

・応募フォーム

-公開の可否：

公開してよい

非公開とする

-公開形式：

ナレッジグラフ推論チャレンジのサイトで公開

独自のサイトで公開してリンクを希望

→公開先 URL (※)：

・応募したプログラム、データ等

-公開の可否：

公開してよい

非公開とする

-公開形式：

ナレッジグラフ推論チャレンジのサイトで公開

独自のサイトで公開してリンクを希望

→公開先 URL (※)：