2C3-OS-06b-2

# 関係知識獲得のための意味表現の学習

Learning meaning representation for relation knowledge acquisition

高瀬 翔 \*1

岡崎 直観 \*1

乾 健太郎 \*1

Sho Takase Naoaki Okazaki

Kentaro Inui

# \*1東北大学

Tohoku University

Vector representation is a common approach for expressing the meaning of a relational pattern. Most previous work regarded a relational pattern as a single 'word' for the vector acquisition. However, this approach suffers from the data sparseness problem because relational patterns are productive, i.e., produced by combinations of words. In order to address this problem, we propose a novel method for computing the meaning of a relational pattern based on the semantic compositionality of constituent words. We extend the Skip-gram model to handle semantic compositions of relational patterns by using Recursive Neural Network (RNN).

# 1. はじめに

近年,ニューラルネットワークを用いて単語や句の意味ベクトルを学習する手法がめざましい発展を遂げている [Mikolov 13, Socher 12] . 単語や句の意味ベクトルの学習・計算は自然言語処理において非常に重要であり,High-fat diet increases the risk of disease. という文から high-fat diet と disease の間に因果関係があるというように,エンティティ間の意味的関係をコーパスから抽出する関係抽出タスクでも,エンティティ間を結ぶ統語的なパタン(関係パタン,上記の例では increases the risk of) がどのような意味であるか(上記の例では意味的に cause に近く,因果関係を表す)の計算のため,意味ベクトルを得る必要がある.

一般的に,関係パタンの意味ベクトルは,関係パタンを一単語とみなし,単語の意味ベクトルの獲得と同様の手法で行われる [Nakashole 12] . しかしながら,この手法では,複数の単語から構成される関係パタンについて,データスパースネス問題(出現頻度の低下)により,意味ベクトルの質が著しく低下してしまう.この問題に対処するには,関係パタンの意味を構成要素から計算する(例えば increase と risk の意味からincrease the risk of の意味を計算する)必要がある.

関係パタンの意味を構成要素から計算する単純な手法として,単語ベクトルの加法構成性を利用する手法がある [Mitchell 10, Mikolov 13] . 特に Mikolov らの提案した Skip-gram モデル [Mikolov 13] で学習した単語の意味ベクトルは  $v_{\rm king}-v_{\rm man}+v_{\rm woman}\approx v_{\rm queen}$  のような意味の演算を高精度でできる事で注目を浴びた .

しかしながら,加法構成性はベクトルの線形結合であるため,意味の変性を扱うことができない.例えば,reduce the risk of という関係パタンは risk の意味を reduce が不活性化する働きをし [Hashimoto 12],パタン全体では inhibit の意味に近くなる.Baroni らはこのような意味の変性を行う表現(機能的な表現)はベクトルよりも行列で表現すべきであると示した [Baroni 10].著者らは意味の変性を扱いつつ,関係パタンの意味ベクトルを構成的に計算,学習する手法として,Skipgram モデルに Recursive Neural Network (RNN)を統合した手法を提案した [高瀬 15].

連絡先: 高瀬翔, 東北大学, 宮城県仙台市青葉区荒巻字青葉 6-6-05, 022-795-7091, takase@ecei.tohoku.ac.jp 本研究では,著者らの手法 [高瀬 15] を発展させ,パラメータ数を減らすために $L_1$  正則化を導入する.これにより,性能を悪化させることなくパラメータを削減できる事を,実験を通して明らかにする.また,先行研究 [高瀬 15] では独自の品詞パタンで関係パタンを認識していたが,本研究ではReverb [Fader 11] を用いてより一般的な関係パタンの扱いを可能にした.これにより,関係パタン間の類似度計算タスクにおける評価事例が増え,実験結果の信頼性を向上させた.さらに,提案手法で学習した意味ベクトルを SemEval-2010 task 8 の関係抽出タスクに適用し,その効果を明らかにした.

# 2. 提案手法

図 1 に描かれた,原形化された単語列  $demand\ help\ have\ a$   $major\ impact\ on\ price$  を例にとり,著者らの手法 [高瀬 15] の概要を説明する.既に説明したように,関係パタンを  $help\ have\ a\ major\ impact\ on\ という一単語としてみなす手法(図 1 の上部)では,良質な関係パタンの意味ベクトルを得ることができない.そこで本手法では,動詞による意味の変性を考慮しつつ,構成要素である <math>help\ have\ major\ impact\ n$ ら RNNを用いて関係パタンの意味ベクトルを計算する.

本研究では、関係パタンに頻出する動詞は意味を変性する機能的な表現であると仮定し、行列で表現する.この機能的な表現の定義については議論の余地があるが、本研究は関係パタンの意味ベクトルの構成的な計算手法を探求する第一歩として、以下の仮定を置いている.

- 1. 動詞は機能的な表現となる(意味を変性できる).
- 2. 動詞が機能的な表現であるかどうかは関係パタン中の統計値に基づいて決定する.
- 3. 機能的な表現以外の動詞,名詞,形容詞,副詞は内容的な意味のみを持つとし,内容的な表現と呼び,ベクトルで表現する.

RNN のパラメータの学習, すなわち, 行列とベクトルの学習のために, 良質な単語の意味ベクトルが学習可能であり, かつ, オンライン学習で RNN の組み込みを行いやすい Skip-gram モデルを拡張する.

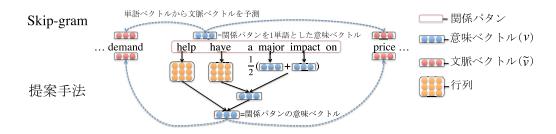


図 1: 提案手法の概要 (Skip-gram モデルとの対比)

## 2.1 Skip-gram モデル

まず、本手法がベースとする Skip-gram モデルを説明する. 単語列からなるコーパスを  $\mathcal{D}=w_1,w_2,...,w_T$ , コーパス中に含まれる単語の集合を V とする. Skip-gram モデルは、式 1 の目的関数を最小化する.

$$J = -\sum_{w \in \mathcal{D}} \sum_{c \in C_w} \log p(c|w) \tag{1}$$

ただし, $C_w$  は単語 w の文脈単語列で,単語 w の p 個前の単語を  $w_{-p}$ ,p 個後の単語を  $w_{+p}$  と表すことにすると, $C_w=\{w_{-h},...,w_{-1},w_{+1},...,w_{+h}\}$ (h は文脈の広さを調整するパラメータ)である.P(c|w) は単語 w から文脈単語 c を予測する確率で,log-bilinear モデルで定式化される.

$$p(c|w) = \frac{\exp\left(\boldsymbol{v}_w \cdot \tilde{\boldsymbol{v}}_c\right)}{\sum_{c' \in V} \exp\left(\boldsymbol{v}_w \cdot \tilde{\boldsymbol{v}}_{c'}\right)}$$
(2)

ここで、 $v_w$  は単語 w のベクトル(次元数 d)、 $\tilde{v}_c$  は文脈 c の ベクトル(次元数 d)である  $^{*1}$ . 式 2 の分母はコーパスに含まれる全ての単語に関する内積の和を求めるものであり、計算コストが高い.そこで、Mikolov らは観測された文脈語 c と人工的にサンプリングした k 個のノイズ(擬似負例単語 z)を識別できるようなロジスティック回帰モデルを学習する、Negative Sampling を提案している.

$$p(c|w) \approx \log \sigma(\boldsymbol{v}_w \cdot \tilde{\boldsymbol{v}}_c) + k \mathop{\mathbb{E}}_{z \sim P_n} [\log \sigma(-\boldsymbol{v}_w \cdot \tilde{\boldsymbol{v}}_z)]$$
 (3)

ここで、 $P_n$  はノイズをサンプリングするための確率分布である。本研究では Mikolov らと同様に、コーパス中での各単語の出現確率を 0.75 乗したものを用いる.

# 2.2 関係パタンの構成的な意味計算

本手法では、内容的な表現を組み合わせて関係パタンの意味を計算する際、ベクトル空間上で加法構成性が成り立つと仮定する。すなわち、内容的な表現列 $p_{n+1},...,p_m$ の意味は、各内容的な表現の意味ベクトルの平均で計算する。

$$\frac{\mathbf{v}_{p_{(n+1)}} + \mathbf{v}_{p_{(n+2)}} + \dots \mathbf{v}_{p_m}}{m-n} \tag{4}$$

これに対し、各機能的な表現  $p_i$  は写像  $f_{p_i}:\mathbb{R}^d\to\mathbb{R}^d$  で意味ベクトルを変換すると仮定する。したがって、関係パタン P の意味ベクトル  $\mathbf{v}_P$  は式 5 で計算する。

$$\mathbf{v}_P = f_{p_1} \left( f_{p_2} \left( \dots f_{p_n} \left( \tanh \left( \frac{\mathbf{v}_{p_{(n+1)}} + \mathbf{v}_{p_{(n+2)}} + \dots \mathbf{v}_{p_m}}{m-n} \right) \right) \right) \right)$$
 (5)

写像  $f_{p_i}$  は RNN に基づいて設計する. すなわち、機能的な表現  $p_i$  の意味変換写像は、行列  $W_i$  ( $d \times d$  次元) と活性化関数でモデル化する.

$$f_{p_i}(\mathbf{v}) = \tanh(W_{p_i}\mathbf{v}) \tag{6}$$

まとめると、本手法は関係パタンの内容的な表現列をベクトルの平均で合成し、機能的な表現の意味を RNN で変換するモデルである。例として、help have a major impact on という関係パタンの意味ベクトルは、図 1 に示したように、majorと impact の意味ベクトルの平均を have および help の行列を用いて変換する事で得られる。こうして得られた関係パタンの意味ベクトルから文脈単語 demandや price を予測できるよう学習を行う。

## 2.3 学習

単語と関係パタンの意味ベクトルは、Negative Sampling に基づく Skip-gram モデルと同様の手順で学習する。ただし、本手法では以下の拡張を行う。

- 1. 学習するのは、単語ベクトル  $v,\tilde{v}$  に加えて、機能的な表現の意味行列 W である。単語ベクトル  $v,\tilde{v}$  は、関係パタンを考慮しない通常の Skip-gram モデルでの学習結果で初期化する。機能的な表現の意味行列 W の要素は平均0、分散 0.01 のガウス分布からサンプリングした値で初期化する。
- 2. 関係パタンの意味ベクトルは式5で計算する。例えば、関係パタンから周辺の文脈語を予測する場合は、式3の $v_w$ の代わりに、式5の $v_P$ を用いる。関係パタンを構成する内容的な表現のベクトルと機能的な表現の行列は、誤差逆伝搬法に基づいて更新する。なお、関係パタンを周辺の単語から予測する際には5式を使わず、各関係パタン毎に割り当てたベクトル $\tilde{v}_P$ を用いる。

<sup>- 1</sup> 学習後に 2 種類のベクトル v と  $\tilde{v}$  が獲得されるが、v のみを単語ベクトルとして採用し、 $\tilde{v}$  は利用しない.

3. 本手法では機能的な表現の意味ベクトルの計算の部分で活性化関数 tanh を適用しているため,関係パタンの意味ベクトルの各次元は (-1,1) の範囲に収まる.一方,元々の Skip-gram モデルでは活性化関数を用いていないため,単語の意味ベクトルの値域に制限がない.本手法では,関係パタンと単語の意味ベクトルの値域を合わせるため,単語ベクトル v にも活性化関数 tanh を適用する.

ところで、 $take\ care\ of\ O\ take\$ や  $have\ access\ to\ O\ have\$ の ようにいくつかの機能的な表現は関係パタンの意味に強い影響を及ぼさないと考えられる。本研究では  $L_1$  正則化を導入することで、機能的な表現の行列を対角行列に近づくようにし、意味計算に強く影響する成分以外を 0 にする事を考える。このとき、目的関数は以下のようになる。

$$J' = -\sum_{w \in \mathcal{D}} \sum_{c \in C_w} \log p(c|w) + \lambda \sum_{W \in \mathbb{W}} r(W)$$
 (7)

ここで, $\mathbb W$  は全ての機能的な表現の行列を表す.関数 r(W) は行列 W の対角成分以外の要素の  $L_1$  ノルムを計算する,すなわち以下のような式で表される.

$$r(W) = \sum_{i \neq j} |W_{i,j}| \tag{8}$$

## 3. 実験

## 3.1 実験設定

単語の意味ベクトルや機能的な表現の行列を学習するためのコーパスとして, $ukWaC^{*2}$ を利用した.このコーパスは,ukドメインから収集した Web ページのテキストを収録しており,約 20 億個の単語からなる.実験では,小文字に変換したレンマを単語とし,動詞の過去分詞形だけは表層形をそのまま用いた  $^{*3}$ .さらに,a,b のような 1 文字からなるトークン,the などの冠詞や what のような疑問詞,前置詞をストップワードとして除去した.

関係パタンを得るために Reverb [Fader 11] を ukWaC に適用し,さらに,無意味な関係パタンを除去するため,確信度 0.9 以上での出現が一度もないもの,yesterday や tonight のような時制表現を含むもの,コーパス中での頻度が 5 未満のものを除去した.また,実験では未知の関係パタンの意味を構成要素から計算可能かを検証するため,評価データに含まれる関係パタンは訓練データから除去した.この結果,55,885 種の関係パタンを学習対象として得た.得られた関係パタンのうち,5 種以上に出現している,be 以外の動詞 697 件を機能的な表現とした.機能的な表現以外の動詞,名詞,形容詞,副詞を内容的な表現とし,関係パタンが機能的な表現のみからなる場合,末尾の動詞を内容的な表現とした.

ukWaC コーパス中に出現する単語や関係パタンのうち,10 回以上出現するものを学習対象とした.なお,機能的な表現として認定された動詞(例えば cause)が関係パタン以外で出現した場合は,単語が出現したと見なし,通常の Skip-gram と同様に単語ベクトルを更新する.ベクトルの次元 d=50,文脈の広さ h=5,Negative Sampling の数 <math>k=5 とし, $10^{-5}$  サブサンプリングで Skip-gram および提案手法の学習を行った.

## 3.2 評価データ

本研究では,関係パタン間の類似度計算および関係抽出という,2つのタスクで実験を行う.

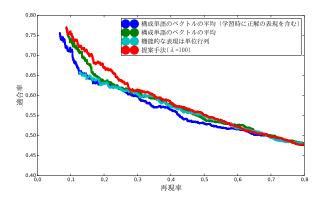


図 2: 各手法での精度と再現率

#### 3.2.1 関係パタン間の類似度計算

関係パタンの意味ベクトルを計算可能か評価するためには、関係パタン間に意味的類似度が付与されたデータセットを用いる事が最適である.しかしながら,このようなデータセットは著者の知る限り存在しないため,代わりに,関係パタン間に含意関係が付与されたデータセット [Zeichner 12]\*4 を用いる.このデータセットでは,Reverbを用いて抽出した関係パタンのペア(例えば prevent と reduce the risk of)について,無意味な関係パタンでないかどうかと,含意関係であるかどうかを付与している(上記の例では含意).本実験では,含意を類似,非含意を非類似とみなし,評価を行う.無意味な関係パタンのペアと項が逆転しているペアを取り除いた結果,5,409ペアを評価用のペアとして得た.

#### 3.2.2 関係抽出

提案手法の関係抽出タスクでの効果を検証するため,SemEval-2010 task 8 のデータセット [Hendrickx 10] を用いる.このタスクは,与えられた 2 つのエンティティ間の関係を適切な関係ラベルに分類するタスクである.関係ラベルは 19 個であり,エンティティの順序を考慮した 9 個の関係ラベル(例えば Cause-Effect)と順序のない Other のラベルからなる.例えば,"The burst has been caused by water hammer pressure" という文中の'burst' と'pressure'という 2 つのエンティティについては,順序を考慮した,Cause-Effect( $e_1,e_2$ ) というラベルが付与される.データセット中には 10,717 の関係インスタンスがあり,うち 8,000 が訓練事例,残りの 2,717 がテスト用のインスタンスである.

#### 3.3 結果

#### 3.3.1 関係パタン間の類似度計算

提案手法と比較手法の適合率—再現率曲線を図 2 に示した.なお, $L_1$  正則化のコスト  $\lambda$  に関しては,値を変えてもほとんど結果が変わらなかったため(表 1 を参照),最も良い結果である  $\lambda=100$  の結果のみを示した.提案手法(赤色)はほとんどの領域において,他の手法よりも高い性能を達成している.特に,強力なベースライン手法である,関係パタンの意味ベクトルの計算に Skip-gram モデルで学習した単語ベクトルの平均を用いる手法(緑色)よりも高い性能を示している.評価データに出現する関係パタンも含め,関係パタンを一単語とみなして意味ベクトルを学習し,既知の関係パタンについては学習した意味ベクトルを直接用いる手法(青色)は最も悪い結果となっている.この結果は,関係パタンを一単語とみなして

<sup>\*2</sup> http://wacky.sslmit.unibo.it

<sup>\*3</sup> 能動態と受動態の区別を付けるため

<sup>\*4</sup> http://u.cs.biu.ac.il/~nlp/resources/downloads/annotation-of-rule-applications/

表 1: 各手法での適合率—再現率曲線における曲線下面積 (AUC)と  $L_1$  正則化によるモデルのパラメータ削減率 (行列のゼロの要素数)

Method	AUC	Sparsity
Method	AUC	Sparsity
構成単語のベクトルの平均	0.557	
(学習時に正解の表現を含む)		_
構成単語のベクトルの平均	0.568	_
機能的な表現は単位行列	0.552	_
提案手法 $(\lambda = 0)$	0.570	0.0%
提案手法 $(\lambda = 1)$	0.570	0.0%
提案手法 $(\lambda = 10)$	0.570	0.7%
提案手法 $(\lambda = 100)$	0.573	14.4%

表 2: 各手法での F1 スコアのマクロ平均 . 下の 3 行は state-of-the-art な手法の , 各論文に記された値 .

F1
~~ -
68.7
76.7
74.8
79.1
82.7

意味ベクトルを直接コーパスから得るよりも,構成要素から計算した方が良い事を示している.また,機能的な表現の写像に対応する行列を単位行列に固定した場合(水色)との比較でも提案手法が優れている事から,各機能的な表現について,個別の行列で意味の変成作用をモデル化するべきであると言える.

提案手法と他の手法とで類似度上位のペアを比較した際に, 提案手法でのみ獲得できているペアとしては inhibit と prevent the growth of や prevent と reduce the risk of などがあった. この事から,提案手法では inhibit や prevent による意味の変 性を扱えている事が分かる.

表 1 には図 2 の各手法での適合率—再現率曲線における曲線下面積を示した.この表からも,提案手法がベースライン手法である構成単語のベクトルの平均手法よりも良くなっている事が分かる.また,表 1 には曲線下面積に加え, $L_1$  正則化でのコスト  $\lambda$  を変化させた際のモデルのパラメータ削減率(機能的な表現の行列におけるゼロ要素の割合)を示した.この表から, $\lambda=100$  の際には 14.4%の削減率でありながら,最も良い性能(AUC:0.573)を達成している事が分かる.この事から,対角要素以外への  $L_1$  正則化により,性能を悪化させることなく,簡素なモデルを獲得できたと言える.

# 3.3.2 関係抽出

提案手法の関係抽出タスクでの性能を表 2 に示した.本実験では,与えられたエンティティペアがある関係ラベルに分類されるか否かを予測する,一対他のロジスティック回帰による分類器を作成した.ベースライン素性としてはエンティティペアの間に出現している単語の品詞,表層系,レンマとエンティティペアのレンマを用いた.これに加え,エンティティペアの間に出現している関係パタンの意味ベクトルを提案手法で計算して素性とし,さらに,エンティティペアの意味ベクトルも素性として利用した.関係パタンとエンティティペアの意味ベクトルを用いる事により,ベースライン素性のみを用いた場合から F1 スコアが大きく向上している事が分かる( $68.7 \rightarrow 76.7$ ).この事から,提案手法によって得た意味ベクトルは,関係抽出

#### に有用である事が分かる.

なお,表2には,比較のため,state-of-the-art な手法である,RNN [Socher 11],MV-RNN [Socher 12],DNN [Zeng 14] の性能も示した.提案手法で計算した意味ベクトルを用いた手法はRNN [Socher 11] よりも良い性能である一方,MV-RNN [Socher 12] およびDNN [Zeng 14] には及んでいないしかしながら,MV-RNN は統語構造を,DNN はWordNetを利用するなど,様々な情報を利用している.提案手法は外部からの知識リソースを用いず,また,SemEval-2010のタスクを解くために意味ベクトルを学習している訳ではないにも関わらず,state-of-the-art な手法に近い性能を達成している.

#### 4. おわりに

本研究では,機能的な表現を行列とし,RNNを利用する事で,関係パタンの意味を構成的に計算する著者らの手法に, $L_1$  正則化を導入した.関係パタン間の類似度計算タスクおよび関係抽出タスクでの実験を通して,手法の有効性を明らかにした.本研究では関係パタンや機能的な表現の判定は非常に簡単なヒューリスティクを用いて行っている.今後は,機能的な表現の計算もモデルに含めた手法を考えたい.また,関係を表す複数の単語列の認識,すなわち,関係パタンの認識とその意味計算を同時に行う手法を考えたい.

謝辞 本研究は,文部科学省科研費課題 26・5820 および課題 23240018 の一環として行われた.また JST 戦略的創造研究推進事業「さきがけ」から部分的な支援を受けて行われた.

## 参考文献

[Baroni 10] Baroni, M. and Zamparelli, R.: Nouns Are Vectors, Adjectives Are Matrices: Representing Adjective-noun Constructions in Semantic Space, in EMNLP 2010, pp. 1183–1193 (2010)

[Fader 11] Fader, A., Soderland, S., and Etzioni, O.: Identifying Relations for Open Information Extraction, in EMNLP 2011, pp. 1535–1545 (2011)

[Hashimoto 12] Hashimoto, C., Torisawa, K., De Saeger, S., Oh, J.-H., and Kazama, J.: Excitatory or Inhibitory: A New Semantic Orientation Extracts Contradiction and Causality from the Web, in EMNLP 2012, pp. 619–630 (2012)

[Hendrickx 10] Hendrickx, I., Kim, S. N., Kozareva, Z., Nakov, P., Ó Séaghdha, D., Padó, S., Pennacchiotti, M., Romano, L., and Szpakowicz, S.: SemEval-2010 Task 8: Multi-Way Classification of Semantic Relations between Pairs of Nominals, in SemEval 2010, pp. 33–38 (2010)

[Mikolov 13] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J.: Distributed Representations of Words and Phrases and their Compositionality, in NIPS 2013, pp. 3111–3119 (2013)

[Mitchell 10] Mitchell, J. and Lapata, M.: Composition in Distributional Models of Semantics, Cognitive Science, Vol. 34, No. 8, pp. 1388–1439 (2010)

[Nakashole 12] Nakashole, N., Weikum, G., and Suchanek, F.: PATTY: A Taxonomy of Relational Patterns with Semantic Types, in EMNLP 2012, pp. 1135–1145 (2012)

[Socher 11] Socher, R., Lin, C. C.-Y., Ng, A. Y., and Manning, C. D.: Parsing Natural Scenes and Natural Language with Recursive Neural Networks, in ICML 2011, pp. 129–136 (2011)

[Socher 12] Socher, R., Huval, B., Manning, C. D., and Ng, A. Y.: Semantic Compositionality through Recursive Matrix-Vector Spaces, in EMNLP 2012, pp. 1201–1211 (2012)

[高瀬 15] 高瀬 翔, 岡崎 直観, 乾 健太郎: 構成性に基づく関係パタンの意味計算, 言語処理学会 第 21 回年次大会, pp. 640-643 (2015)

[Zeichner 12] Zeichner, N., Berant, J., and Dagan, I.: Crowdsourcing Inference-Rule Evaluation, in ACL 2012, pp. 156–160 (2012)

[Zeng 14] Zeng, D., Liu, K., Lai, S., Zhou, G., and Zhao, J.: Relation Classification via Convolutional Deep Neural Network, in COLING 2014, pp. 2335–2344 (2014)