

分散表現に基づく選択選好モデルの文脈化

大野 雅之^{1,a)} 井之上 直也^{1,b)} 松林 優一郎^{1,c)} 岡崎 直観^{1,d)} 乾 健太郎^{1,e)}

概要：述語の選択選好性のモデル化は、述語項構造解析・省略解析を始めとした意味解析において重要な基盤技術の一つである。これまでの研究では、「述語の選択選好性は項となる名詞の意味的な性質にのみ依拠する」という仮定して選好性を学習してきた (e.g., man は arrest の目的語になりうる)。省略解析などの談話解析への応用を考えると、談話内での名詞の言及のされ方まで含めて選好性を計算できることが望ましい (e.g., 悪事を犯した man は arrest の目的語になりうるが、善良な man はなりにくい)。そこで本研究では、ニューラルネットワークに基づく選択選好モデル [Van de Cruys 2014] を拡張し、名詞の意味的な性質に加え、談話内での言及のされ方を分散表現で表現することにより、名詞の出現文脈を考慮した述語の選択選好モデルを提案する。評価実験では、代名詞照応解析への応用を見据え、代名詞に対する先行詞候補のランキング問題に基づく評価を行い、名詞の出現文脈を用いることの有効性を確認した。

キーワード：談話解析，照応・省略解析，文脈，選択選好，分散表現

1. はじめに

述語の選択選好性とは、ある述語のある格が取りうる名詞句に意味的な偏りがあることをいう。例えば、「食べる」という述語の目的格は、「りんご」のように食べ物を意味する名詞句を取ることがほとんどであり、「時計」のように食べ物ではないものを取ることは極めて稀である。自然言語処理の分野において、述語の選択選好性の知識は述語項構造解析や省略解析を始めとした意味解析に有効な情報とされ、これを利用した多くの研究がなされてきた [1], [2], [3]。

述語の選択選好性は述語によってそれぞれ異なるため、人手によりその性質を全てを書き下すのは現実的でない。このため先行研究では、大規模コーパスから、述語と、その述語の格を埋める名詞（以後、項と呼ぶ）の用例を獲得し、この共起の統計値から選択選好のモデルを構築するアプローチが採られてきた。また、共起用例をそのまま保持するだけでは、出現頻度の低い述語に関して選好性の知識を正しく得られないという問題が起こるため（データスパースネス問題）、シソーラス [4] や単語の分散表現化 [5]、用例のクラスタリング [6], [7], [8], [9]、類似性判定 [10], [11] といった方法で、獲得した述語と項の共起用例をいかに一

般化するか、という点でさまざまな工夫がなされてきた。

一方で、選択選好性が述語項構造解析や照応解析の際に用いられる知識であることを踏まえて現実的な実用上の状況と考えた場合、実際にはそれぞれの名詞には文章上で様々な言及がなされる。選択選好性によって項としての尤もらしさを判定する際には、単にその名詞の意味的性質だけでなく、その名詞に対して文脈上でどのような言明がなされてきたかを踏まえてその意味的性質を考えるほうが自然である。例えば、次のような照応の問題を解いているという状況を考える。

(1) [John_(i)] attacked [Bob_(j)]. Police arrested him_(i).

この例において、先行詞の候補は John_(i) と Bob_(j) の二種類であると仮定すると、

“attack” された人より “attack” した人のほうが
“arrest” される人として尤もらしい

という選好性により John が him の先行詞であると判断される。しかし、これまでの選択選好モデルでは、述語の選択選好性が名詞自身との関係の性質だけから決まると仮定しており、上のような文脈の選好性は無視されてきた。これは、(1) の例では

(2) Police arrested [John].

(3) Police arrested [Bob].

の二つの尤もらしさを比較していることになるが、文脈なしにはこれらの項としての尤もらしさに大きな差は現れな

¹ 東北大学大学院 情報科学研究科
Tohoku University

a) masayuki.ono@ecei.tohoku.ac.jp

b) naoya-i@ecei.tohoku.ac.jp

c) y-matsu@ecei.tohoku.ac.jp

d) okazaki@ecei.tohoku.ac.jp

e) inui@ecei.tohoku.ac.jp

い。一方、次のように前方文脈により *John* や *Bob* に付加されている情報を加味して *arrest* の目的語としての選好性を考慮できれば、正しい解を導き出せるはずである。

- (4) Police arrested [*John, who attacked Bob*].
- (5) Police arrested [*Bob, whom John attacked*].

以上のような背景から、本研究では、項自身の意味の性質に加えて、項が談話内で言及されてきた文脈情報を用いて述語の選択選好性を計算するモデルを提案する。先に述べたデータスパースネス問題に対応するために、分散表現に基づく Van de Cruys [5] のモデルをベースとして、項の意味の性質を先行文脈から構成的に計算し、述語の選択選好性を計算する手法を提案する。2 節で述べるように、分散表現に基づいて項の出現文脈をモデル化し、述語の選択選好性を計算するモデルは、我々の知る限り初めての試みである。

本稿では、まず述語の選択選好性の関連研究について述べたあと(2 節)、分散表現に基づいて項の周辺文脈から項の意味計算を行う手法について説明する(3 節)。評価実験では(4 節)、提案モデルが項の先行文脈を考慮して項の意味の性質を適切に計算できていることを確認するために、代名詞照応解析問題に基づく評価・分析を行った結果を報告する。

2. 関連研究

本稿の冒頭で述べたように、述語の選択選好性モデルの研究は、大規模コーパスから自動獲得した述語と項の共起用例を、何らかの方法で一般化するアプローチが主流である。Resnik [4] は人手で作成した名詞クラスター (WordNet [12] の synset) に基づいて項を一般化し、確率モデルを構築している。Rooth ら [6] は EM クラスタリングアルゴリズムによる確率的潜在変数モデルを提案している。Rooth らのモデルは、潜在変数を用いることで訓練コーパスに出現しなかった述語と項の組に関する選択選好知識の獲得を可能にしている。確率的潜在変数モデルは他に、Latent Dirichlet Allocation (LDA) を用いた手法 [7], [8] が提案されている。Kawahara ら [9], [13] は、Chinese Restaurant Process (CRP) に基づく共起用例のクラスタリングを行い、動詞の意味フレームを導出する手法を提案している。Erk ら [10], [11] は分布仮説に基づき、単語の共起情報の類似度を測ることで未知の述語と項の組に関する選好性を計算している。Van de Cruys [5] は、データスパースネス問題を解消する手段として、Mikolov ら [14] の研究に代表される、分散表現を用いた一般化を行っている。データスパースネス問題に対する分散表現の有効性はさまざまなタスクで報告されており、データスパースネス問題に対する事実上標準的な手段となりつつある。Van de Cruys は、ニューラルネットワークを用いて、選択選好性の獲得と、

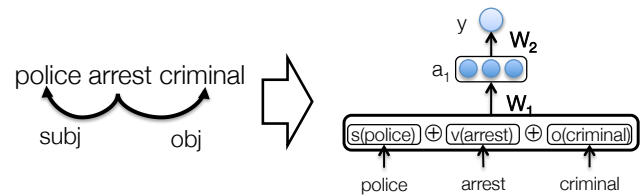


図 1 SVO モデルのネットワーク構造

述語と項の意味ベクトルの学習を同時に行う手法を提案しており、現状の選択選好性モデルとしては最高性能を示している。

最近では、一般化の技術と同時に、周辺文脈をどの程度見て述語と項の一般化を行うか、という点も重要な課題とされている。研究の初期段階では、「述語の意味は動詞のみで決まる」という前提が置かれていたが、これでは述語の主たる意味が項に依存する場合(「実施する」等の機能動詞など)に、選好性を正しく計算できない可能性があった(例えば、「セールを実施する」の主語には店が入る一方で、「国勢調査を実施する」の主語には政府機関が入る)。先に述べたモデルのうち、Ritter ら、Van de Cruys らのモデル [5], [8], [15] は、このような問題に対応するため、主語・動詞、または動詞・目的語の 2 つ組を述語とみなし、選択選好性知識の一般化を行っている。しかしながら、1 節で述べた、the police arrested him の him が John か Bob かという問題にはやはり無力である。

我々の提案モデルは、項の周辺文脈をさらに広く見るという点で、先行研究と異なる。Van de Cruys の分散表現ベースのモデルを拡張することにより、項の周辺文脈の多様性をうまく捉えて、述語の選択選好性をモデル化できると考えられる。

また、冒頭で述べたように、照応解析の分野では、Narrative Chain [16] に代表される事象間関係知識 (X bankrupt → acquire X) [17], [18] を利用し、先行詞候補と照応詞の参与している事象の間の関連性を計算し、それを照応解析の手がかりとすることで、照応解析の性能が向上することが報告されている [2], [3], [19]。これらの研究で用いられる事象間関係知識においては、項は離散的な意味クラスで表現され、述語の汎化は行われていない。また、解析器側でも、述語の完全一致により事象間の関連性を計算するため、データスパースネス問題がより顕著であると予想される。一方で、我々の枠組みは、述語と項を分散表現に基づいて表現するため、未知のテスト事例に対してより頑健であることが期待される。分散表現を用いて事象間関係知識を獲得する手法として Modi ら [20] の研究があるが、事象間の順序関係を特定するモデルであり、事象間関係の有無の識別は行っていないため、本研究とは方向性が異なる。

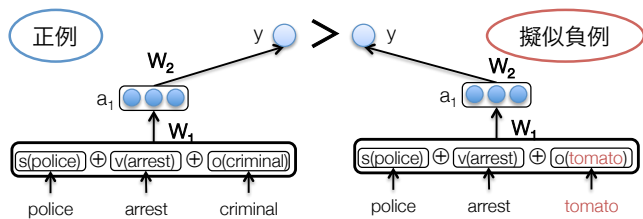


図 2 SVO モデルの学習 (目的語 $w_o = \text{criminal}$ を $\tilde{w}_o = \text{tomato}$ に置き換えた例)

3. Van de Cruys の SVO モデル

本研究では、単語分散表現とニューラルネットワークを用いた選択嗜好モデルである Van de Cruys の SVO モデル (以下 SVO モデル) を拡張する形で文脈情報の埋め込みをはかる。このため、本節ではこの SVO モデルについて簡単に紹介する。

SVO モデルは 1 層の隠れ層を持つフィードフォワード型ニューラルネットワークで構成されており、(主語 w_s , 動詞 w_v , 目的語 w_o) の三つ組に対し、それぞれに対応する単語ベクトル $s(w_s), v(w_v), o(w_o) \in \mathbb{R}^d$ を入力とし、*1 入力の組合せが選択嗜好を満たすか否かを識別するスコアを出力する。例えば、(police, arrest, criminal) という三つ組に対しては、図 1 のように対応する単語ベクトルから選択嗜好のスコアを計算する。

隠れ層 $a_1 \in \mathbb{R}^h$ と出力スコア $g[(w_s, w_v, w_o)]$ は式 (1), (2) によって計算する。

$$g[(w_s, w_v, w_o)] = \mathbf{W}_2 \mathbf{a}_1(w_s, w_v, w_o) \quad (1)$$

$$\mathbf{a}_1(w_s, w_v, w_o) = f(\mathbf{W}_1(s(w_s) \oplus v(w_v) \oplus o(w_o)) + b_1) \quad (2)$$

ここで、 \oplus はベクトルの連結 (concat) であり、 $\mathbf{W}_1 \in \mathbb{R}^{h \times 3d}$, $\mathbf{W}_2 \in \mathbb{R}^{1 \times h}$ は各層の重み行列、 b_1 はバイアス項である。活性化関数 $f(\cdot)$ には \tanh を用い、出力値 $g[(w_s, w_v, w_o)]$ が選択嗜好のスコアとなる。

SVO モデルの学習は、Collobert ら [21] が提案した教師なし学習を用いる。図 2 に例を示す。まず、訓練コーパスから得られた (主語, 動詞, 目的語) の組を正例とし、この組に対し、単語の出現頻度分布に基づき (1) 「主語」、(2) 「目的語」、(3) 「主語と目的語の両方」を別の単語に置き換えた擬似負例をそれぞれ作成する。図 2 は目的語 w_o を \tilde{w}_o に置き換えている例である。次に、これらの正例と擬似負例についての選択嗜好スコアを式 (1) を用いて計算し、式 (3) を最小化するように、モデルパラメータ (単語ベクトル s, v, o と重み行列 $\mathbf{W}_1, \mathbf{W}_2$, バイアス項 b_1) を学習する。

*1 Van de Cruys のモデルでは、単語 w が主語、動詞、目的語のいずれとして出てくるかによってそれぞれ異なる単語ベクトル $s(w), v(w), o(w)$ となるように学習される。

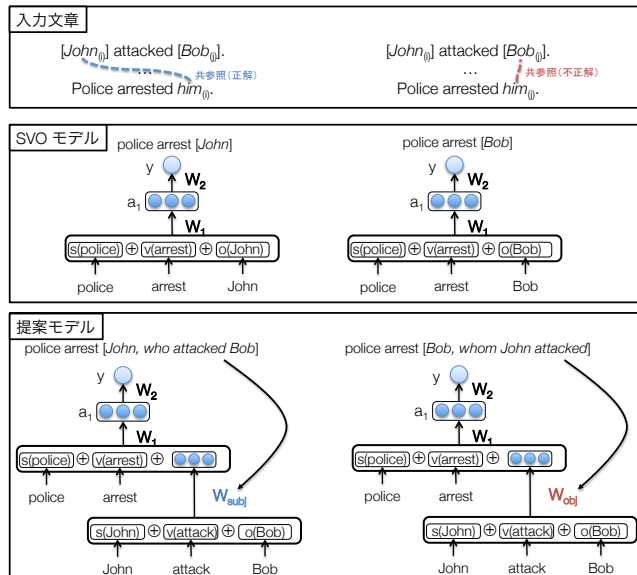


図 3 提案手法の概要

$$\begin{aligned} & \max(0, 1 - g[(w_s, w_v, w_o)] + g[(\tilde{w}_s, w_v, w_o)]) \\ & + \max(0, 1 - g[(w_s, w_v, w_o)] + g[(w_s, w_v, \tilde{w}_o)]) \\ & + \max(0, 1 - g[(w_s, w_v, w_o)] + g[(\tilde{w}_s, w_v, \tilde{w}_o)]) \quad (3) \end{aligned}$$

ここで、 \tilde{w}_s, \tilde{w}_o はそれぞれランダムにサンプルし置き換えた主語と目的語である。つまり、コーパスから得られた三つ組に対するスコアと、ランダムな置換により不自然になった擬似負例に対するスコアの差が 1 以上になるようにモデルの最適化を行う。

4. 文脈を考慮した選択嗜好モデル

1 節で述べたとおり、我々は談話内での言及による名詞の意味的性質の変化を考慮した選択嗜好モデルを設計する。今、例として 1 節の例文 (1) と同様の問題を解いていると仮定する。我々の目的とする文脈の埋め込みは、下の例文のように、前方文脈の代わりに名詞が関係節等によって修飾されている文を考えて、この修飾関係を表現した分散表現ベクトルを SVO モデルにおける目的語ベクトルの部分に埋め込んで嗜好性を評価することだと考えると直感的に分かりやすい。

$$(6) [John_{(i)}] attacked [Bob_{(j)}]. Police arrested him_{(i)}.$$

$$(7) Police arrested [John, who attacked Bob].$$

$$(8) Police arrested [Bob, whom John attacked].$$

嗜好性に関わる性質に変化を与える修飾関係としては述語表現、副詞表現、同格、A of B など様々な要素が考えられるが、本稿では、項の周辺文脈のモデル化への第一歩として、事象間関係を応用した照応解析の先行研究 [2], [3], [19] に倣い、選択嗜好性の判定対象となる名詞が項として参加している (主語, 動詞, 目的語) の三つ組の述語-項関係を文

脈として採用する．ただし，ただし，対象名詞がその三組において，主語となっているか目的語になっているかでその名詞が帯びる意味が違っているので，これを区別する．上記の例であれば，例文 (7) の John は，(John, arrest, Bob) の主語としての John であり，例文 (8) の Bob は，(John, arrest, Bob) の目的語としての Bob であるという情報を表現する．また，対象の名詞の文脈としては，一般に文章中にはその名詞を含む複数の言明が書かれていると考えられ，意味の変化はこれらの重ね合わせとして計算されることが自然だと考えられるが，今回は後述の方法で最も選好性の高くなる述語項関係の一つを選択し，この述語項関係が対象の名詞の性質に影響を与える文脈であるとする．

このような仮定の結果，我々の考える文脈の埋め込みは図 3 のようなネットワークで表現される．提案手法では，3 節で説明した Van de Cruys [5] の SVO モデルにおいて，主語ベクトルと目的語ベクトルの部分に，それぞれ単語ベクトルと述語項関係を表現した文脈ベクトルのいずれかを選択できるように拡張する．このようにすることで，対象の名詞に対して前方文脈として述語による言明が存在する場合は文脈ベクトルを用い，文脈がない場合は単語ベクトルを用いることができるようにする．

述語項構造を表現する文脈ベクトルは，文脈情報の意味計算を構成要素から計算する方法 [14], ? に習い，構成性に基いてモデル化を行う．

ここで，いくつか記号の定義を行う．まず，選択選好性の計算対象となる名詞 a に関する先行文脈を， $c_a = \langle c_{a,s}, c_{a,v}, c_{a,o} \rangle$ と書く． c_a は，主語 $c_{a,s}$ ・ 動詞 $c_{a,v}$ ・ 目的語 $c_{a,o}$ からなる述語項構造である．また， c_a は a に関する言及であるため， $c_{a,s} = a$ ，または $c_{a,o} = a$ であり， a が埋め込まれた格関係を $c_{a,c} \in \{subj, obj\}$ と書く．例として，以下の談話を用いて説明する．

(9) A man attacked a woman. (...) A police arrested him.

いま，arrest を選択選好性の計算対象とし，候補 $a=man$ についての適合度合いを計算したいとする．このとき，man の先行文脈 c_a は，A man attacked a woman の 1 つだけ存在し， $c_{a,s} = man$ ， $c_{a,v} = attack$ ， $c_{a,o} = woman$ ， $c_{a,c} = subj$ である．

本研究では，SVO モデルにおける項の入力ベクトル $s(w_s)$ ， $o(w_o)$ を先行文脈を考慮して計算した意味ベクトル $s_{cmp}(w_s, c_s)$ ， $o_{cmp}(w_o, c_o)$ に置き換えた以下の式で選好性を計算する．

$$g[(w_s, w_v, w_o, c_s, c_o)] = \mathbf{W}_2 \mathbf{a}_1(w_s, w_v, w_o, c_s, c_o) \quad (4)$$

$$\mathbf{a}_1(w_s, w_v, w_o, c_s, c_o) =$$

$$f(\mathbf{W}_1(\mathbf{S}_{cmp}(w_s, c_s) \oplus \mathbf{v}(w_v) \oplus \mathbf{O}_{cmp}(w_o, c_o)) + b_1) \quad (5)$$

先行文脈をもとにした意味計算では，先行文脈における項の埋め込み格 $c_{a,c}$ が意味計算に大きく影響するため，これらを区別した合成が必要となる．例えば，“a man attacks a woman” と “a woman attacks a man” という二つの先行文脈では，それぞれ *man* の意味は大きく変わるはずである．そこで，意味ベクトルの生成関数 s_{cmp} を，単に先行文脈の各要素の意味ベクトルの和ではなく，次のように定義する．

$$\mathbf{S}_{cmp}(w_s, c_s) = \begin{cases} f(\mathbf{W}_{subj}(\mathbf{s}(c_{s,s}) \oplus \mathbf{v}(c_{s,v}) \oplus \mathbf{o}(c_{s,o}))) & \text{if } c_{s,c} = \text{subj}; \\ f(\mathbf{W}_{obj}(\mathbf{s}(c_{s,s}) \oplus \mathbf{v}(c_{s,v}) \oplus \mathbf{o}(c_{s,o}))) & \text{if } c_{s,c} = \text{obj}; \\ s(w_s) & \text{otherwise;} \end{cases} \quad (6)$$

つまり，先行文脈の述語 $c_{s,v}$ と，その項 $c_{s,s}, c_{s,o}$ の各ベクトルに対し，先行文脈における項の埋め込み格が *subj* の場合（つまり，a man attacks），行列 \mathbf{W}_{subj} を用いてベクトルの変換を行い，項の埋め込み格が *obj* の場合（つまり，attack a man），行列 \mathbf{W}_{obj} を用いてベクトルの変換を行う．このように，埋め込まれた位置によって計算に用いる行列を変えることで，“a man attacked a woman” と “a woman attacked a man” という，二つの先行文脈の違いを考慮して意味ベクトルを計算することが可能になる．

目的語の意味ベクトルの生成関数 o_{cmp} についても，同様に定義する．

$$\mathbf{O}_{cmp}(w_o, c_o) = \begin{cases} f(\mathbf{W}_{subj}(\mathbf{s}(c_{o,s}) \oplus \mathbf{v}(c_{o,v}) \oplus \mathbf{o}(c_{o,o}))) & \text{if } c_{o,c} = \text{subj}; \\ f(\mathbf{W}_{obj}(\mathbf{s}(c_{o,s}) \oplus \mathbf{v}(c_{o,v}) \oplus \mathbf{o}(c_{o,o}))) & \text{if } c_{o,c} = \text{obj}; \\ o(w_o) & \text{otherwise;} \end{cases} \quad (7)$$

構成性に基づくベクトル合成手法は多く提案されているが，それらの多くが動詞句や文の意味ベクトルを合成することを目的としており，本研究が表現したい “John attacked Bob” から “John who attacked Bob” の意味を持つベクトルを合成には適用できない．出現した文脈に基づいた単語ベクトルの合成手法として， J_i らは Recursive Neural Network (RNN) をベースとした手法（以後，Entity Vector と呼ぶ）を提案している [22]．Entity Vector は RNN で再帰的に単語ベクトルを組み上げ文ベクトルを計算し，文ベクトルを再帰的に各単語へ逆伝搬させることで，各単語ベクトルに単語が出現した文の意味を付加している． J_i らは Entity Vector を談話関係認識の素性のひとつとして用いており，Entity Vector が出現文脈の情報を表現できているかの評価はしていない．そのため本研究では，提案モデルの他に Entity Vector によって項の先行文脈の意味ベクトルを合成したモデルの評価も行なう．

4.1 学習

パラメータの学習は、SVO 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。

- 学習するのは、単語ベクトル s, v, o と 重み行列 W_1, W_2 、バイアス項 b_1 に加えて、先行文脈の意味ベクトルの計算に用いる重み行列 W_{subj}, W_{obj} である。
- 学習には、SVO モデルの学習に用いた主語-動詞-目的語の三つ組と共参照関係から作成した文脈付き事例の両方を用いる。
- 先行文脈の意味ベクトルの計算部分で、活性化関数 \tanh を適用しているため、先行文脈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 SVO モデルでは単語の意味ベクトルに活性化関数を適用していないため、値域の制限はない。本手法では、先行文脈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 s, v, o にも活性化関数 \tanh を適用する。

項の先行文脈を考慮した学習を行うために、訓練事例を次の手順で作成する。

まず、訓練コーパスに対して構文解析・共参照解析を行い、Type A: 〈主語, 動詞, 目的語〉, Type B1: 〈主語, 動詞, 目的語, 主語の先行文脈〉, および Type B2: 〈主語, 動詞, 目的語, 目的語の先行文脈〉の共起事例を獲得し、正例とする。例えば、下記のような談話を考える。

(10) [*The old man_(i)*] attacked [*a boy_(j)*]. A policeman arrested *the man_(i)*.

ここで、*the man_(i)* と *The old man_(i)* は共参照関係である。この場合、Type A の訓練事例として 〈policeman, arrest, man〉を、Type B2 の訓練事例として 〈policeman, arrest, man, 〈man, attack, boy, subj〉〉を獲得する。

本研究では、訓練コーパスとして、大規模 Web コーパス ClueWeb12^{*2} の一部 (約 2.2 億文書, 6.9 億文) を用い、Stanford CoreNLP [23]^{*3} を用いて構文解析及び共参照解析を行ったのち、Type A, Type B1, Type B2 の 3 種類の訓練事例を獲得した。共参照解析結果に関しては、可能な限り正確な自動解析結果を用いるため、(1) 同一文内で、かつ (2) 表層形が一致する単語間の共参照関係のみを先行文脈の同定に用いた。先行文脈の同定精度を確認するため、共参照解析の結果を 100 事例サンプリングし、人手により評価したところ、87.0 % の精度であることがわかった。

獲得結果からノイズとなりうる共起事例を取り除くため、いくつかのフィルタリング処理を行った。(1) に関しては、品詞付きの単語出現頻度上位 5 万単語を含み、かつ共起頻度 20 以上の共起事例を得た。学習時間の都合上、20 分の 1 をサンプリングし、4,177,864 事例 (1,282,994 種類)

を学習に用いた。(2) に関しては、(1) のフィルタリング後の共起事例に含まれる単語で表現可能な 1,140,266 事例を学習に用いた。

また、本研究では、上記の手続きにより獲得された正例に対して、二種類の方法により擬似負例を生成する。

(1) ランダム負例 3 節の擬似負例の生成手続きに従う。例えば、談話 (10) では、Type A の正例 〈policeman, arrest, man〉に対して、負例 〈policeman, arrest, book〉が、Type B2 の正例 〈policeman, arrest, man, 〈man, attack, boy, subj〉〉に対して、負例 〈policeman, arrest, banana, 〈monkey, eat, banana, obj〉〉が生成される。

(2) 重み入れ替え負例 Type A の正例に対する負例は 3 節の擬似負例の生成手続きに従う。Type B1, Type B2 の正例については、先行文脈における文脈が付与された単語の格関係を入れ替えることにより負例を生成する。例えば、談話 (10) では、正例 〈policeman, arrest, man, 〈man, attack, boy, subj〉〉に対して、〈policeman, arrest, man, 〈man, attack, boy, obj〉〉を負例として用いる。ランダムに負例を生成する場合に比べて、本手続きでは、より高い精度で負例を得ることができると考えられる。

また、学習の目的関数についても、下記のようなバリエーションを考える。3 節では、正例と擬似負例の間の相対的な自然さを学習していたが、選択選好モデルとしては、入力の組合せが自然か否かの 2 値分類が行えるような絶対的な自然さを学習することも考えられる。このため、式 (8) の損失関数を最小化するような、ロジスティック回帰的な学習も行う。

$$-\log \sigma(g[(w_s, w_v, w_o)]) - \log(1 - \sigma(g[(\tilde{w}_s, w_v, w_o)])) \\ - \log(1 - \sigma(g[(w_s, w_v, \tilde{w}_o)])) - \log(1 - \sigma(g[(\tilde{w}_s, w_v, \tilde{w}_o)])) \quad (8)$$

ここで、 σ はシグモイド関数である。 \tilde{w}_s, \tilde{w}_o はそれぞれ、前述の負例生成手続きに従って生成された主語、目的語の負例である。

5. 評価実験

本評価実験では、提案モデルにより項の先行文脈の情報を反映した述語の選択選好性の計算がどの程度正しく行えたかを確認する。本評価実験では、照応解析の先行研究 [2], [3], [19] から得られた知見より、「代名詞が参与している事象と先行詞が参与している事象には関連がある」ことを利用し、先行文脈を見ることの効果を照応解析性能の向上の度合いで評価する。

5.1 タスク

文書中に出現した代名詞に対して先行詞候補が複数提示されている状況で、先行詞候補の集合を先行詞らしさの順に順序付けするランキング問題を考える。評価指標に

^{*2} <http://lemurproject.org/clueweb12/>

^{*3} <http://nlp.stanford.edu/software/corenlp.shtml>

は、一般的にランキングモデルの評価で用いられる Mean Reciprocal Rank (MRR) を用いる。より具体的には、下記の式でスコアを計算する。

$$\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i} \quad (9)$$

ここで、 $|Q|$ は問題総数、 $\text{rank}(i)$ は i 番目の問題において、式 (4) によって計算された正解先行詞の順位である。

なお、述語の選択選好性モデルの先行研究では、*pseudo-disambiguation test* と呼ばれるタスクによる評価が一般的である。*pseudo-disambiguation test* では、コーパス中に出現した〈主語、動詞、目的語〉を正例とし、これのうちの一つの要素をランダムに入れ替えた事例を負例とした、二値分類タスクである。本稿では、より現実的な問題の上で、選択選好性モデルの文脈化の効果を評価するため、代名詞照応解析の問題上で評価を行う。詞照応解析の先行研究では、照応解析性能の評価指標として CEAF, BLANC などのさまざまな指標が提案されているが [24], [25], etc., 本研究の評価は純粋に選択選好性モデルの評価であるため、ランキングレベルの評価に留める。今後の課題として、提案モデルの結果を既存の照応解析器の解析手がかりの一部として用い、既存研究との比較を行う予定である。

評価データには、OntoNotes 4.0 [26] を用いた。OntoNotes は、新聞記事、放送原稿、Web ページなど複数ジャンルからなるコーパスで、品詞タグ、構文木、述語項構造、代名詞・名詞句間の共参照関係、固有表現などが付与されている。

以下では、下記の談話を用いて、実験対象となる代名詞および先行詞候補の抽出方法に関して説明する。

- (11) In $his_{(p)}$ 40-minute $speech_{(i)}$, $Chen_{(j)}$ declared the $determination_{(k)}$ of the $people_{(l)}$ of $Taiwan_{(m)}$ to continue forward... (中略)... $Chen_{(n)}$ visited...(中略), and $he_{(p')}$ stated ... (ectb.1025 改)

評価対象の代名詞として、提案モデルにより先行詞のランク付けが行える問題、すなわち他動詞の主語もしくは目的語となっている代名詞を対象とした ($he_{p'}$ は対象であるが、 $his_{p'}$ は対象でない)。また、下記の基準により先行詞候補を抽出し、抽出結果の中に正解の先行詞が存在する問題を評価の対象とした。

- 先行詞候補として、(1) 代名詞の出現した文より前 3 文以内に出現した名詞句の主辞、(2) 品詞が名詞である (代名詞は含まない)、かつ (3) 主辞が訓練事例に出現したものを、を対象とした。例えば、 $he_{(p')}$ に対して、 $his_{(p)}$ 以外のすべての名詞 i, j, k, l, m が先行詞候補となる。
- 先行詞候補集合の中に共参照関係にある複数の先行詞候補が存在する場合、それらを別々の先行詞候補とし

表 1 ベースラインとの比較

設定	MRR
SVO モデル (ベースライン)	0.3866
重み付き和+ランダム負例	0.3968
重み付き和+重み入れ替え負例	0.4186
重み付き和+ランダム負例+回帰	0.3932
重み付き和+重み入れ替え負例+回帰	0.4281
Entity Vector+ランダム負例	0.4001
Entity Vector+重み入れ替え負例	0.4092
Entity Vector+ランダム負例+回帰	0.3788 ^{*4}
Entity Vector+重み入れ替え負例+回帰	0.4179

て扱った。例えば、 $Chen_{(j)}$ と $Chen_{(n)}$ を別々の先行詞候補として順位付けを行う。MRR における評価の際は、正解の先行詞候補に付けられた順位のうち、最も高い順位を $rank$ として評価した

- 先行詞候補の先行文脈を抽出する際、訓練データに出現しなかった単語を含む先行文脈は、「先行文脈なし」とみなして処理した。例えば、 $Chen_{(j)}$ の先行文脈は $c_j = \langle \text{Chen, declared, determination, subj} \rangle$ であるが、少なくとも 1 つの単語が訓練事例の中に現れなかった場合、 $c_j = \phi$ とする

最終的には、OntoNotes コーパスに出現する代名詞のうち、4,021 問の代名詞を評価対象とした。

5.2 実験設定

単語ベクトルの次元数は $d = 50$ に設定し、モデルパラメータは全て乱数で初期化した。最適化には初期学習率を 0.1 とした AdaGrad [27] を用い、1,000 事例ごとのミニバッチ学習を、学習データ全体に対するイテレーション回数を 20 回として行った。

また、先行文脈を見ることの効果を見積もるため、ベースラインモデルを構築した。より具体的には、訓練事例のうち〈主語、動詞、目的語〉のみを与えて訓練を行い、先行詞候補のランキング時に項の先行文脈を一切与えないモデルを構築した。このベースラインモデルは、3 節で説明した Van de Cruys ら [5] の SVO モデルに相当する。

また、4 節で提示した (1) 2 つの擬似負例の生成手法 (ランダム負例、重み入れ替え負例)、(2) 2 種類の損失関数 (マージン最大化、回帰)、(3) 2 種類の項の先行文脈の合成方法 (重み付き和、Entity Vector) のそれぞれを組み合わせた 8 個のモデルを構築し、性能を測定した。

5.3 実験結果

先行詞候補ランキング問題の MRR に基づく評価結果を表 1 に示す。表 1 より、提案モデルは一つの設定を除くすべての設定で SVO モデルよりも高い性能を達成しており^{*4}、項の先行文脈の情報を反映した選択選好性の計算が

^{*4} 一部の設定ではベースラインを下回ったが、Wilcoxon 検定で有

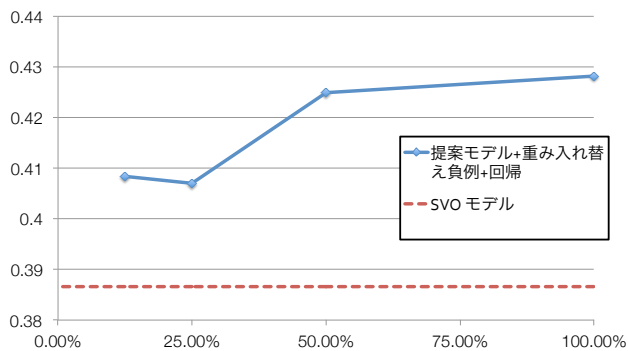


図 4 訓練に用いる共参照事例数の変化による学習曲線

行えていることが確認できた。

設定のバリエーションについては、「重み付き和 + 重み入れ替え負例 + 回帰」の設定が最も高い性能を示した。以下では、この原因について考察する。

第一に、重み入れ替え負例による擬似負例の生成により、ランダムに単語・文脈付き単語をサンプルするよりも、精度よく不自然な負例を作り出すことができたためと考えられる。

第二に、回帰に基づく損失関数によって、述語の選択選好性を絶対的な基準で学習できたためと考えられる。マージン最大化に基づく損失関数は、各学習事例毎に、正例と擬似負例の間のマージンを最大化するという相対的な学習となっており、対となって生成されなかった擬似負例については低いスコアが出ることは保証されない。例えば、people eat book は明らかに述語 eat の選択選好性を満たしていないが、マージン最大化に基づく損失関数では、これと対となって出てきた正例（例えば、people eat food）とのマージンの確保のみを考える。一方、回帰に基づく損失関数では、訓練事例毎に正例と擬似負例のスコアの差を最大化するのではなく、正例は ∞ 、負例は $-\infty$ に近づくよう、絶対的なスコア付けを学習する。このため、学習事例に対となって生成されなかった擬似負例に対して低いスコアを付与することができ、高い性能を達成したと考えられる。

5.4 分析

学習に用いる文脈情報付き事例の量と、提案モデルの照応解析性能の関係について調査した。図 4 は、学習に用いた文脈付き訓練事例の量を 1/2 ずつ変化させた際の「提案モデル + 重み付き和 + 回帰」の MRR を表している。^{*5} 図 4 より、使用する文脈付き訓練事例の量を増やすことによるモデルの性能向上を確認できたが、訓練事例量 50% から 100% への変化では、性能の向上率は限定的であるこ

意差を確認出来ない程度であった。

^{*5} SVO モデルについては、学習事例を 26,500,536 事例(7,323,250 種類)に増やして解析性能を計測したが、現行の MRR (0.3866) からの向上は微小であった。

表 2 次元数と MRR

設定	25 次元	50 次元	100 次元
baseline (SVO モデル)	0.3915	0.3866	0.3751
提案モデル+重み入れ替え負例+回帰	0.4195	0.4281	0.4215

とがわかった。このことより、モデルの性能をさらに向上させるためには、提案モデルの入力形式を拡張し、述語の主語・目的語以外の項（例えば、in restaurant などの前置詞句）や修飾句（incorrectly などの副詞、否定詞など）など、本稿で捨象している情報を扱える枠組みを構築する必要があると考えられる。

次に、単語ベクトルの次元数を変化させた場合に性能がどう変化するかを調査した。結果を表 2 に示す。表 2 より、性能は単語ベクトルの次元数にはあまり依存しないことがわかった。

提案モデルの振る舞いを確認するために、提案モデルにより正解先行詞の順位が改善された 10 事例を、人手により分析した。改善が確認された事例の中には、項の先行文脈で言及された事例と、代名詞を項に持つ述語の間に関連性があり、提案モデルによりそれらの関連を適切に捉えることができたと考えられる事例が存在した。以下に、その一例を示す。

- (12) All these were the highest levels in history. Since being put into operation five years ago, *the Tianjin Port Bonded Area_(i)* has completed the construction of China's first goods distribution center, functions like a customs port, opened up the special use the railway line from ... Moreover, *the bonded area_(j)* has implemented a series of preferential policies towards enterprises entering *the area_(k)*: *it_(pro)* has established a system of "no custom accounting" and established manuals and management... (chtb_0099)

ここで、解析対象となる代名詞は *it_(pro)* であり、正しい先行詞候補は *the Tianjin Port Bonded Area_(i)*, *the bonded area_(j)*, *the area_(k)* の 3 つである。SVO モデルによるランク付けでは、それぞれ 36 位, 38 位, 39 位であったが、提案モデルによるランク付けでは、それぞれ 1 位, 2 位, 34 位に改善されていた。SVO モデルでは、*it_(pro)* の述語である「確立する (establish)」の主語に *area* が入りやすいか、といった基準でランキングが行われるのに対して、提案モデルでは、「工事を終えた X (X which completed construction) (先行詞 *i* の場合)」「一連の何かを施行した X (X which implemented series) (先行詞 *j* の場合)」が establish の主語になりやすいか、という尺度でランキングが行われ、提案モデルにより項の先行文脈と述語の選択選好性の関連を適切に捉えられたと考えられる。

6. おわりに

本稿では、述語の選択選好性モデルを談話解析に用いる場合に特に重要となる、項の先行文脈を考慮した述語の選択選好性計算モデルを提案した。述語の選択選好性モデルに関する先行研究では、これまで項自身の意味的性質に基づいて選好性の学習・計算が行われていたのに対し、本研究では、先行文脈における項に対する言及から、構成的に項の意味を組み上げる分散表現ベースの枠組みを提案した。提案モデルを先行詞候補のランキング問題により評価・分析した結果、提案した枠組みにより、正しい先行詞候補を上位にランキングできるようになることを確認した。

今後の課題として、提案モデルで扱うことのできる述語の周辺文脈と項の先行文脈の種類を増やすことが挙げられる。主語・動詞・目的語以外にも、形容詞・副詞・否定詞による修飾など、重要な周辺文脈要素は多数存在しており、これらに対応することで、さらにモデルの性能が向上することが期待される。また、本稿では、ただひとつの述語項構造を項の先行文脈として扱ったが、1節で述べたように、先行文脈における項に対する言及は一般的には複数存在する。このため、recurrent な分散表現モデルを用いて、先行文脈の複数の言及の内容や、言及の順序を反映した上で項の意味計算を行うような機構を構築していく予定である。

謝辞

本研究は JSPS 科研費 15H01702, 15H05318, 15K16045 の助成を受けたものである。また、本研究は、JST, CREST の支援を受けたものである。

参考文献

- [1] Bergsma, S.: Discriminative Learning of Selectional Preference from Unlabeled Text, *EMNLP*, pp. 59–68 (online), available from <http://www.aclweb.org/anthology/D08-1007> (2008).
- [2] Rahman, A. and Ng, V.: Resolving Complex Cases of Definite Pronouns: The Winograd Schema Challenge, *Proceedings of EMNLP-CoNLL*, pp. 777–789 (2012).
- [3] Peng, H., Khashabi, D. and Roth, D.: Solving Hard Coreference Problems, *NAACL*, pp. 809–819 (2015).
- [4] Resnik, P.: Selectional constraints: An information-theoretic model and its computational realization, *Cognition*, Vol. 61, No. 1, pp. 127–159 (1996).
- [5] Van de Cruys, T.: A neural network approach to selectional preference acquisition, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 26–35 (2014).
- [6] Rooth, M., Riezler, S., Prescher, D., Carroll, G. and Beil, F.: Inducing a Semantically Annotated Lexicon via EM-Based Clustering, *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, College Park, Maryland, USA, Association for Computational Linguistics, pp. 104–111 (1999).
- [7] Séaghdha, D. O.: Latent variable models of selectional preference, *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, pp. 435–444 (2010).
- [8] Ritter, A., Etzioni, O. et al.: A latent dirichlet allocation method for selectional preferences, *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, pp. 424–434 (2010).
- [9] Kawahara, D., Peterson, D. W. and Palmer, M.: A Step-wise Usage-based Method for Inducing Polysemy-aware Verb Classes, *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 1030–1040 (2014).
- [10] Erk, K.: A simple, similarity-based model for selectional preferences, *ACL*, Vol. 45, No. 1, p. 216 (2007).
- [11] Erk, K., Padó, S. and Padó, U.: A flexible, corpus-driven model of regular and inverse selectional preferences, *Computational Linguistics*, Vol. 36, No. 4, pp. 723–763 (2010).
- [12] Fellbaum, C.: *WordNet: An Electronic Lexical Database*, Bradford Books (1998).
- [13] Kawahara, D. and Kurohashi, S.: Case frame compilation from the web using high-performance computing, *NAACL*, pp. 1–7 (2006).
- [14] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. and Dean, J.: Distributed representations of words and phrases and their compositionality, *Advances in neural information processing systems*, pp. 3111–3119 (2013).
- [15] Van de Cruys, T., Poibeau, T. and Korhonen, A.: A tensor-based factorization model of semantic compositionality, *Conference of the North American Chapter of the Association of Computational Linguistics (HTL-NAACL)*, pp. 1142–1151 (2013).
- [16] Chambers, N. and Jurafsky, D.: Unsupervised Learning of Narrative Schemas and their Participants, *ACL*, pp. 602–610 (2009).
- [17] Lin, D. and Pantel, P.: DIRT: discovery of inference rules from text, *KDD '01: Proceedings of the seventh ACM SIGKDD international conference*, pp. 323–328 (2001).
- [18] J. Berant, T. A. and Goldberger, J.: Global Learning of Typed Entailment Rules, *ACL*, pp. 610–619 (2008).
- [19] Inoue, N., Ovchinnikova, E., Inui, K. and Hobbs, J.: Coreference Resolution with ILP-based Weighted Abduction, *COLING*, pp. 1291–1308 (2012).
- [20] Modi, A. and Titov, I.: Inducing neural models of script knowledge, *CoNLL-2014*, p. 49 (2014).
- [21] Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K. and Kuksa, P.: Natural language processing (almost) from scratch, *The Journal of Machine Learning Research*, Vol. 12, pp. 2493–2537 (2011).
- [22] Ji, Y. and Eisenstein, J.: One Vector is Not Enough: Entity-Augmented Distributed Semantics for Discourse Relations (2015).
- [23] Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J. and McClosky, D.: The Stanford CoreNLP Natural Language Processing Toolkit, *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 55–60 (2014).
- [24] Luo, X.: On Coreference Resolution Performance Metrics, *HLT/EMNLP*, pp. 25–32 (2005).
- [25] Recasens, M. and Hovy, E. H.: BLANC: Implementing the Rand Index for Coreference Evaluation, *Journal of Natural Language Engineering* (2010).
- [26] Hovy, E., Marcus, M., Palmer, M., Ramshaw, L. and

- Weischedel, R.: OntoNotes: the 90% solution, *Proceedings of the human language technology conference of the NAACL, Companion Volume: Short Papers*, Association for Computational Linguistics, pp. 57–60 (2006).
- [27] Duchi, J., Hazan, E. and Singer, Y.: Adaptive subgradient methods for online learning and stochastic optimization, *The Journal of Machine Learning Research*, Vol. 12, pp. 2121–2159 (2011).