

分散表現による文脈情報を用いた選択選好モデル

大野 雅之 井之上 直也 松林 優一郎 岡崎 直観 乾 健太郎
 東北大学 大学院情報科学研究科
 {masayuki.ono, naoya-i, y-matsu, okazaki, inui}@ecei.tohoku.ac.jp

1 はじめに

本研究では、分散表現を用いた構成的な意味計算により、先行文脈の述語表現から得られる、項の付加的な意味情報を考慮した選択選好モデルを提案する。

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の意味的な偏りのことをいう。例えば、「食べる」という述語の目的格は「りんご」などの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。

選択選好性は述語ごとにそれぞれ異なるため、人手によりその性質を全て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（データスパースネス問題）ため、シソーラスの利用 [1] や単語の分散表現化 [2, 3]、用例のクラスタリング [4, 5] といった方法で、獲得した述語と項の共起用例の一般化が行われてきた。中でも、分散表現を用いた一般化手法が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。

一方、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れば、選好性を評価する現実的な状況においては、評価対象となる名詞に文章中で様々な言及がなされているはずである。そのような名詞に対し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はなく、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 $John$ と Bob 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。これは例文 (1) では、“Police arrested $[John]$.” と “Police arrested $[Bob]$.” の二つの尤もらしさを比較しているこ

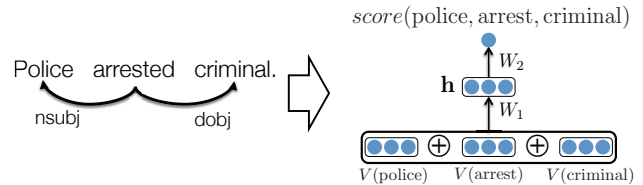


図 1: (s =police, v =arrest, o =criminal) に対する選択選好スコアの計算例

とに相当するが、文脈なしにはこれらの尤もらしさに大きな差は現れない。一方、例文 (2), (3) のように先行文脈により $John$ や Bob に付加されている情報を加味して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。

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

(3) 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的性質に加えて、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。本研究の具体的な成果は、次の二点である。

- 分散表現に基づく Van de Cruys のモデルを拡張し、先行文脈における項に対する言及から、構成的に項の意味的性質を計算する選択選好モデルを提案した。
- 代名詞照応における先行詞候補のランキング問題による評価結果から、提案モデルが項の先行文脈と述語の選択選好性の関連を捉えられていることを確認した。

2 Van de Cruys の SVO モデル

まず、本研究がベースとする Van de Cruys の SVO モデル（以下 SVO モデル）について説明する。SVO モデル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、項と述語の意味ベクトルの学習を同時に行い、(主語 s , 動詞 v , 目的語 o) の三つ組に対し、組合せの尤もらしさを表すスコアを計算する。図 1 に選択選好スコアの計算例を示す。選択選好スコア $score(s, v, o)$ は

$$score(s, v, o) = W_2 \mathbf{h} \quad (1)$$

$$\mathbf{h} = f(W_1(V(s) \oplus V(v) \oplus V(o)) + b) \quad (2)$$

と計算する。ここで $V(w) \in \mathbb{R}^d$ は単語 w のベクトル^{*1}、 $\mathbf{h} \in \mathbb{R}^h$ は隠れ層、 $W_1 \in \mathbb{R}^{h \times 3d}$, $W_2 \in \mathbb{R}^{1 \times h}$ は各層の

*1 元論文では単語 w が主語、動詞、目的語のいずれかによって異なる単語ベクトルを用いていたが、同一のベクトルを用いたほうが高い性能が得られたため本稿では同一のベクトルを用いた。

[John] attacked Bob. ... Police arrested [John].

主語	動詞	目的語
police	arrest	[John, who attacked Bob]

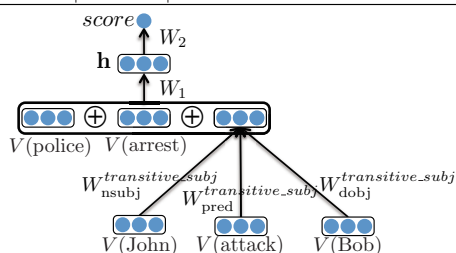


図 2: 提案手法の概要

重み行列, b はバイアス項である. また, \oplus はベクトルの連結を表し, 活性化関数 $f(\cdot)$ には \tanh を用いる.

SVO モデルの学習には Collobert ら [6] が提案した教師なし学習を用いる. 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「主語」, 「目的語」, 「主語と目的語の両方」をそれぞれ別の単語に置き換えた擬似負例を作成する. 次に, これらの正例と擬似負例についての選択選好スコアを式 (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 $V(w)$ と重み行列 W_1, W_2 , バイアス項 b) を学習する.

$$\begin{aligned} & \max(0, 1 - \text{score}(s, v, o) + \text{score}(\tilde{s}, v, o)) \\ & + \max(0, 1 - \text{score}(s, v, o) + \text{score}(s, v, \tilde{o})) \\ & + \max(0, 1 - \text{score}(s, v, o) + \text{score}(\tilde{s}, v, \tilde{o})) \end{aligned} \quad (3)$$

ここで, \tilde{s} , \tilde{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, ランダム置換により不自然になった三つ組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.

3 文脈を考慮した選択選好モデル

次に, SVO モデルを項の先行文脈を考慮したモデルに拡張する. 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による付加要素を考慮した選好性の計算は, 例文 (5), (6) のように, 先行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, SVO モデルにおける項ベクトルの部分に, この修飾関係を表現した分散表現ベクトルを埋め込んで選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.

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

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

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

項の意味的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 [7, 8, 9] に倣い, 選択選好性の判定対象となる名詞が項として出現する述語 (動詞, 形容詞, コピュラ) の述語項関係を文脈として採用する.

述語項関係を表現するベクトルについては, Hashimoto ら [10] が構成要素から計算する手法を提案しており, 本研究もそれに倣い, 構成性に基づいて文脈付きベクトルを計算する. 我々の手法では, 述語の種類によって名詞と述語の意味的関係が異なると考え, 述語の種類毎に異なる合成関数を考える. また, 他動詞のように述語が複数の項をとる場合, 対象名詞が先行文脈の動詞においていずれの格関係になっているかで名詞が帯びる意味が異なるため, これを区別する. 上記の例であれば, 例文 (5) の John は (John, arrest, Bob) の主語としての John であり, 例文 (6) 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.

また一般に, 文章中には, 対象の名詞を項に取り, その名詞の文脈とみなすことのできる複数の述語が書かれていると考えられる. したがって, 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, 今回は埋め込みの結果最も選好性の高くなる述語項関係を一つ選択し, この述語項関係が対象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 2 のようなネットワークで表現される. 提案手法では, 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付きベクトルのいずれかを選択できるように拡張する. これにより, 対象の名詞に対する先行文脈として述語による言明が存在する場合は文脈付きベクトルを用い, 文脈がない場合は単語ベクトルを用いることができる.

3.1 定式化

前述のとおり, 我々の手法では, 述語の種類及び, 述語との格関係の違いを区別して文脈付きベクトルの合成を行う. そこで, 述語の種類と格関係の組み合わせにより, 文脈の種類を { 他動詞-主語, 他動詞-目的語, 自動詞, 形容詞, コピュラ } の 5 種類に分け, 文脈タイプに応じた変換行列を用いて単語 w に対する文脈付きベクトルを生成することで, 各述語表現における役割を考慮した意味ベクトルの計算を可能にする.

w に対する先行文脈 c_w による文脈付きベクトルの生成関数 $V'(w, c_w)$ を以下の式で定義する.

$$V'(w, c_w) = \begin{cases} V(w) & \text{if } c_w = \phi; \\ \tanh(W_{\text{pred}}^t V(p) + \sum_i W_{\text{arg}, i}^t V(a_i)) & \text{otherwise;} \end{cases} \quad (4)$$

ここで, $W_{\text{pred}}^t \in \mathbb{R}^{d \times d}$ は文脈タイプ t の文脈における述語 p に作用する行列であり, $W_{\text{arg}, i}^t \in \mathbb{R}^{d \times d}$ は文脈タイプ t の文脈における項 a_i に作用する行列である. 我々の定義する文脈タイプにおいては, 他動詞の場合には二つの項を取り, それ以外は一つの項を取る. 我々は, 当

面の設定として5種類の文脈タイプを用意したが、この定式化により特定の項数を持つ任意の文脈タイプの意味ベクトル計算が可能である。先行文脈が存在しない場合 ($c_w = \phi$) には、単語自身のベクトルを用いる。

提案モデルは、SVO モデルにおける項の入力ベクトル $V(s)$, $V(o)$ を先行文脈を考慮したベクトル $V'(s, c_s)$, $V'(o, c_o)$ に置き換えた以下の式で選好性を計算し、従来の選択選好モデルでは捉えることが出来なかった文脈を考慮した選好性の計算を実現する。

$$\max_{c_s, c_o} \text{score}(s, v, o, c_s, c_o) = W_2 \mathbf{h} \quad (5)$$

$$\mathbf{h} = f(W_1(V'(s, c_s) \oplus V(v) \oplus V'(o, c_o)) + b) \quad (6)$$

3.2 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。

- 単語ベクトル $V(w)$ と 重み行列 W_1, W_2 , バイアス項 b 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 $W_{\text{pred}}^t, W_{\text{arg}, i}^t$ の学習も行う。
- 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。
- 文脈付きベクトルと単語ベクトルの値域を合わせるため、単語ベクトル $V(w)$ 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。

(7) [*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 B1 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 [11]*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。可能な限り正確な共参照解析結果を用いるため、同一文内で、かつ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*4。

*2 <http://lemurproject.org/clueweb12/>

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

*4 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、コーパス全体で単語の品詞を考慮した出現頻度上位10万単語を語彙として用いた。また、Type A に関しては、共起頻度30以上の共起事例を対象とし、学習時間の都合上、30分の1をサンプリングし、5,431,385事例(1,689,780種類)を学習に用いた。Type B1, Type B2 に関しては、2,473,311事例を学習に用いた。

4 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。

4.1 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 ($\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}(i)}$) を用いる。ここで、 $|Q|$ は問題総数、 $\text{rank}(i)$ は i 番目の問題において、式(5)によって計算された正解先行詞の順位である。

評価データには OntoNotes 4.0 [12] 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、構文木、述語項構造、代名詞・名詞句間の共参照関係等が付与されている。

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。

(8) In *his_(p)* 40-minute *speech_(i)*, *Chen_(j)* declared the *determination_(k)* of the *people_(l)* ... (中略)... *Chen_(m)* 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、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て扱った。例えば、 $Chen_{(j)}$ と $Chen_{(m)}$ 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

表 1: 先行詞候補ランキング問題での MRR

設定	MRR
SVO モデル (ベースライン)	0.4157
提案モデル	0.4514

脈は $c_j = \langle \text{Chen, declared, determination, subj} \rangle$ であるが, 訓練事例の中に現れなかった単語を含む場合, $c_j = \phi$ とした.

最終的には, OntoNotes コーパスに出現する代名詞のうち, 4,368 問の代名詞を評価対象とした.

4.2 実験設定

単語ベクトル, 隠れ層の次元数はそれぞれ $d = 50, h = 100$ 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 Adam [13] 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, 2 節で説明した Van de Cruys の SVO モデルに相当する.

4.3 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 SVO モデルよりも高い性能を達成しており, 項の先行文脈の情報を反映した選択選好性の計算が行えていることが確認できた.

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 20 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.

(9) 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... (中略)... *It*_(pro) has built up an installation capacity of 7,000 sets of program ... (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) である. SVO モデルによるランク付けでは 36 位であったが, 提案モデルによるランク付けでは 1 位に改善された. SVO モデルでは, $it_{(pro)}$ の述語である「建設する (build)」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)」が build 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.

5 おわりに

本稿では, 述語の選択選好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モデルを提案した. 述語の選択選好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クできるようになることを確認した. 本稿では, ただひとつの述語項関係を項の先行文脈として扱ったが, 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] Philip Resnik. Selectional constraints: An information-theoretic model and its computational realization. *Cognition*, Vol. 61, No. 1, pp. 127–159, 1996.
- [2] Katrin Erk, Sebastian Padó, and Ulrike Padó. A flexible, corpus-driven model of regular and inverse selectional preferences. *Computational Linguistics*, Vol. 36, No. 4, pp. 723–763, 2010.
- [3] Tim Van de Cruys. A neural network approach to selectional preference acquisition. In *EMNLP*, pp. 26–35, 2014.
- [4] Mats Rooth, Stefan Riezler, Detlef Prescher, Glenn Carroll, and Franz Beil. Inducing a semantically annotated lexicon via em-based clustering. In *ACL*, pp. 104–111. Association for Computational Linguistics, 1999.
- [5] Diarmuid O Séaghdha. Latent variable models of selectional preference. In *ACL*, pp. 435–444, 2010.
- [6] Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. Natural language processing (almost) from scratch. *The Journal of Machine Learning Research*, Vol. 12, pp. 2493–2537, 2011.
- [7] N. Inoue, E. Ovchinnikova, K. Inui, and J. Hobbs. Coreference Resolution with ILP-based Weighted Abduction. In *COLING*, pp. 1291–1308, 2012.
- [8] A. Rahman and V. Ng. Resolving Complex Cases of Definite Pronouns: The Winograd Schema Challenge. In *EMNLP-CoNLL*, pp. 777–789, 2012.
- [9] Haoruo Peng, Daniel Khashabi, and Dan Roth. Solving Hard Coreference Problems. In *NAACL*, pp. 809–819, 2015.
- [10] Kazuma Hashimoto, Pontus Stenetorp, Makoto Miwa, and Yoshimasa Tsuruoka. Jointly learning word representations and composition functions using predicate-argument structures. In *EMNLP*, pp. 1544–1555, 2014.
- [11] Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. The Stanford CoreNLP natural language processing toolkit. In *ACL*, pp. 55–60, 2014.
- [12] Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. Ontonotes: the 90% solution. In *NAACL-Short*, pp. 57–60, 2006.
- [13] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.