

# 関係代数の意味論に基づく含意関係認識

Textual Entailment Recognition based on a Semantics of Relational Algebra

田 然      岡崎 直観      乾 健太郎  
Ran Tian    Naoaki Okazaki    Kentaro Inui

東北大学  
Tohoku University

Dependency-based Compositional Semantics (DCS) is a framework for constructing meanings of sentences from dependency-like trees. Toward a seamless application of DCS to textual entailment recognition, we connect DCS with a vector-based composition model, showing that vector-based composition brings an ability to calculate similar phrases as similar vectors, whereas DCS provides structured representations that facilitate inference.

## 1. はじめに

自然言語文を意味計算や命令実行、論理推論などを可能にする表現に変換することは、意味理解の主要目標の一つである。このための形式意味論は、単語をある種の機能的ユニットとみなし、論理規則に沿ってそれらのユニットを組み上げることで文の意味を表す。依存関係に基づく構成的意味論 (Dependency-based Compositional Semantics, DCS) [4] もその一つである。一方、分布意味論に基づくベクトル表現を使って、単語の意味を表すだけでなく、フレーズや文の意味を構成的に計算するモデルも研究されてきた [6, 1, 2, 8, 7, 3]。本稿は、ベクトル表現を構成的に計算するモデルとして、DCS を実現できることを示す。これによって、記号的に扱われてきた単語や文の意味を表す論理式はベクトル表現を持ち、意味的に似たフレーズの似たベクトル表現が期待できる。加えて DCS の形式意味論は、命令実行や推論を可能にする論理構造を保証してくれる。

## 2. DCS

DCS は単語の外延から出発して意味を構築する。外延とは、その単語が指す「もの」の集合であり、ものは複数の特徴 (「フィールド = 値」のような構造を特徴と呼ぶ) を持つ組によって表現される。例えば、単語 *ban* の外延は全ての *ban* イベントからなる集合で、各イベントにはそのイベント参加者の組、すなわち禁止を発令する側 ( $\text{SUBJ} = \text{Canada}$ ) や禁止される物 ( $\text{COMP} = \text{Thalidomide}$ ) が記録されていると考えられる。このように外延は  $\text{ban} = \{(\text{SUBJ} = \text{Canada}, \text{COMP} = \text{Thalidomide}), \dots\}$  のように表現され、「カナダがサリドマイドの使用を禁止した」イベントは *ban* の外延に属することを意味する (図 1a)。

集合演算によって新しい外延が作られ、意味の合成を表す。例えば「*pet fish*」は *pet* の外延と *fish* の外延の交わりで表現されると考えられ、ここで「交わり」は一つの集合演算である。もう一つ、「射影」の集合演算は、組からある特定のフィールドの値を取り出して集合を作る操作である。例えば  $\pi_{\text{COMP}}(\text{ban})$  は外延 *ban* の元の  $\text{COMP}$  フィールドの値からなる集合を表す (すなわち  $\{\text{Thalidomide}, \dots\}$  である)。本稿では、物事の「名前」を示すフィールドとして  $\text{ARG}$  を仮定する。例えば  $\pi_{\text{ARG}}(\text{drug})$  は薬の名前の集合になる。

DCS では、更に「射影の逆像」という集合演算も必要である。射影  $\pi_N$  と値の集合  $V$  に対して、逆像を  $\pi_N^{-1}(V) := \{x \mid \pi_N(x) \in V\}$  と定義する。例えば、

$$D_1 := \pi_{\text{SUBJ}}^{-1}(\pi_{\text{ARG}}(\text{man}))$$

は ( $\text{SUBJ} = x, \dots$ ) のような形の組からなる集合で、ただし  $x$  は *man* の名前になる値に限る (言い換えれば  $x \in \pi_{\text{ARG}}(\text{man})$ )。これによって  $\text{sell} \cap D_1$  は「男が (何かを) 売る」イベント全体の集合を表現できる。同様に「*banned drugs*」の外延は次のように計算できる (図 1a)：

$$D_2 := \text{drug} \cap \pi_{\text{ARG}}^{-1}(\pi_{\text{COMP}}(\text{ban}))$$

すると、外延

$$D_3 := \text{sell} \cap D_1 \cap \pi_{\text{COMP}}^{-1}(\pi_{\text{ARG}}(D_2))$$

は、「男が禁止された薬を売る」イベント全体の集合を表す。

上記のような計算は、DCS 木によって統制されている。文「*a man sells banned drugs*」の DCS 木を図 1b に示した。公式的に、DCS 木は各ノードが内容語の外延、各エッジの両端がフィールド名ラベル付きの木として定義される。仮にノード  $x$  に子  $y_1, \dots, y_n$  があるとして、対応するエッジ  $(x, y_1), \dots, (x, y_n)$  はそれぞれ  $(P_1, L_1), \dots, (P_n, L_n)$  でラベルされているとすると、 $x$  をルートとする部分木の外延  $\llbracket x \rrbracket$  は再帰的に

$$\llbracket x \rrbracket := x \cap \bigcap_{i=1}^n \pi_{P_i}^{-1}(\pi_{L_i}(\llbracket y_i \rrbracket)) \quad (1)$$

のように計算される。例えば図 1b の DCS 木に沿って計算すると、「*a man sells banned drugs*」の意味を表す外延  $D_3$  は上記のように計算される。DCS 木は依存構造木からルールで変換でき、更に高度な意味現象 (量子化や最上級など) への拡張も可能である [9]。

## 3. DCS のベクトル化

任意の内容語  $w$  に対して、その外延をベクトル  $v_w$  (クエリベクトル)、外延の典型的な元をベクトル  $u_w$  (答えベクトル) でモデリングする。答え  $u$  がクエリ  $v$  に属す確率が  $\exp(v \cdot u)$  に比例するように、ベクトル  $u$  と  $v$  を学習する。学習のソースとして、コーパスを構文解析して得られる DCS 木のコレク

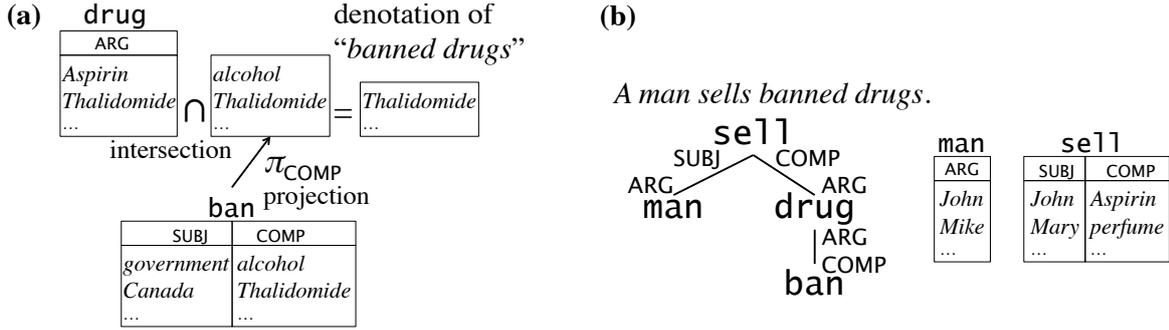


図 1: (a) 「banned drugs」の外延の計算 (b) 「a man sells banned drugs」の DCS 木

ションを使う。そのアイデアは、コーパスに大量にある肯定文の DCS 木の外延は、大体空集合でないことに基づく。例えば、「kids play」はある子供が存在して実際遊んでいることを意味する。よって、play の外延のある元は  $\pi_{\text{SUBJ}}^{-1}(\pi_{\text{ARG}}(\text{kid}))$  にも属し、kid の外延のある元は  $\pi_{\text{ARG}}^{-1}(\pi_{\text{SUBJ}}(\text{play}))$  にも属す。このことは訓練シグナルとして、 $\mathbf{u}_{\text{play}}$  と  $\pi_{\text{SUBJ}}^{-1}(\pi_{\text{ARG}}(\text{kid}))$  のクエリベクトルの内積、及び  $\mathbf{u}_{\text{kid}}$  と  $\pi_{\text{ARG}}^{-1}(\pi_{\text{SUBJ}}(\text{play}))$  のクエリベクトルの内積の増大に寄与する。大規模なコーパスでこれを最適化すると、「典型的な」play や kid の元は  $\mathbf{u}_{\text{play}}$  や  $\mathbf{u}_{\text{kid}}$  によってモデリングされるようになる。

一般的に下記の定理が成り立つ。

**定理 1** ある DCS 木の外延が空でないとして仮定する。その木のノード  $x$  から  $y$  へのパスに対して、経由するエッジが順に  $(P, L), \dots, (K, N)$  とラベル付いているとすると、 $y$  の外延のある元は  $\pi_N^{-1}(\pi_K(\dots(\pi_L^{-1}(\pi_P(x))\dots))$  にも属する。

よって、DCS 木の任意二つのノードに対して、一方から他方へのパスは一つの訓練サンプルを与える。

このような DCS のベクトル化は skip-gram モデル [5] と関連が深い。Skip-gram モデルは、単語  $w$  に対してターゲットベクトル  $\mathbf{v}_w$  とコンテキストベクトル  $\mathbf{u}_w$  を学習する。ある文脈で単語  $y$  と  $x$  が共起する確率が  $\exp(\mathbf{v}_x \cdot \mathbf{u}_y)$  に比例するように、ベクトルの学習を行う。ここで、もし「ある文脈」を「同じ DCS 木」に置き換えると、skip-gram モデルはほぼ DCS のベクトル化になる。唯一の相違点は、skip-gram モデルにおけるターゲットベクトル  $\mathbf{v}_x$  が DCS のベクトル化では単語に対するベクトルではなく、 $x$  から  $y$  へのパスに対応するクエリベクトルになる点である。パスでは統語役割の作用 ( $\pi_{\text{COMP}}$  や  $\pi_{\text{SUBJ}}^{-1}$  など) も考慮しているので、DCS のベクトル化は skip-gram モデルに統語役割を統合したモデルとも考えられる。

### 3.1 加法構成

Skip-gram モデルで訓練された単語ベクトルは、足し算によって意味の合成ができることで知られている。[10] では、この加法構成性に関する一つの理論分析を与えた。簡単にいうと、ターゲットベクトル  $\mathbf{v}_w$  は、 $w$  周辺のコンテキスト単語の分布をエンコードしているとみなせる。別の単語  $x$  が与えられると、 $\mathbf{v}_w$  を二つのコンポーネントに分解でき、一つは  $x$  とシェアしているコンテキスト単語の分布をエンコードし、もう一つは  $x$  とシェアしない  $w$  独自のコンテキスト単語の分布をエンコードする。ベクトル  $\mathbf{v}_w$  と  $\mathbf{v}_x$  が加算されると、それぞれ独自のコンテキスト単語分布をエンコードするコンポーネントは互いに打ち消す作用を持つ。なぜなら、独自のコンテキスト単語分布は、互いにほぼ独立だからだ。結果、 $w$  と  $x$  がシェ

アしているコンテキスト単語の分布をエンコードするコンポーネントが強化される。 $\frac{1}{2}(\mathbf{v}_w + \mathbf{v}_x)$  をこのコンポーネントの近似とみなした時、誤差に対する上界が証明できた。DCS のベクトル化においても、同様のメカニズムが存在する。DCS 木において、二つのパスが同じコンテキスト単語をシェアするのは、同じノード  $y$  にたどり着く時である。これは、 $y$  の外延のある元は両方のパスの外延に属することを意味する。例えば、「kids play balls」を言われた時、一つの play イベントが  $\pi_{\text{SUBJ}}^{-1}(\pi_{\text{ARG}}(\text{kid}))$  にも  $\pi_{\text{COMP}}^{-1}(\pi_{\text{ARG}}(\text{ball}))$  にも属する。足し算によってこのシェアしている共通部分が強化されるので、パスの外延の交わりはそのクエリベクトルの足し算によって近似的に表現することができる。

### 3.2 射影

任意の外延  $X_1, X_2$  と射影  $\pi_N$  に対して、

$$\pi_N(X_1 \cap X_2) \subseteq \pi_N(X_1) \cap \pi_N(X_2). \quad (2)$$

が成り立つ。そして「 $\subseteq$ 」が度々「 $=$ 」になる（例えば  $\pi_N$  が単射や  $X_1$  が  $\pi_N^{-1}(V)$  の形をしている時）ことを考えると、「交わり」をベクトルの足し算で実現するなら「射影」は線形写像で実現するのが自然である。そうすると

$$(\mathbf{v}_1 + \mathbf{v}_2)M_N = \mathbf{v}_1M_N + \mathbf{v}_2M_N \quad (3)$$

が任意のベクトル  $\mathbf{v}_1, \mathbf{v}_2$  と行列  $M_N$  に対して成り立ち、(2) と平行である。そして、 $\pi_N$  を行列  $M_N$  で実現するとすると、 $\pi_N^{-1}$  は逆行列  $M_N^{-1}$  で実現される。なぜなら、任意の集合  $V$  に対して  $\pi_N(\pi_N^{-1}(V)) = V$  だからである。

### 3.3 DCS 木のクエリベクトル

DCS において意味構成に必要な三つの演算、交わりと射影と射影の逆像に対して全てベクトル演算が定められた。これで DCS 木のクエリベクトルを (1) に習って次のように計算することができる。

$$\mathbf{v}_{[x]} := \mathbf{v}_x + \frac{1}{n} \sum_{i=1}^n \mathbf{v}_{[y_i]} M_{L_i} M_{P_i}^{-1}. \quad (4)$$

ベクトル表現を使うことで意味的に似たフレーズは近いベクトルになり、更に DCS の論理構造によって推論などもできる枠組みを実現できた。

## 4. おわりに

DCS の形式意味論を含意関係認識に応用する研究は [11] からなされてきた。本稿は、その枠組みに分布意味論から得ら

---

れる類似語・類似フレーズの言語知識を自然に取り入れる方法の一つ記述した。実験的検証やシステムの評価は追って報告する。

## 参考文献

- [1] Marco Baroni and Roberto Zamparelli. Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. In *Proceedings of EMNLP*, 2010.
- [2] Edward Grefenstette and Mehrnoosh Sadrzadeh. Experimental support for a categorical compositional distributional model of meaning. In *Proceedings of EMNLP*, 2011.
- [3] Kazuma Hashimoto, Pontus Stenetorp, Makoto Miwa, and Yoshimasa Tsuruoka. Jointly learning word representations and composition functions using predicate-argument structures. In *Proceedings of EMNLP*, 2014.
- [4] Percy Liang, Michael I. Jordan, and Dan Klein. Learning dependency-based compositional semantics. *Computational Linguistics*, 39(2), 2013.
- [5] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In *Advances in NIPS*, 2013.
- [6] Jeff Mitchell and Mirella Lapata. Composition in distributional models of semantics. *Cognitive Science*, 34(8), 2010.
- [7] Denis Paperno, Nghia The Pham, and Marco Baroni. A practical and linguistically-motivated approach to compositional distributional semantics. In *Proceedings of ACL*, 2014.
- [8] Richard Socher, Brody Huval, Christopher D. Manning, and Andrew Y. Ng. Semantic compositionality through recursive matrix-vector spaces. In *Proceedings of EMNLP*, 2012.
- [9] Ran Tian, Yusuke Miyao, and Takuya Matsuzaki. Logical inference on dependency-based compositional semantics. In *Proceedings of ACL*, 2014.
- [10] Ran Tian, Naoaki Okazaki, and Kentaro Inui. The mechanism of additive composition. arXiv:1511.08407, 2015.
- [11] 田 然 and 宮尾 祐介. 関係代数に基づく推論の含意関係認識への応用. In 人工知能学会第 27 次年次大会発表論文集, 2013.