

# 論理式による意味表現と証明プロセスに着目した 文の類似度学習方法の提案

谷中 瞳<sup>1</sup> 峯島 宏次<sup>2,4</sup> Pascual Martínez-Gómez<sup>3</sup> 戸次 大介<sup>2,4</sup>

<sup>1</sup> 東京大学 <sup>2</sup> お茶の水女子大学 <sup>3</sup> 産業技術総合研究所 <sup>4</sup> JST, CREST

hitomiyana@ecc.u-tokyo.ac.jp, minesima.koji@ocha.ac.jp,  
pascual.mg@aist.go.jp, bekki@is.ocha.ac.jp

## 1 はじめに

文の持つ意味に基づいて文の類似度を計算できれば、膨大な文書から知りたい情報を含む文書を特定でき、文書検索や文書要約への幅広い応用が期待される。従来の文の類似度計算における代表的な言語モデルとして、文中の単語の出現頻度に基づいて文をベクトルで表現する Bag of Words モデルがある。しかし、Bag of Words モデルでは、語順が無視されるという問題点がある。たとえば、“A cat is watching a bird.”と“A bird is watching a cat.”という2つの文は文中の単語は全く同じであり、どちらも「見る」という動作が共通するという点では類似しているが、主語と目的語が逆であるため、同じ意味を示さない。

一方で、含意関係認識といった深い意味処理を必要とするタスクでは、推論を扱いやすい論理式を用いて文の複雑な意味を表現する研究 [1][2][3] が行われてきた。そこで、本研究では文全体の意味を論理式で表現し、推論を用いて文の類似度を学習する方法を検討する。この方法を実現するために、組合せ範疇文法に基づく統語解析と意味合成によって文を高階述語論理式に変換し、自動推論を行う統合的システムである ccg2lambda [2] を利用する。

## 2 先行研究

### 2.1 推論を利用した文の類似度学習

文を一階述語論理式に変換し、推論を利用して文の類似度を学習する先行研究として、The Meaning Factory [3], UTexas [4] がある。The Meaning Factory は述語の一致率と証明結果を特徴量に用いて決定木学習により類似度を導出する。UTexas は Probabilistic Soft Logic に基づき、単語の分散表現を用いて証明の

確信度を計算し、これを特徴量に用いて加法回帰モデルにより類似度を導出する。

いずれの手法も一階述語論理式の証明から導出した特徴量を組み合わせることで、表層的な特徴量のみを用いた場合と比較して、精度が向上している。本研究では証明論的意味論 [5] の観点から文の意味の類似性を定義し、証明結果だけではなく、証明の過程にも着目して体系的に特徴量を設計し、文の類似度学習を行う。証明の過程にも着目することによって、より詳細に文間の意味関係を捉えられる可能性がある。また、先行研究で意味表現として用いられている一階述語論理は、一般化量量子のような複雑な意味表現を扱うことが難しいという問題がある。本研究では高階述語論理を用いることによって、一般化量量子を含む文の意味表現も扱うことが可能である。

### 2.2 ccg2lambda:高階述語論理に基づく自動推論システム

ccg2lambda [2] は統語構造からの意味合成に適した文法体系である組合せ範疇文法 (Combinatory Categorical Grammar, CCG) [6] によって統語解析を行い、ラムダ計算による意味合成によって文を高階述語論理式に変換し、自動推論を行うシステムである。

CCG は語彙化文法の一つであり、語の統語・意味に関する情報を記述する辞書と、語の統語的な振る舞いと語の意味合成の計算方法を同時に指定する組合せ規則から構成される。CCG において、統語範疇は N, NP, S, CONJ 等の基底範疇と二項演算子 /, \ によって再帰的に定義される複合範疇の二種類がある。

語の意味はラムダ項によって表現され、組合せ規則によって得られたラムダ項を  $\beta$  簡約することで文の意味表示である論理式が得られる。意味表示は Neo-Devvidsonian Event Semantics [7] に基づく。Neo-

Devidsonian Event Semantics は動詞をイベントを項にとる 1 項述語として分析する意味論であり、副詞や前置詞などの修飾表現をイベントに対する述語とみなすことで、時制のような複雑な意味を含む論理式や語彙知識から導出した公理を簡潔に記述できる。

高階述語論理による証明は、高階論理・型理論に基づく定理証明支援系である Coq [8] を用いる。Coq の証明系は自然演繹 [9] に基づくものであり、タクティクスと呼ばれる証明手続きを用いることで証明の検証を行い、Ltac と呼ばれるタクティクスの記述言語を用いることで証明探索の手続きを定義する。cgg2lambda は Coq の一階述語論理の部分系に対する自動推論を含む自動証明機能と高階の公理とを組み合わせることで、自然言語の効率的な自動推論を可能にする。含意関係認識システムの評価用データセットである FraCaS [10] を用いて cgg2lambda の評価を行った結果、教師あり学習を用いた手法と同等の正答率だったことが報告されており [2]、文の類似度計算への応用が見込まれる。

### 3 cgg2lambda を用いた文の類似度学習

#### 3.1 システムの構成

本研究では文  $A, B$  間の意味の類似度を文  $A, B$  の同値関係の証明可能性を介して定義する。同値関係を証明できる可能性が高いほど、文  $A, B$  は同じ意味を持つと考えられる。そこで、文  $A, B$  の同値関係の証明可能性を自動推論を用いて定量化し、特徴量に用いることで、文  $A, B$  間の類似度学習を実現する。

本研究の提案システムの全体像を図 1 に示す。本研究ではまず、CCG パーザによる統語解析によって、自然言語の英文  $A, B$  を構文木に変換する。cgg2lambda では 2 種類の CCG パーザを組み合わせることによって、構文的曖昧性が解消され、精度が向上したことが報告されている [11]。そこで、先行研究に準じて C&C [12] と EasyCCG [13] という 2 種類の CCG パーザを用いる。次に、ラムダ計算による意味合成によって、構文木から高階述語論理式  $A', B'$  に変換する。

次に、Coq を用いて、はじめは公理なしで自然演繹による証明を試みる。まず、 $A' \Rightarrow B', B' \Rightarrow A'$  の証明を行う。証明に失敗した場合は、結論を否定した  $A' \Rightarrow \neg B', B' \Rightarrow \neg A'$  の証明を行う。これは結論の否定が証明可能な場合は、結論も結論の否定も証明不可能な場合と比較して、論理式が共通の部分を多く含むためである。実際に、評価に用いる文の意味的類似度

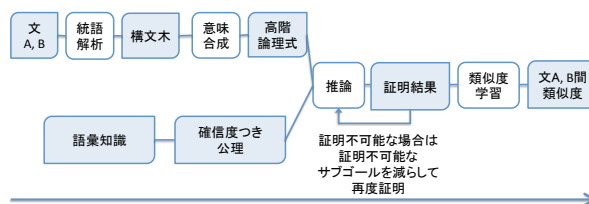


図 1: 提案システムの全体像

データセットである SemEval2014 Task1 の Sentences Involving Compositional Knowledge (SICK) [14] では、含意関係の正解ラベルが「矛盾」である文ペアの 7 割において、正解の類似度スコアが高く設定されている。いずれも証明に失敗した場合は、語彙知識から公理を生成して再度証明を行う。まず、生成した公理を用いて  $A' \Rightarrow B', B' \Rightarrow A'$  の証明を行い、証明に失敗した場合は結論を否定して再度証明を行う。いずれも証明に失敗した場合は、 $A' \Rightarrow B', B' \Rightarrow A'$  の証明の途中で証明不可能と判定されたサブゴールを強制的に削除して、再度証明を行う。

最後に、各証明の過程と結果から特徴量を設計し、決定木学習によって類似度の学習を行う。決定木学習におけるハイパーパラメータはグリッドサーチを用いて最適化する。予測精度の評価には平均二乗誤差 (MSE) を評価指標として、5 分割の交差検定を用いる。次節で公理の作成方法と、類似度学習に用いる特徴量の設計方法を説明する。

#### 3.2 語彙知識からの公理の作成

語彙知識から前提と結論の述語間の意味的關係をチェックし、公理を作成する。作成する公理は、証明の途中で証明不可能と判定されたサブゴールに関して、前提と結論で項をシェアしている述語だけに着目して候補を絞り込む。語彙知識は WordNet を用いる。形態変化、派生形、同義語、反意語、上位語、概念間の類似性、下位語の順に述語間の関係をチェックし、いずれかにマッチした場合は公理を作成する。

#### 3.3 証明を用いた特徴量の設計

各証明を用いて、前提と結論における述語の一致率、証明に使用した公理の数、公理の確信度、証明可能なサブゴールの割合、証明の結果という 5 種類の特徴量を導出する。述語の一致率を除く 4 種類の特徴量は  $A' \Rightarrow (\neg)B', B' \Rightarrow (\neg)A'$  の証明から導出し、全部で

ID	文 1	文 2	含意関係	類似度
23	There is no biker jumping in the air.	A lone biker is jumping in the air	No	4.2
1412	Men are sawing logs.	Men are cutting wood.	Yes	4.5
1493	A man is playing an electric guitar.	A man is playing guitar	Yes	4.9
9963	The animal is grazing on the grass.	The cop is sitting on a police bike.	Unknown	1

表 1: SICK データセットの例.

$1 + 4 \times 2 = 9$  個の特徴量を使用する. 特徴量はいずれも  $[0-1]$  の範囲の値をとる.

- 前提と結論における述語の一致率  
一般的に, 前提と結論との間で述語が一致しているほど, 証明可能性が高い. そこで, 前提と結論で述語が一致している割合を特徴量に用いる.
- 証明に使用した公理の数, 公理の確信度  
公理を用いることによって証明可能な場合を考慮して, 導入した公理の数と確信度を特徴量に用いる. WordNet から導出した公理は述語間に共通する上位概念への最短経路の長さを公理の確信度とする. 複数の公理を導入した場合は, 導入した各公理の確信度の平均を特徴量として採用する. また, 公理を用いずに証明可能だった場合は特徴量を 1 とする.
- 証明可能なサブゴールの数の割合  
証明に失敗したサブゴールを強制的に削除することによって証明可能な場合を考慮して, サブゴールのうち証明可能なサブゴールの数の割合を特徴量に用いる. 具体的には, 前提のプールに現れる論理式の数を  $m$ , 削除せずに残ったサブゴールの数を  $n$  としたとき,  $n/m$  を特徴量として導出する. また, サブゴールを削除せずに証明可能だった場合は特徴量を 1 とする.
- 証明の結果  
3.1 で述べた通り, 結論または結論の否定が証明可能な場合と, どちらも証明不可能である場合をそれぞれ区別するために, 結論が証明可能な場合は 1, 結論の否定が証明可能な場合は 0.5, どちらも証明不可能だった場合は 0 を特徴量とする.

## 4 実験と評価

### 4.1 実験環境

3.1 で述べたように, 本研究の提案手法は, SICK[14] を用いて評価した. SICK は, 表 1 に示すように, 2 つの文の意味的な類似度が人手で  $[1-5]$  の範囲の値でスコア付けされている. 訓練データは 5000 の文ペア, テストデータは 4927 件の文ペアを用いる. 評価には, 提案手法によって計算された類似度スコアと, 正解スコアとの Pearson の相関係数  $\gamma$ , Spearman の相関係数  $\rho$ , MSE を用いる. 類似度スコアは正解スコアの範囲  $[1-5]$  になるように正規化した. また, SICK では, 含意関係の正解ラベルが「矛盾」である文ペアも正解スコアが高い傾向にあり, 含意関係の正解ラベルが「不明」である文ペアは広範囲の正解スコアをとる. そこで, 含意関係の正解ラベルが「含意」の場合は 5, 「矛盾」または「不明」の場合は 3 として類似度スコアに置き換えた結果をベースラインとする. ベースラインに加え, 2.1 で紹介した The Meaning Factory[3], UTexas[4] を比較対象とする.

### 4.2 実験結果・考察

表 2 に, 提案手法と比較対象の相関係数を示す. 既存手法と比べると, 提案手法は UTexas の結果を上回ったが, The Meaning Factory の結果を下回った. しかし, UTexas では 17 種類, The Meaning Factory では 32 種類の特徴量を使用しているが, 本研究では比較手法よりも少ない特徴量で同等の結果を出すことができた. 提案手法では, パラフレーズに関する語彙知識がないために公理を生成できず証明に失敗し, 類似度スコアが正解スコアと離れた例が数件見られた. The Meaning Factory は語彙知識として提案手法で用いた WordNet に加え, 単語の分散表現や Paraphrase Database (PPDB) を使用しており, 提案手法もパラフレーズに関する語彙知識を追加することによって改善が見込まれる.

	$\gamma$	$\rho$	MSE
The Meaning Factory	0.827	0.772	0.322
UTexas	0.714	0.674	0.499
Baseline	0.653	0.745	0.808
提案手法	0.735	0.720	0.693

表 2: 提案手法と比較対象の手法の相関係数.

表 3 に, 各特徴量における相関係数を示す. 証明可能なサブゴールの割合を特徴量に用いることによって精度が向上しており, 証明の過程に関する情報は類似度学習に有効な特徴量であることが示唆された.

	$\gamma$	$\rho$	MSE
述語の一致率	0.515	0.495	0.875
公理の確信度	0.451	0.462	0.923
公理の数	0.223	0.232	0.998
サブゴールの割合	0.517	0.465	0.879
証明の結果	0.432	0.444	0.920
全特徴量 (サブゴール除)	0.705	0.696	0.725
全特徴量	0.735	0.720	0.693

表 3: 各特徴量を用いたときの相関係数.

## 5 おわりに

本研究では `cgc2lambda` を用いて文を論理式に変換し, 自動推論によって体系的に文の意味類似度学習に必要な特徴量を導出する手法を提案した. 今後の課題として, 証明に必要な語彙知識を効果的に選択する方法を検討する. また, 同値関係の証明可能性を定量化するための指標として, 今回は単純に述語の一致率や証明可能なサブゴールの割合を特徴量に用いたが, 複雑な証明図であるほど, 証明にかかるコストが高いと考えられる. そこで, 証明に要したステップの数や証明に適用した推論規則を用いて証明可能性を定量化する方法を今後検討する.

## 参考文献

[1] 田中リベカ, 峯島宏次, Pascual Martínez-Gómez, 宮尾祐介, 戸次大介. 日本語 CCG パーザに基づく意味解析・推論システムの提案. 言語処理学会第 22 回年次大会, pp. 757–760, 2016.

[2] Koji Mineshima, Pascual Martínez-Gómez, Yusuke Miyao, and Daisuke Bekki. Higher-order logical inference with compositional semantics.

Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP'15), pp. 2055–2061, 2015.

- [3] Johannes Bjerva, Johan Bos, Rob Goot, and Malvina Nissim. The Meaning Factory: Formal semantics for recognizing textual entailment and determining semantic similarity. *SemEval 2014: International Workshop on Semantic Evaluation*, pp. 642–646, 2014.
- [4] Islam Beltagy, Stephen Roller, Gemma Boleda, Katrin Erk, and Raymond J. Mooney. UTexas: Natural language semantics using distributional semantics and probabilistic logic. *SemEval 2014: International Workshop on Semantic Evaluation*, pp. 796–801, 2014.
- [5] Daisuke Bekki and Koji Mineshima. Context-passing and underspecification in Dependent Type Semantics. In Stergios Chatzikyriakidis and Zhaohui Luo, editors, *Modern Perspectives in Type Theoretical Semantics*, Studies of Linguistics and Philosophy. Springer, 2017.
- [6] Mark Steedman. *The Syntactic Process*. MIT Press, 2000.
- [7] Terence Parsons. *Events in The Semantics of English: a Study in Subatomic Semantics*. MIT Press, 1990.
- [8] Yves Bertot, Pierre Castéran, Gérard Huet, and Christine Paulin-Mohring. *Interactive Theorem Proving and Program Development. Coq'Art: The Calculus of Inductive Constructions*. Springer, 2004.
- [9] Prawitz Dag. *Natural Deduction – A Proof-Theoretical Study*. Almqvist & Wiksell, Stockholm, 1965.
- [10] Robin Cooper et al. Using the framework. Technical Report Technical Report LRE 62-051 D-16, The FraCaS Consortium, 1996.
- [11] Pascual Martínez-Gómez, Koji Mineshima, Yusuke Miyao, and Daisuke Bekki. On-demand injection of lexical knowledge for recognizing textual entailment. Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, (EACL'17), 2017.
- [12] Stephen Clark and James R. Curran. Wide-coverage efficient statistical parsing with CCG and log-linear models. *Computational Linguistics*, Vol. 33, No. 4, pp. 493–552, 2007.
- [13] Mike Lewis and Mark Steedman. A\* CCG parsing with a supertag-factored model. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP'14), pp. 990–1000, 2014.
- [14] Marco Marelli, Stefano Menini, Marco Baroni, Luisa Bentivogli, Raffaella Bernardi, and Roberto Zamparelli. A SICK cure for the evaluation of compositional distributional semantic models. Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), pp. 216–223, 2014.