

# Neural Wani : 依存型理論のための 自動定理証明器 wani の高速化に向けて

宮川夏菜子 大洞日音\* 戸次大介

お茶の水女子大学

{miyagawa.nanako, daido.hinari, bekki}@is.ocha.ac.jp

## 概要

本研究は、依存型理論 (DTT) に基づく自動定理証明器である wani の高速化を目的とし、証明探索クエリを入力として、それを導いた推論規則の確率分布を出力するニューラルモデルを提案する。先行研究では証明項を含む型判定クエリを入力していたが、証明探索への適用には証明項のないクエリに対する予測が不可欠である。本研究で提案するモデルは、実験の結果、低頻度クラスの再現率に課題を残すものの、高頻度クラスにおいて高い適合率を示し、証明探索の高速化に十分な実用性を有するニューラル判定器であることが示された。

## 1 はじめに

計算言語学では、自然言語の構造的な意味に関わる計算過程を明らかにする取り組みが長年行われている。計算言語学の目的の一つは言語現象の背後にある計算機構の解明であるが、手計算で計算過程を明らかにするのは困難である。そこで、計算を自動化するパイプラインシステムの構築が求められてきた。その中で、Tomita et al. [1] は CCG 統語・意味解析器 lightblue<sup>1)</sup> [2] と依存型理論 (Dependent Type Theory: DTT) [3] のための定理自動証明器 wani [4, 5] を組み合わせた自然言語推論システムを提唱・実装した。この自然言語推論システムでは、自然言語の文を入力とし、統語解析、意味解析、型検査、照応・前提解決、自動定理証明を経て、推論の結果を出力する。

### 1.1 wani

上述の自然言語推論システムに組み込まれている定理自動証明器 wani は、証明体系として DTT を採

用している。DTT の型判定 (judgment) は、以下の形式をもつ。

$$\Gamma \vdash M : A$$

$\Gamma$  を環境 (context),  $M$  を項 (term),  $A$  を型 (type) としたとき、上記の型判定は、型理論の観点からは「環境  $\Gamma$  のもとで項  $M$  が型  $A$  を持つ」ことを意味する。同時に、証明論の観点からは「前提  $\Gamma$  から帰結  $A$  に至る証明  $M$  が存在する」ことを意味する (カーリー・ワード対応)。証明項  $M$  が未知である以下のような型判定を証明探索クエリといい、証明探索クエリに対して証明項を探索するタスクを証明探索と呼ぶ。

$$\Gamma \vdash ? : A$$

wani では証明探索を解くために、前向き推論と後ろ向き推論を組み合わせて探索を行う。前向き推論では、「前提から何が導けるか」という点に着目し、上から下へと証明木を構築する。一方、後ろ向き推論では、「与えられた命題を示すには何が必要なのか」という点に着目し、下から上へと証明木を構築する。その際、示したい命題に規則を適用し、それを示すために必要な条件をミニゴールとして割り出す。

### 1.2 依存型意味論 (DTS)

§1節で述べた自然言語推論システムでは、意味表示・意味合成の枠組みとして依存型意味論 (Dependent Type Semantics: DTS) [6, 7] を採用している。DTS は DTT に基づく自然言語の意味の理論である。DTT では、関数型と直積型の一般化として  $\Pi$  型と  $\Sigma$  型を用い、それぞれ  $(x : A) \rightarrow B$ ,  $(x : A) \times B$  と表現する。後件の型  $B$  が前件の型  $A$  に依存できる性質を持つことから、DTS では先行する文の意味に依存した文の意味を表すことができる。これにより、DTS では照応解決・前提束縛といった複雑な言語現象を扱うことが可能となる。

\* 本研究は著者が独自に実施したものであり、Amazon の見解や立場を反映するものではない。

1) <https://github.com/DaisukeBekki/lightblue/>

また、DTS は証明論的意味論であり、文の意味を推論規則を媒介した検証条件として定義している。これは形式意味論における標準的な理論であるモデル理論的意味論と比較して、モデルを参照することなく証明探索のみで推論の成否を計算できるという実装上の利点を持つ。一方で、DTS が基盤とする DTT のような高階論理は、一階述語論理に比べて記述力が強い反面、決定不能性という計算上の課題を抱えている。Coq 等を用いた先行研究 [8] により、実効的な速度での推論が示されつつあるものの、実用的な自然言語推論システムの構築には、wani のような専用の定理証明器のさらなる高速化が不可欠である。そこで、証明探索の効率化により wani の計算時間を短縮するアプローチとして、筆者らは Neural Wani [9] を提案している。

## 2 先行研究

Neural Wani は、DTT の証明探索にニューラルモデルを導入することで、推論の高速化を図る試みである。従来の wani における後ろ向き推論は深さ・時間制限付きの深さ優先探索に基づいており、不適切な推論規則が探索の初期段階で適用されると、計算時間が増大するという課題があった。そこで Neural Wani では、型判定を入力とし、次に適用すべき推論規則をニューラルモデルにより予測する手法をとる。予測確率の高い規則から優先的に探索を行うことで、証明にかかる時間の短縮を目指している。

先行研究 [9] では、Neural Wani の予備的検討として、DTT の型判定に対するベクトル埋め込み手法の比較・評価を行っている。その結果に基づき、DTT の証明探索に適した埋め込み手法を提案した。

## 3 提案手法

### 3.1 先行研究との差異

先行研究 [9] では、証明項が既に存在する型判定を入力とし、推論規則を予測するモデルを構築していた。DTT において証明項は推論規則の適用履歴を反映していることから、証明項が含まれる状態での予測は容易であり、先行研究では micro-F1 スコア 0.982 という高い精度が報告されている。しかし、§1.1 節で述べたように、実際の証明探索（後ろ向き推論）の過程では、証明項は未知であり、証明項を含まない型判定である証明探索クエリから適切な推論規則を導く必要がある。

後ろ向き推論では、現在の型判定に対して推論規則を適用することで、新たな部分ゴールとして、証明項を含まない型判定が生成される。この各段階において規則名を再帰的に予測することで、証明図全体の構築が可能となる。したがって、証明項が未知の状態から規則を予測する本研究の設定は、Neural Wani における効率的な探索を実現するために不可欠なステップである。

### 3.2 実装（型判定の埋め込み）

本研究では、ニューラルモデルに入力する証明探索クエリの埋め込み表現として、「EO~区切り」を採用する。先行研究 [9] において、4 つの埋め込み手法（() 区切り、SEP 区切り、EO~区切り、区切り用トークンなし）を比較した結果、EO~区切りが最も高い micro-F1 スコアを達成している。また、証明項という手がかりがない本タスクは、先行研究よりも複雑であると考えられることから、各要素の役割と境界を明示できる EO~区切りが適していると考えられる。

具体的なデータの構成は以下の通りである。なお、本研究では証明項が未知であることを前提とすることから、入力データから term は除外する。

```
signature = [(entity, type)]
context = [type, entity]
term = var + ← 本研究では除外
type = entity
```

型判定は以下のような形式で埋め込まれる。

```
[Word1, ‘,’ , Type, EOPre, EOSig,
Type, EOPre, Word1, EOPre, EOCon,
Word1, EOPre, EOTyp]
```

### 3.3 ニューラルモデルの構築

本研究では先行研究 [9] と同様、ニューラルモデルを LSTM [10] を用いて構築した。ハイパーパラメータは表 1 の通りである。

表 1 ハイパーパラメータ

エポック数	10
学習率	$5e-4$
レイヤーの数	1
ドロップアウト	なし
インプットサイズ	512
隠れ状態のサイズ	512
ステップ数	32

### ファイル 1 SYN001+1.p の例

```

%-----
% File      : SYN001+1 : TPTP v9.2.1. Released
              v2.0.0.
% Domain    : Syntactic
% ... (ヘッダー情報の省略) ...
%-----

fof(pe12,conjecture,
    ( ~ ~ p
      <=> p ) ).
%-----

```

## 4 評価実験

### 4.1 学習・評価用データセット

本実験では、学習・評価用データとして、TPTP (Thousands of Problems for Theorem Provers) Problem Library<sup>2)</sup> [11] を使用した。具体的には、ファイル 1 のような SYN (Syntactic) ドメインの FOF (first-order form) 問題を対象とし、以下の手順でデータの抽出を行った。

1. 字句解析器生成系 Alex および 構文解析器生成系 Happy を用いて、TPTP の問題を DTT の証明探索問題へと変換する。
2. wani により証明探索を実行し、証明図を出力する。
3. 得られた証明図から、型判定と、適用された規則名のペアを抽出する。
4. 抽出したペアのうち、wani の後ろ向き推論で用いられる規則のみを対象とし、かつ形成則に該当するペアを除外する。

上記の手続きの結果、変換および証明探索に成功した 50 問から、型判定と規則名のペアを計 1,608 件獲得した。実験では、これらを訓練データ (80%)、検証データ (10%)、テストデータ (10%) に分割して使用した。

#### 4.1.1 後ろ向き推論で使われる規則名に絞る理由

Neural Wani の主目的は、証明探索において計算コストの高い箇所をニューラルモデルによって効率化することにある。wani の推論システムにおいて、前向き推論に適用可能な規則は表 2 で示す通り、限定的である。一方、後ろ向き推論の各ステップで適用

可能な規則の候補は表 3 のように多岐にわたる。また、どんな型に対しても制限なく適用できる規則 (( $\Pi E$ ) や ( $DNE$ ) 等) があることによって探索空間が爆発的に広がりやすい。探索空間が広く、計算時間のボトルネックとなりやすいのはこの後ろ向き推論のステップであることから、本研究では後ろ向き推論で用いられる規則の予測に焦点を当てる。

表 2 前向き推論の規則一覧

(Membership)  
( $\Sigma E$ )  
( $=I$ )  
( $=E$ )

表 3 後ろ向き推論の規則一覧

( $\Pi I$ ) ( $\Sigma I$ ) (Membership)  
( $\Pi E$ ) ( $\Sigma F$ ) ( $DNE$ )  
( $\Pi F$ ) ( $\top I$ ) ( $EFQ$ )  
( $=F$ )

#### 4.1.2 形成則を除外する理由

DTT の形成則は、以下のような形で表される。

$$\frac{\overline{x : A}^i}{\vdots} \frac{A : s_1 \quad B : s_2}{(x : A) \rightarrow B : s_2} (\Pi F), i$$

where  $(s_1, s_2) \in \{(type, type), (type, kind)\}$

図 1 ( $\Pi F$ ) の例

$$\frac{A : type \quad M : A \quad N : A}{M =_A N : type} (=F)$$

図 2 ( $=F$ ) の例

形成則の下段 (結論部) に現れる型判定は、その型として type もしくは kind を持つ。ある型判定がこれらの型を持つ場合、適用すべき規則は (( $CON$ ) などの例外を除き) 構文的に一意に決定される。すなわち、形成則の適用は決定的であり、ニューラルモデルによる確率的な予測を必要としないことから、本実験の対象からは除外した。

2) <http://www.tptp.org/>

## 4.2 実験結果

表4 分類性能の全体評価

	Prec	Rec	F1	Supp
micro	0.863	0.863	0.863	161
macro	0.880	0.432	0.446	161
w-avg	0.867	0.863	0.846	161

表5 ラベルごとの分類性能

	Prec	Rec	F1	Supp
(Membership)	0.894	0.944	0.918	125
( $\Pi I$ )	0.800	0.444	0.571	9
( $\Pi E$ )	0.708	0.773	0.739	22
( $\Sigma I$ )	1.000	0.000	0.000	4
(DNE)	1.000	0.000	0.000	1

## 4.3 考察

表4に示す通り、全体評価において micro-F1 スコアは 0.863 という高い値を示した一方で、macro-F1 スコアは 0.446 に留まった。この乖離は、本実験に用いたデータセットにおけるラベル分布の不均衡に起因するものである。micro-F1 スコアは全サンプルに対する正解率を反映することから、多数派クラスの性能に強く依存する。これに対し、各クラスを均等に評価する macro-F1 スコアが低値であることは、少数派クラスの識別において課題が残っていることを示唆している。

しかしながら、本研究の最終目標である Neural Wani による証明探索の高速化という観点からは、今回の実験結果は以下に述べる二つの理由で許容し得る。

第一に、少数クラスにおける適合率 (Precision) の高さである。表5によれば、( $\Pi I$ ) の適合率は 0.800 であり、( $\Sigma I$ ) や (DNE) も (Recall は 0 であるものの) 不適切な予測は限定的である。Neural Wani のアルゴリズムでは、モデルが出力した確率に基づいて適用規則の順序を決定する。少数クラスに対して「確信度が高い場合にのみ予測する」という保守的な性質は、探索において誤った規則を優先的に選択するリスクを低減させ、結果として探索回数を抑制する。

第二に、micro-F1 スコア (0.863) の高さが全体の探索効率に与える影響である。DTT の証明探索にお

ける型判定の大部分は、出現頻度の高い (Membership) や ( $\Pi E$ ) によって構成される。これらの主要な規則において高い識別性能を維持できていることは、探索ステップの大部分において、モデルが最適な推論規則を第一候補として提示できることを意味する。これは、深さ優先探索における平均的な計算時間の短縮に直結する。

したがって、本モデルは「汎用的な分類器」としては改善の余地があるものの、「証明探索の高速化の補助ツール」としては十分に実用的な性能を有する可能性がある。

## 5 おわりに

本研究では、DTT の証明項を含まない型判定から次に適用すべき推論規則を予測する LSTM モデルを構築し、その性能を評価した。実験の結果、データ不均衡に起因する少数クラスの識別能力に課題を残したものの、全体として 0.863 の高い micro-F1 スコアを達成した。これは、証明探索の大部分のステップにおいて適切なガイドが可能であることを示唆している。今後の展望としては、本モデルを実際に定理自動証明器 wani の探索アルゴリズムに組み込み、実際の計算時間の短縮率を定量的に評価することが挙げられる。

## 謝辞

本研究の一部は JST CREST JPMJCR2565 および JSPS 科研費 JP23H03452 の支援を受けたものである。

## 参考文献

- [1] Asa Tomita, Mai Matsubara, Hinari Daido, and Daisuke Bekki. Natural language inference with CCG parser and automated theorem prover for DTS. In Proceedings of the Workshop on the Bridges and Gaps between Formal and Computational, pp. 1–7, 2025.
- [2] Daisuke Bekki and Ai Kawazoe. Implementing variable vectors in a CCG parser. Logical Aspects of Computational Linguistics (LACL 2016), pp. 52–67, 2016.
- [3] Per Martin-Löf. Intuitionistic type theory, Vol. 9. Bibliopolis Naples, 1984.
- [4] Hinari Daido and Daisuke Bekki. Development of an automated theorem prover for the fragment of DTS. In the 17th International Workshop on Logic and Engineering of Natural Language Semantics (LENLS17), 11 2020.
- [5] 大洞日音. DTS の部分体系を用いた定理自動証明器への等号型の導入, 2022. お茶の水女子大学, 修士論文.
- [6] Daisuke Bekki. Representing anaphora with dependent types. In N. Asher and S. Soloviev, editors, Logical Aspects of Computational Linguistics - 8th International Conference, LACL 2014, Toulouse, France, June 18-20, 2014. Proceedings, Vol. 8535 of Lecture Notes in Computer Science, pp. 14–29. Springer, 2014.
- [7] 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, pp. 11–41. Springer, Cham, 2017.
- [8] Pascual Martínez-Gómez, Koji Mineshima, Yusuke Miyao, and Daisuke Bekki. ccg2lambda: A compositional semantics system. In Proceedings of ACL 2016 System Demonstrations, pp. 85–90, Berlin, Germany, August 2016.
- [9] 宮川夏菜子, 田上青空, 戸次大介. 依存型理論のための自動定理証明器 Neural Wani の開発に向けて. 第 39 回人工知能学会全国大会論文集, No. 3G5-GS-6-04, 5 2025.
- [10] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. Neural Computation, Vol. 9, No. 8, pp. 1735–1780, 1997.
- [11] Geoff Sutcliffe. The tptp problem library and associated infrastructure: from cnf to th0, tptp v6. 4.0. Journal of Automated Reasoning, Vol. 59, No. 4, pp. 483–502, 2017.