

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

宮川 夏菜子 (指導教員: 戸次大介)

1 はじめに

本研究で扱う理論言語学に基づいた自然言語推論のパイプライン [1] では、日本語 CCG パーザ (統語解析器) lightblue [2] と依存型理論 (Dependent Type Theory; DTT) [3] の部分体系における自動定理証明器 wani [4, 5] を接続している。lightblue は自然言語の文を入力として受け取り、統語解析を行った結果として意味表示を出力する。そこから得られた意味表示を wani に入力することで、照応・前提解決、推論を行う。

本研究では、パイプライン中の定理証明器における証明探索をニューラルネットワークによって補助することで、高い説明力と高速性を備えたシステム Neural Wani を構築することを目指す。本稿ではその Neural Wani 開発に向けて、前段階となるモデルを構築し、定理証明器とニューラルネットワークとの融合方法について検討した。

2 自動証明器 wani の証明探索

wani による証明探索は、lightblue の照応・前提解決、および推論において呼び出される。以下、証明探索の説明にあたり、型判定、規則について説明したのち、証明探索について述べる。

型判定は「 $\Gamma \vdash M : A$ 」という形式で表される。これは「環境 Γ のもとで項 M は型 A を持つ」という意味である。

そして、規則は型判定を導出するものである。図 1 の (Var) 規則のように、前提となる型判定がない状態で規則を適用し、新たに型判定を導出できる規則がある。一方、図 1 の (ΣE) 規則のように、既にある型判定に規則を適用することで、新たな型判定を導出するものもある。

型判定「 $\Gamma \vdash M : A$ 」は、論理の視点からは「前提 Γ から命題 A に至る証明 M が存在する」とも解釈できる。また、 Γ から A に至る証明を探索する問題「 $\Gamma \vdash ? : A$ 」は証明探索クエリと呼ばれる。実際の証明探索の例を見る。図 2 の「?」部分を求めることが証明探索にあたり、この探索結果が図 1 である。下段の $\pi_1(s_0)$ は証明探索によって求められた証明項であり、前提部から命題が導き出された過程を示す。

$$\frac{s_0 : \left[\begin{array}{l} x_0:\text{entity} \\ x_1:\text{entity} \\ \text{木/き}(x_1, x_0) \end{array} \right], s_1 : \text{entity} \vdash s_0 : \left[\begin{array}{l} x_0:\text{entity} \\ x_1:\text{entity} \\ \text{木/き}(x_1, x_0) \end{array} \right]}{s_0 : \left[\begin{array}{l} x_0:\text{entity} \\ x_1:\text{entity} \\ \text{木/き}(x_1, x_0) \end{array} \right], s_1 : \text{entity} \vdash \pi_1(s_0) : \text{entity}} \quad (\Sigma E)$$

図 1: 証明探索の結果の例

$$s_0 : \left[\begin{array}{l} x_0:\text{entity} \\ x_1:\text{entity} \\ \text{木/き}(x_1, x_0) \end{array} \right], s_1 : \text{entity} \vdash ? : \text{entity}$$

図 2: 証明探索クエリの例

wani は DTT の部分体系に基づく自動定理証明器であり、証明探索を行う。wani での証明探索は枝刈りを活用した深さ優先探索であり、深さ制限と時間制限が設定されている。大洞 [5] によると、制限を緩めることで探索の成功率が上昇している。しかし制限を設けなければ一部の証明では証明探索が終了しない。

DTT は一階述語論理を超えた記述力を持つため、DTT の推論は決定不能である。このことは wani が DTT の証明探索を行う上で避けられない課題である。

3 提案手法

本研究の目指すところは、ニューラルモデルで wani の証明探索を補助する Neural Wani の実装である。Neural Wani では、与えられた型判定を導出した規則をニューラルネットワークを用いて予測することによって、wani の証明探索を高速化する。ニューラルモデルは wani とは相補的な特徴を持つ。wani では探索過程における最悪計算量が大きいのが、規則に基づくため正確である。一方、本研究で提案するニューラルモデルでは探索を行わずニューラルネットワークによって予測することで高速に結果を返すものの、正しい結果を出すことは必ずしも保証されない。

本研究では、Neural Wani 実装の第一歩として、証明項がわかっている状態の型判定から導出した規則を予測するという分類問題を解くモデルを構築した。これは Neural Wani の実装の部分問題として位置づけられる。

4 評価

本研究では上述のニューラルモデルを LSTM[6] を用いて構築した。LSTM の入力である型判定は階層構造を持ち、埋め込みには複数の手法が考えられる。埋め込み方による予測性能の差を比較するため、型判定の埋め込み手法を 4 種類設定し比較実験を行った。

1. 要素の始めと終わりそれぞれに区切り用トークンとして開き括弧と閉じ括弧を用いる。
2. 要素の終わりに SEP という区切り用トークンを用いる。
3. 要素の終わりに各要素の特徴を表す詳細な区切り用トークンを用いる。(例: signature の要素の終わりに EOSig (End of Signature) の区切り用トークンを用いる。)
4. 区切り用トークンを用いない。

これら 4 パターンの分割方法をそれぞれ実装し、分割方法によって評価を行う。

実験に用いるデータには、wani の証明探索の結果を用いる。本研究では以下の 2 つのデータセットを使用した。1 つ目は定理自動証明器の評価のための包括的なライブラリである TPTP library [7] のデータを wani で処理したものを使用した。2 つ目は、日本語推論テストセット JSeM [8, 9] を lightblue で処理したものをを用いた。

本実験では、規則ごとに¹450個のデータを集め、training データに80%、validation データに10%、test データに10%を用いた。

表1のとおりハイパーパラメータを設定した。機械学習フレームワークとしては²Hasktorch を用いた。

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 | Nothing |
| インプットサイズ | 256 |
| 隠れ状態のサイズ | 256 |
| ステップ数 | 32 |

4.1 実験結果

| | Prec | Rec | F1 | Supp | | Prec | Rec | F1 | Supp |
|-------|-------|-------|-------|------|-------|-------|-------|-------|------|
| micro | 0.963 | 0.963 | 0.963 | 379 | micro | 0.953 | 0.953 | 0.953 | 379 |
| macro | 0.963 | 0.894 | 0.886 | 379 | macro | 0.966 | 0.793 | 0.794 | 379 |
| w-avg | 0.965 | 0.963 | 0.961 | 379 | w-avg | 0.954 | 0.953 | 0.947 | 379 |

表2: () 区切り

| | Prec | Rec | F1 | Supp | | Prec | Rec | F1 | Supp |
|-------|-------|-------|-------|------|-------|-------|-------|-------|------|
| micro | 0.982 | 0.982 | 0.982 | 379 | micro | 0.966 | 0.966 | 0.966 | 379 |
| macro | 0.987 | 0.907 | 0.933 | 379 | macro | 0.975 | 0.934 | 0.948 | 379 |
| w-avg | 0.982 | 0.982 | 0.981 | 379 | w-avg | 0.966 | 0.966 | 0.965 | 379 |

表4: EO~区切り

表3: SEP 区切り

| | Prec | Rec | F1 | Supp | | Prec | Rec | F1 | Supp |
|-------|-------|-------|-------|------|-------|-------|-------|-------|------|
| micro | 0.966 | 0.966 | 0.966 | 379 | micro | 0.966 | 0.966 | 0.966 | 379 |
| macro | 0.975 | 0.934 | 0.948 | 379 | macro | 0.975 | 0.934 | 0.948 | 379 |
| w-avg | 0.966 | 0.966 | 0.965 | 379 | w-avg | 0.966 | 0.966 | 0.965 | 379 |

表5: 区切り用トークンなし

4.2 考察

micro-F1 スコアはEO~区切りの場合に0.982を達成した。EO~区切りが最も高いmicro-F1スコアを示した理由として、詳細な区切り用トークンの情報を学習できたことが挙げられる。()区切りやSEP区切りは区切り用トークンがノイズになってしまった可能性がある。

他の指標と比べてmacro-F1スコアが低い原因としては、全体で14個しかない規則(SigmaI, IqF)などで学習がうまくいかず、recallが0.5となることがあった影響が考えられる。しかしながら、IqE規則はデータ数が42個と少ないにもかかわらず、F1スコアがトークンの区切り方によらず0.889~1に収まっていることから、極端にデータ数が少ない規則を除き、安定した結果が得られたことがわかる。

5 おわりに

Neural Waniの実装の第一歩として型判定から規則を予測するニューラルネットワークを構築した。

Neural Waniを実現するための次のステップとして、証明項がない状態で規則を予測することが挙げられる。証明探索ではまず初めに証明項のない型判定に対し規則を予測することが求められる。これは現状の全ての要素がある型判定から規則を予測する問題とはギャップがある。

Neural Waniを実現することができれば、探索できる問題数を増やすことができると考えられる。型判定

から規則を予測するニューラルモデルを導入することで適切な規則から優先的に探索を開始できるようになる。今後、ニューラルモデルの組み込みによる探索時間の変化を検証していきたい。

謝辞 本研究の一部は、JST CREST JPMJCR20D2の支援を受けたものである。

参考文献

- [1] 富田朝, 大洞日音, 戸次大介. CCG 統語解析器 light-blue と自動定理証明器 wani による自然言語推論の試み. 人工知能学会全国大会論文集 (掲載予定), 2025.
- [2] Daisuke Bekki and Ai Kawazoe. Implementing variable vectors in a CCG parser. *Logical Aspects of Computational Linguistics (LACL 2016)*, Vol. Proceedings, 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)*, 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. Master's thesis, お茶の水女子大学, 2022.
- [6] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation MIT-Press*, 1997.
- [7] 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.
- [8] Ai Kawazoe, Ribeka Tanaka, Koji Mineshima, and Daisuke Bekki. A framework for constructing multilingual inference problem. *Proceedings of 1st International Workshop on the Use of Multilingual Language Resources in Knowledge Representation Systems (MLKRep2015)*, 2015.
- [9] Ai Kawazoe, Ribeka Tanaka, Koji Mineshima, and Daisuke Bekki. An inference problem set for evaluating semantic theories and semantic processing systems for japanese. In *LENLS12*, pp. 67-73, 2015.

¹一部の規則についてはデータ数が450個に満たない場合があり、その場合は利用可能な全てのデータを用いた。

²<https://github.com/hasktorch/>

付録

A 各規則の評価指標

| | Prec | Rec | F1 | Supp |
|--------|-------|-------|-------|------|
| Var | 0.977 | 0.933 | 0.955 | 45 |
| Con | 0.935 | 0.956 | 0.945 | 45 |
| EnumF | 0.978 | 1.000 | 0.989 | 45 |
| PiE | 0.953 | 0.911 | 0.932 | 45 |
| TypeF | 1.000 | 1.000 | 1.000 | 24 |
| PiF | 1.000 | 0.933 | 0.966 | 45 |
| SigmaF | 0.882 | 1.000 | 0.938 | 45 |
| PiI | 1.000 | 1.000 | 1.000 | 31 |
| SigmaI | 1.000 | 0.000 | 0.000 | 2 |
| IqE | 0.833 | 1.000 | 0.909 | 5 |
| SigmaE | 1.000 | 1.000 | 1.000 | 45 |
| IqF | 1.000 | 1.000 | 1.000 | 2 |

表 6: () 区切り

| | Prec | Rec | F1 | Supp |
|--------|-------|-------|-------|------|
| Var | 0.978 | 1.000 | 0.989 | 45 |
| Con | 1.000 | 1.000 | 1.000 | 45 |
| TypeF | 1.000 | 1.000 | 1.000 | 24 |
| PiF | 1.000 | 0.978 | 0.989 | 45 |
| SigmaF | 0.938 | 1.000 | 0.968 | 45 |
| PiI | 0.969 | 1.000 | 0.984 | 31 |
| PiE | 1.000 | 0.911 | 0.953 | 45 |
| SigmaE | 0.957 | 1.000 | 0.978 | 45 |
| SigmaI | 1.000 | 0.500 | 0.667 | 2 |
| EnumF | 1.000 | 1.000 | 1.000 | 45 |
| IqF | 1.000 | 0.500 | 0.667 | 2 |
| IqE | 1.000 | 1.000 | 1.000 | 5 |

表 8: E0~区切り

| | Prec | Rec | F1 | Supp |
|--------|-------|-------|-------|------|
| Var | 0.978 | 1.000 | 0.989 | 45 |
| Con | 0.938 | 1.000 | 0.968 | 45 |
| TypeF | 1.000 | 1.000 | 1.000 | 24 |
| PiF | 0.917 | 0.978 | 0.946 | 45 |
| SigmaF | 0.953 | 0.911 | 0.932 | 45 |
| PiI | 0.912 | 1.000 | 0.954 | 31 |
| PiE | 0.889 | 0.889 | 0.889 | 45 |
| SigmaI | 1.000 | 0.000 | 0.000 | 2 |
| SigmaE | 1.000 | 0.956 | 0.977 | 45 |
| EnumF | 1.000 | 0.978 | 0.989 | 45 |
| IqF | 1.000 | 0.000 | 0.000 | 2 |
| IqE | 1.000 | 0.800 | 0.889 | 5 |

表 7: SEP 区切り

| | Prec | Rec | F1 | Supp |
|--------|-------|-------|-------|------|
| Var | 1.000 | 0.978 | 0.989 | 45 |
| SigmaE | 0.978 | 0.978 | 0.978 | 45 |
| Con | 1.000 | 0.978 | 0.989 | 45 |
| PiF | 0.933 | 0.933 | 0.933 | 45 |
| TypeF | 1.000 | 1.000 | 1.000 | 24 |
| EnumF | 0.957 | 1.000 | 0.978 | 45 |
| SigmaF | 0.933 | 0.933 | 0.933 | 45 |
| PiI | 0.967 | 0.935 | 0.951 | 31 |
| PiE | 0.936 | 0.978 | 0.957 | 45 |
| SigmaI | 1.000 | 1.000 | 1.000 | 2 |
| IqF | 1.000 | 0.500 | 0.667 | 2 |
| IqE | 1.000 | 1.000 | 1.000 | 5 |

表 9: 区切り用トークンなし