

病名知識の公理補完を用いた日本語症例テキスト間の論理推論

村上夏輝（指導教員：戸次 大介）

1 はじめに

医療分野には画像所見や電子カルテといった電子化されたテキストが多く存在しており、これらを用いた医療自然言語処理の研究は日本語でも活発に行われつつある [5, 7]. その一方で、それらの研究の多くは機械学習を用いたものであり、機械学習を用いたアプローチは否定や量化といった深い意味を考慮することに課題があると言われている. これらの複雑な言語現象を伴う体型的な推論を行うために、論理ベースの手法が提案されてきており、その一つに症例テキストの含意関係認識の研究である意味解析・論理推論システム Medc2l [6] がある. 日本語症例テキストには複合語が多く含まれており、複合語内の言語現象を分析する手法が求められていた. Medc2l は、組合せ範疇文法 (Combinatory Categorical Grammar, CCG)[4] に基づく高精度な構文解析と、高階論理に基づく自動推論システムを組み合わせた高度な意味解析・推論システムである ccg2lambda [6] をベースとした推論システムである. Medc2l では ccg2lambda を拡張し、複合語の解析を可能にすることで、ccg2lambda を医療テキストへの応用可能にした. しかし、既存の論理ベースの推論システムでは、入力文に病名などの医療ドメイン特有の言い換え表現が含まれていた場合、含意関係を正しく判定することができなかった. そこで本研究では、医療ドメインの知識補完モジュールを推論システムに追加し、定理証明の際に不足している病名の知識を補完する手法を提案する. 具体的には、病名の固有表現抽出器と万病辞書 [2] を用いて定理証明に不足する病名の言い換え表現の候補を特定する. 特定した病名に付与される意味現象タグの組み合わせに応じた公理を生成することで、病名の言い換え知識を公理として補完する手法を提案する. また、病名の知識補完が必要な推論テストセットを構築し、提案手法の有効性を評価する.

2 提案手法

提案手法の全体像を図 1 に示す. 本研究では、医療ドメインの知識補完を行うモジュールを Medc2l に追加する手法により、病名の知識補完を可能にする.

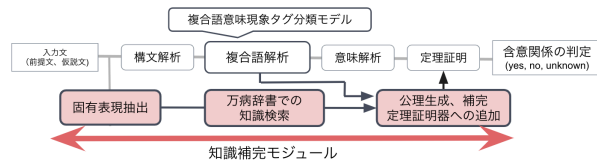


図 1: 提案手法の全体像

知識補完モジュールの概要を説明する. 知識補完モジュールでは、まず入力文に対して、病名の固有表現抽出器を用いて病名の抽出を行う. 抽出された病名に対し、万病辞書を用いて検索を行う. 病名と出現形が一致すれば、万病辞書内の病名に関する知識を公理として定理証明支援系 Coq [1] に追加する. この際に、追

加する公理の形は形態素間の意味関係によって異なるため、石田らの複合語意味現象タグ分類モデルによって付与される複合語の意味現象タグを確認し、意味現象タグと万病辞書の知識を用いて公理を導出してから補完を行う. 導出された公理を用いて、定理証明支援系 coq は前提文の意味表示と仮説文の意味表示間の含意関係の有無の証明を行う.

2.1 病名の固有表現抽出

2.1.1 病名の固有表現抽出データセットの構築

固有表現抽出器の学習を行うために、症例テキストを用いて病名の固有表現抽出データセットを構築した. 症例テキストには、J-Stage でオープンアクセス公開されている症例報告論文 PDF から OCR 抽出した症例報告コーパスである J-MedStd-CR¹ を用いる. J-MedStd-CR に含まれる 2629 文のうち、2303 文を学習用のデータセットとしてアノテーションをし、残りの 326 文は評価用データセットとして用いている. アノテーションは、文中に現れる万病辞書の出現形と、テキストのどの範囲に出現形が含まれているかを表すスパンを人手で付与した. 以下にアノテーションの例を示す.

```
1 "text" : " 切断前から両側下肢に筋萎縮を認めた",
2 "entities": [{
3     "name" : " 筋萎縮",
4     "span" : [10,13],
5     "type" : " 万病辞書"}]
```

病名のアノテーションは、最大文字数のもののみを万病辞書の出現形として採用している.

症例テキストのうち、病名の出現形が含まれていた文は 1345 文であった. 1345 文には 1181 種類の病名の出現形が 2551 件出現していた.

2.1.2 固有表現抽出器の構築・実験

固有表現抽出器を構築するにあたって、表 1 に示す 3 つの日本語 RoBERTa[3] を用いて固有表現抽出器を構築し、実験を行った. 各汎用言語モデルの事前学習コーパス、トークナイザーも合わせて表 1 に示す.

表 1: 実験に使用する RoBERTa モデル

モデル	事前学習コーパス	Tokenizer
Japanese RoBERTa base(理研)	Wikipedia	MeCab + BPE
japanese-roberta-base(rinna)	Wikipedia + CC-100	Juman++ + sentencepiece
roberta-large-japanese(早稲田)	Wikipedia + CC-100	Juman++ sentencepiece

固有表現抽出器は、症例テキスト内の 2303 文を学習データとして用い、そのうち 85%を開発データ、15%を検証データとして汎用言語モデルのファインチューニングを行い構築した. 評価データには、同じく J-MedStd-CR の症例テキスト内からランダムで抽出した 326 文を用いた. この 326 文は学習データとの重複

¹<https://sociocom.naist.jp/medtxt/cr/>

はない。

表 2: 病名の固有表現抽出の実験結果

モデル	理研モデル	rinna モデル	早稲田モデル
accuracy	97.2%	96.4%	97.0%
病名の fl-score	81.0%	77.3%	74.6%

実験結果を 2 に示す。病名の予測精度としては理研モデルのスコアが一番高かった。

この実験を経て、提案手法に組み合わせる固有表現抽出器には最も性能の良かった理研モデルを採用した。

2.2 公理補完

本研究では、定理証明に不足する言い換え表現の候補を特定したのちに、病名の言い換え知識を公理として定理証明支援系 coq に補完し、証明に用いる。公理補完においては、病名の言い換え知識を Medc21 における意味表示に合わせて公理として生成する必要がある。提案手法では、出現形と標準病名それぞれに付与される意味現象タグの組み合わせによって生成する公理を定義することで、公理補完を可能にした。Medc21 の複合語意味現象タグ分類モデルで付与される意味現象タグの一部を表 3 に示す。

表 3: 意味現象タグの例

意味現象タグ	意味現象	例
EN	エンティティ	DVT
MLEN	修飾語	隆起性 病変
PA	体の部位	脳 梗塞
GA	ガ格名詞句	胸水 貯留
WO	ヲ格名詞句	骨髄 抑制
NI	ニ格名詞句	骨 転移
EV	イベント	胸水 貯留

表 4: 公理補完を行った病名のタグの組み合わせ

公理	病名の例	意味現象タグ
出現形 ⇒ 標準病名	出現形 ⇒ 標準病名	出現形 ⇒ 標準病名
* ⇒ *	歩行 不能 ⇒ 歩行 困難	EV EN ⇒ EV EN
MLEN * ⇒ *	慢性 偽性 腸閉塞症 ⇒ 偽性 イレウス	MLEN MLEN EN ⇒ MLEN EN
PA * ⇒ *	脳 血管炎 ⇒ 脳動脈炎	PA EN ⇒ EN
PA * ⇒ MLEN *	血流 障害 ⇒ 末梢循環 障害	PA EN ⇒ MLEN EN
MLEN * ⇒ PA *	びまん性 色素 沈着 ⇒ 皮膚 色素 沈着	MLEN GA EV ⇒ PA GA EV
GA EV ⇒ WO EV	血沈 亢進 ⇒ 赤沈 促進	GA EV ⇒ WO EV
EN ⇒ MLEN+ EN	PBC ⇒ 原発性 胆汁性 肝硬変	EN ⇒ MLEN MLEN EN
EN ⇒ PA EN	PE ⇒ 肺 塞栓症	EN ⇒ PA EN
EV ⇒ EN	肝線維化 ⇒ 肝線維腫	EV ⇒ EN

表 4 に、公理補完を行った病名のタグの組み合わせとその例を示す。この公理補完は事前に公理補完の組み合わせについて調査を行い、その調査結果に合わせて公理補完を可能にした。* は、出現形と標準病名それぞれに付与される意味現象タグの組み合わせが一致していることを表している。生成した公理は、自動証明の際に呼び出し可能な公理として定理証明器に追加し、推論に用いる。

3 評価実験

3.1 公理補完が必要な推論テストセットの構築

本研究では、病名の知識補完が必要なテストセットを構築し、提案手法の評価を行う。表 5 に推論テストセットの例を示す。テストセットの構築には、症例報告コーパスである J-MedStd-CR 内の病名が現れる文のうち、万病辞書の病名の出現形と標準病名が異なる

文を用いる。簡略化した文を仮説文、仮説文の病名箇所を標準病名に変えたものを前提文とした文ペアを人手で作成し、合計 149 ペアからなる推論テストセットを構築した。

表 5: 推論テストセットの例

前提文	仮説文
病理診断は t u b i であった。	⇒ 病理診断は癌であった。
V A P には有効である。	⇒ 人工呼吸器関連肺炎には有効である。
患者はホーマンズ徴候陽性であった。	⇒ 患者は深部静脈血栓症陽性であった。

(太字箇所は推論に必要な病名知識)

3.2 実験結果

構築した 149 件の推論テストセットを用いて、石田らのシステムと、石田らのシステムに知識補完モジュールを追加した提案システムとで推論の精度比較を行った。表 6 に推論の評価結果を示す。石田らのシステムでは正しく推論を行える例がなかったのに対して、提案システムでは 149 件中 106 件のテストセットについて正しく推論を行うことができた。

表 6: 推論の評価結果

推論システム	精度
石田らのシステム	0/149 (0.0%)
石田らのシステム + 知識補完モジュール	106/149 (71.1%)

4 おわりに

本研究では、Medc21 で提案された意味解析・論理推論システムに対して、医療ドメインの知識補完モジュールを追加し、推論システムを構築した。具体的には、病名の固有表現抽出を行ったのち、抽出された病名に対して万病辞書を用いて知識検索を行い、検索結果に基づいた公理を石田らのシステムの定理証明器に追加するという知識補完モジュールを作成した。また、固有表現抽出を行うモデルの学習のために病名の固有表現抽出データセット、医療ドメインの知識が補完された推論システムの評価を行うための推論テストセットを構築した。構築したテストセットを用いて推論システムの評価を行った結果、推論テストセット 149 件のうち、106 件について推論を正しく行うことができ、提案手法の有効性を示した。

謝辞: 本研究の一部は、政策科学総合研究事業（臨床研究等 ICT 基盤構築・人工知能実装研究事業）21AC5001 および JST さきがけ JPMJPR21C8、JSPS 科研費 JP20K19868 の支援を受けたものである。

参考文献

- [1] Bertot, Y. and Castéran, P.: *Interactive Theorem Proving and Program Development: Coq/Art: The Calculus of Inductive Constructions*, Springer Science & Business Media (2013).
- [2] Ito, K., Nagai, H., Okahisa, T., Wakamiya, S., Iwao, T. and Aramaki, E.: J-MeDic: A Japanese Disease Name Dictionary based on Real Clinical Usage, in Calzolari, N., Choukri, K., Cieri, C., Declerck, T., Goggi, S., Hasida, K., Isahara, H., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., Piperidis, S. and Tokunaga, T. eds., *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan (2018), European Language Resources Association (ELRA).
- [3] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. and Stoyanov, V.: RoBERTa: A Robustly Optimized BERT Pretraining Approach, *ArXiv*, Vol. abs/1907.11692, (2019).
- [4] Steedman, M. J.: *The Syntactic Process*, The MIT Press (2000).
- [5] 荒牧英治, 岩尾友秀, 若宮翔子, 伊藤薫, 矢野憲, 大江和彦症例検索システムの試行運用に基づいた利用状況に関する基礎的検討, *医療情報学*, Vol. 38, No. 4, pp. 245-256 (2018).
- [6] 石田真穂, 谷中暁, 戸次大介日本語症例テキストの複合語解析・推論システム Medc21, *自然言語処理*, Vol. 30, No. 3, pp. 935-958 (2023).
- [7] 土井俊祐, 木村隆, 関根正樹, 鈴木隆弘, 高林克己, 田村俊世学会ホームページにおける類似症例検索システムの実運用と評価, *生体医工学*, Vol. 49, No. 6, pp. 870-876 (2011).