日本語症例テキストの複合語解析・推論システム Medc2l

石田 真捺（指導教員：戸次 大介）

1 はじめに

医療分野には電子カルテや退院サマリといった症例テキストが蓄積されており，これらを新たな知識の発見に繋げるために，自然言語処理技術を応用する研究が試みられている [3, 5, 7, 1, 10, 13, 6, 12]．一方で，日本語の症例テキストを用いた否定や量化といった構成素の構造を考慮した高度な意味解析については発展途上である．その理由の一つとして，日本語の症例テキストに多く含まれている複合語の構文解析や意味解析が難しいという問題がある．

複合語を含む症例テキストの例を (1) に示す．

(1) 非持続性心室頻拍が認められたため，アミオダロン併用した．

(2) a. 心室頻拍は持続性ではない．
b. アミオダロンを併用した．

(1) の「非持続性心室頻拍」からは，「持続性ではない心室頻拍」が認められたこと，「アミオダロン併用」からは「アミオダロンを併用した」ことがわかる．このように複合語には，複合語を構成する要素間の様々な意味関係が非明示的に含まれている．これらの意味関係を同定することができれば，複合語が現れる (1) のような文から，複合語が現れない (2a) や (2b) のような文への含意関係が認識可能となる．

そこで本論文では，ccg2lambda [4] を改良して，症例テキストの高度な意味解析と推論を扱える推論システム Medc2l を提案する．具体的には，ccg2lambda に複合語解析モジュールを追加することで，複合語を含む症例テキストに対して頑健に意味解析と推論ができるようにする．

2 症例テキストの意味解析・推論システム

ccg2lambda を用いた症例テキストの意味解析・推論システムの全体像を図 1 に示す．

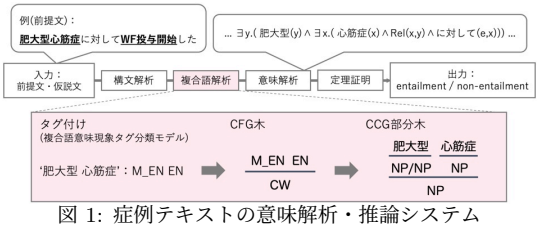

図 1: 症例テキストの意味解析・推論システム

入力となる前提文・仮説文は，日本語形態素解析器 Janome[1]によって形態素解析が行われ，CCG 構文解析器 depccg [9] によって，CCG 構文木に変換される．次に，変換された CCG 構文木の品詞タグを確認し，連続した名詞もしくは接頭辞の最大列からなる部分木を複合語として抽出する．抽出した複合語の表層形を複合語意味現象タグ分類モデルに渡し，複合語内の形態素間の意味関係を表す意味現象タグを付与する．複合語意味現象タグ分類モデルには，複合語内の形態素の表層の系列を意味現象タグの系列に変換する系列ラベリングモデルを用いる．そして，意味現象タグを終端記号とする文脈自由文法 (CFG) に基づいて，各複合語の CFG 木を構築する．設計した CFG 規則を表 1 に示す．CFG 木を CCG 木へ変換する規則群を定義することで，各複合語の CFG 木は CCG 部分木に変換される．その後，CCG 構文木内の各複合語を，上の手順で得た CCG 部分木に置き換えたのち，CCG 木において語に意味情報を割り当てる意味テンプレートに基づいて，ラムダ計算を用いて意味合成を行う．表 2 に意味テンプレートを示す．この過程によって得られた意味表示（論理式）の対に対し，定理証明支援系である Coq [2] を用いて含意関係を判定する．

表 1: CFG 規則

CW	→	EN | T_EN | Q_EN | E_EN | EV | EV' | S_EV
NEG1	→	NEG
NEG2	→	NEG
M_EN	→	M_EN | M_EN' M_EN | PA' M_EN | NEG1 M_EN
M_EN'	→	M_EN | PA' M_EN' | NEG1 M_EN'
PA	→	PA | PA' PA' | M_EN PA
PA'	→	PA
WO	→	WO | M_EN WO' | PA WO' | EV_wo WO'
WO'	→	WO | M_EN WO' | PA WO'
NI	→	NI | M_EN NI | PA NI
GA	→	GA | M_EN GA | PA GA
EV	→	EV | GA EV | WO EV | NI EV | EV' EV' | NEG2 EV
EV'	→	EV
EV_wo	→	WO' EV' | NI EV' | GA EV'
S_EV	→	S_EV | EV S_EV | NEG2 S_EV
EN	→	EN | S_EV EN | EV EN | M_EN EN | PA EN | GA EN
Q_EN	→	Q_EN | EV Q_EN
T_EN	→	T_EN | EV T_EN
E_EN	→	E_EN | S_EV E_EN | EV E_EN

表 2: 複合語に関する意味テンプレート

非終端記号	統語範疇	意味表示
EV EV_wo S_EV	$S[ev]$	$\lambda K.\exists e.(K(Surf,e))$
EV'	$S[ev]/S[ev]$	$\lambda S.\lambda K.(\exists e.(K(Surf,e)) \wedge S(K))$
GA	$S[ga]/S[ev]$	$\lambda S.\lambda K.\exists x.(Surf(x) \wedge S(\lambda J.\lambda E.K(J,E) \wedge (Nom(E)=x)))$
WO WO'	$S[wo]/S[ev]$	$\lambda S.\lambda K.\exists x.(Surf(x) \wedge S(\lambda J.\lambda E.K(J,E) \wedge (Acc(E)=x)))$
NI	$S[ni]/S[ev]$	$\lambda S.\lambda K.\exists x.(Surf(x) \wedge S(\lambda J.\lambda E.K(J,E) \wedge (Dat(E)=x)))$
EN	$NP[en]$	$\lambda N.\lambda F.\lambda x.(N(Surf,x) \wedge F(x))$
Q_EN	$NP[q\_en]$	$\lambda N.\lambda F.\lambda x.(N(Surf,x) \wedge F(x))$
T_EN	$NP[t\_en]$	$\lambda N.\lambda F.\lambda x.(N(Surf,x) \wedge F(x))$
E_EN	$NP[e\_en]$	$\lambda N.\lambda F.\lambda x.(N(Surf,x) \wedge F(x))$
M_EN M_EN'	$NP[m\_en]/NP[en]$	$\lambda M.\lambda N.\lambda F.\lambda y.(N(Surf,y) \wedge M(N,\lambda x.Rel(x,y) \wedge F(x)))$
PA PA'	$NP[pa]/NP[en]$	$\lambda M.\lambda N.\lambda F.\lambda y.(N(Surf,y) \wedge M(N,\lambda x.PartOf(x,y) \wedge F(x)))$
NEG1	$(NP[neg]/NP[en])$ $/(NP[m\_en]/NP[en])$	$\lambda L.\lambda M.\lambda N.\lambda F.L(M,\lambda E.\lambda y.\neg N(E,y),F)$
NEG2	$S[neg]/S[ev]$	$\lambda S.\lambda K.\neg S(K)$

2.1 複合語意味関係データセット

症例テキスト内の複合語に対して意味現象タグのアノテーションを行い，複合語意味現象タグ分類モデルの学習に必要な複合語意味関係データセットを構築した．

症例テキストには，J-MedStd CR: 症例報告 (Case Reports) コーパス[2]の頻度バランスサブセット 224 件を用い，テキストに含まれる複合語 3443 件に対して人手で意味現象タグを付与した．表 3 に 14 種類の意味現象タグの出現件数を示す．アノテーションには，アノテーションツールである brat[3]を利用した．

3 評価実験

3.1 症例テキストの推論データセットの構築

提案システムの評価実験にあたり，J-MedStd CR: 症例報告コーパス 91 件から症例テキストを抽出し，1054 件の推論データセットを人手で作成した．表 4 に推論

[1]https://mocobeta.github.io/janome/

[2]https://sociocom.naist.jp/j-medstd/cr/

[3]http://brat.nlplab.org/

表 3: 意味現象タグの件数

意味現象	タグ	件数
複合語	CW	3443
後ろの形態素の一部	PP	2309
イベント	EV	1434
イベント (サ変動詞をヲ格とするもの)	S_EV	78
ガ格名詞句	GA	271
ヲ格名詞句	WO	484
ニ格名詞句	NI	127
エンティティ	EN	1865
エンティティ(量系)	Q_EN	43
エンティティ(傾向系)	T_EN	31
エンティティ(作用系)	E_EN	15
修飾語	M_EN	876
体の部位	PA	547
否定	NEG	41

表 4: 推論データセットの例

前提文	仮説文	正解ラベル
肝機能改善傾向がみられる.	肝機能が改善する傾向がみられる.	entailment
pembrolizumab を投与開始した.	pembrolizumab を投与した.	entailment
急性虫垂炎のため当科紹介となった.	虫垂炎のため当科紹介となった.	entailment
VCM の臨床効果は作用時間に依存する.	VCM の臨床効果は作用する時間に依存する.	entailment
肝機能改善傾向がみられる.	肝機能が改善した.	non-entailment
pembrolizumab を投与開始した.	pembrolizumab を投与終了した.	non-entailment
急性虫垂炎のため当科紹介となった.	急性虫垂炎のため当科から紹介となった.	non-entailment
VCM の臨床効果は作用時間に依存する.	VCM の臨床効果は作用する時期に依存する.	non-entailment

データセットの例を示す．推論データセットに含まれる複合語は 709 件である．症例報告コーパスから抽出した複合語を含む症例テキストを前提文とし，前提文をもとに仮説文を作成した．データセットには，前提文が仮説文を含意している (entailment) ペアが 614 件，前提文が仮説文を含意していない (non-entailment) ペアが 440 件含まれている．

3.2 実験設定

作成した推論データセットを用いて，深層学習の含意関係認識モデルと提案システムとで精度比較を行なった．深層学習のモデルには，標準的な日本語の RTE データセットである JSICK [8]（学習データ 5 千件），JSNLI [11]（学習データ約 53 万件）で entailment・non-entailment の二値分類[4]タスクとしてファインチューニングした日本語 BERT（日本語 BERT-JSICK，日本語 BERT-JSNLI）を用いた．提案システムは，正解の意味現象タグを用いて意味解析を行った場合（Gold タグ），BiLSTM モデル, BERT モデルを複合語意味現象タグ分類モデルに用いた場合の 3 条件で評価を行った．

3.3 実験結果と考察

表 5 に意味現象タグの予測結果と推論の評価結果を示す．意味現象タグの予測結果については，推論データセットに含まれる複合語 709 件に対するタグ予測の正答率を示している．全ての複合語に対し正解の意味現象タグが付与されていた場合，提案システムは 1054 件中 816 件のペアについて，含意関係を正しく判定し，深層学習のモデルの性能を超える性能を示した．

表 6 に，推論データセット内の複合語に付与された意味現象タグごとの推論全体の正答率を示す．1 件の複合語に対し複数の意味現象タグが付与されるため，各正答率には重複したデータセットも含まれる．このうち，NEG が付与された複合語を含むデータセット 66 件の正答数については，日本語 BERT-JSICK は 35 件，日本語 BERT-JSNLI は 38 件と少ないのに対し，提案システムでは 64 件の文ペアについて正しい含意関係を予測した．

[4]contradiction は non-entailment として扱った．

表 5: 意味現象タグの予測結果，および推論の評価結果 (正答率)

	タグ予測	推論全体	entailment	non-entailment
提案システム（Gold タグ）	-	77.4%	72.3%	84.5%
提案システム（BiLSTM）	91.3%	67.2%	64.3%	71.1%
提案システム（日本語 BERT）	88.6%	61.6%	58.6%	65.7%
日本語 BERT-JSICK	-	64.0%	95.0%	20.9%
日本語 BERT-JSNLI	-	69.4%	92.8%	36.6%

表 6: 複合語に付与された意味現象タグごとの推論全体の正答率

		提案システム			日本語 BERT	
意味現象タグ	件数	Gold タグ	BiLSTM	BERT	JSICK	JSNLI
EV	556	73.0%	62.6%	54.0%	67.5%	68.9%
S_EV	110	77.3%	70.9%	53.6%	63.6%	71.8%
GA	177	73.5%	63.8%	59.3%	66.1%	68.4%
WO	288	71.2%	59.4%	52.1%	71.2%	71.9%
NI	84	81.0%	61.9%	46.4%	59.5%	60.7%
EN	440	83.2%	72.5%	70.0%	60.2%	71.1%
Q_EN	18	83.3%	61.1%	55.6%	50.0%	55.6%
T_EN	40	70.0%	55.0%	67.5%	47.5%	37.5%
E_EN	17	76.5%	58.8%	64.7%	52.9%	58.8%
M_EN	436	84.4%	70.9%	67.2%	63.5%	73.9%
PA	168	78.0%	58.3%	59.5%	61.3%	72.6%
NEG	66	97.0%	81.8%	86.4%	53.0%	57.6%

4 おわりに

本論文では，日本語の高度な意味解析・推論システム ccg2lambda に複合語解析モジュールを追加することで，複合語を含む症例テキストに対して頑健に意味解析と推論ができる推論システム Medc2l を構築した．さらに，日本語の症例テキストを用いて症例テキストの複合語に関する推論データセットを構築し，提案システムの評価を行った．実験の結果，提案システムは深層学習の含意関係認識モデルと同等またはそれ以上の性能を示した．とくに，深層学習のモデルは複合語を含む多くの問題に対して entailment と予測する傾向であったのに対して，提案システムは non-entailment のケースを正しく予測する傾向が見られた．

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] Aramaki, E., Yano, K. and Wakamiya, S.: MedEx/J: A One-Scan Simple and Fast NLP Tool for Japanese Clinical Texts, in *Proceedings of MEDINFO2017*, pp. 285–288 (2017).

[2] Bertot, Y., Castéran, P., Huet, G. and Paulin-Mohring, C.: *Interactive Theorem Proving and Program Development. Coq'Art: The Calculus of Inductive Constructions*, Springer (2004).

[3] Bethard, S., Savova, G., Palmer, M. and Pustejovsky, J.: SemEval-2017 Task 12: Clinical TempEval, in *Proceedings of SemEval-2017*, pp. 565–572, Vancouver, Canada (2017), Association for Computational Linguistics.

[4] Martínez-Gómez, P., Mineshima, K., Miyao, Y. and Bekki, D.: ccg2lambda: a compositional semantics system, in *Proceedings of ACL System Demonstrations*, pp. 85–90 (2016).

[5] Romanov, A. and Shivade, C.: Lessons from Natural Language Inference in the Clinical Domain, in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 1586–1596, Brussels, Belgium (2018), Association for Computational Linguistics.

[6] Usui, M., Aramaki, E., Iwao, T., Wakamiya, S., Sakamoto, T. and Mochizuki, M.: Extraction and Standardization of Patient Complaints from Electronic Medication Histories for Pharmacovigilance: Natural Language Processing Analysis in Japanese, *JMIR Med Inform*, Vol. 6, No. 3 (2018).

[7] Yadav, P., Steinbach, M., Kumar, V. and Simon, G.: Mining Electronic Health Records (EHRs): A Survey, *ACM Comput. Surv.*, Vol. 50, No. 6 (2018).

[8] Yanaka, H. and Mineshima, K.: Compositional Evaluation on Japanese Textual Entailment and Similarity, *Transactions of the Association for Computational Linguistics* (2022), to appear.

[9] Yoshikawa, M., Noji, H. and Matsumoto, Y.: A* CCG Parsing with a Supertag and Dependency Factored Model, in *Proceedings of ACL (Volume 1: Long Papers)*, pp. 277–287, Vancouver, Canada (2017).

[10] 荒牧英治, 若宮翔子, 矢野憲, 永井宥之, 岡久太郎, 伊藤薫：病名アノテーションが付与された医療テキスト・コーパスの構築, 自然言語処理, Vol. 25, No. 1, pp. 119–152 (2018).

[11] 吉越卓見, 河原大輔, 黒橋禎夫：機械翻訳を用いた自然言語推論データセットの多言語化, 第 244 回自然言語処理研究会, pp. 1–8 (2020).

[12] 矢田竣太郎, 田中リベカ, Cheng, F., 荒牧英治, 黒橋禎夫：汎用的な臨床医学テキストアノテーション仕様およびガイドラインの策定：重篤肺疾患ドメインに着目して, 自然言語処理, Vol. 29, No. 4, pp. 1165–1197 (2022).

[13] 東山翔平, 関和広, 上原邦昭：医療用語資源の語彙拡張と診療情報抽出への応用, 自然言語処理, Vol. 22, No. 2, pp. 77–105 (2015).