

ccg2lambdaを用いた医療言語処理における否定スコープ問題へのアプローチ

佐久間香那（指導教員：戸次大介）

1 はじめに

医療分野では電子カルテなどの電子化されたテキストが多く存在しており、英語に限らず日本語でもこうした電子テキストを用いた自然言語処理研究が発展しつつある。

石田ら [6] は、意味解析・推論システム ccg2lambda[1] を拡張し、日本語症例テキストの高度な意味解析と論理推論を扱える推論システムを提案した。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する意味解析と含意関係認識 (Recognizing Textual Entailment) を可能にした。含意関係認識とは、前提文と仮説文が与えられたとき、前提文が仮説文を含意するかどうかを判定するタスクである。Murakami et al.[2] は、知識補完 (knowledge injection) モジュールを加えることで石田らのシステムを拡張した。ここでいう知識とは、病名のパラフレーズ (言い換え) に関する知識であり、例えば、「DVT は深部静脈血栓症である」というように標準病名「深部静脈血栓症」を別表記「DVT」と関係づけるものである。Murakami et al. はこうした知識を自動で補完するモジュールを追加することで、前提文「DVT 発症を認めた」が仮説文「深部静脈血栓症を発症した」を含意するといった正しい判定を与えるシステムを構築した。

石田らおよび Murakami et al. による先行研究の中で残された課題として、否定のスコープに関する問題がある。本研究では、複合語を含む否定を wide scope で解析するための unary rule の追加、公理の導入を行う。

2 先行研究

石田らは、ccg2lambda に複合語解析モジュールを追加し、意味テンプレートを改良することで、複合語を含む症例テキストの意味解析と推論を可能にした Medc2l を提案した。

ccg2lambda とは、組合せ範疇文法 (Combinatory Categorical Grammar: CCG) [3, 5] に基づく高精度な構文解析と、定理証明支援系 Coq [4] 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。テキストの意味を高階論理式 (意味表示) によって表現し、論理式間の含意関係を自動判定する。

複合語解析モジュールは、症例テキストに含まれる複合語を抽出し、複合語を構成する形態素間の意味的な関係を表す意味現象タグを付与したのち、意味現象タグに基づく複合語の構文解析、意味解析というステップで複合語を解析する。意味解析の際に、統語範疇をキーとして意味表示を割り当てるために意味テンプレートを用いる。また、文の意味から考えると合成されるべき意味表示が、統語範疇の問題により合成できない場合がある。この問題に対処するため、石田らは CCG 構文木内の統語範疇が変化している個所で意味表示を変換する規則である unary rule を定義している。

Medc2l に残った課題のうちの一つが、否定のスコープに関する問題である。Medc2l では、複合語を含む否定を常に narrow scope で解析しているため、「患者は歩行未獲得であった」は「患者は歩行を獲得しなかった」を正しく含意する一方、「患者は歩行未獲得であった」は「患者は歩行しなかった」を含意しないと誤って判定してしまう。

3 提案手法

3.1 unary rule の提案

複合語の中の否定辞 (「未」「非」など) には、narrow scope での否定と wide scope での否定の 2 つの用法がある。「歩行未獲得」を例に説明すると、narrow scope では、「獲得」のみに否定がかかる。一方で、wide scope では、「歩行」と「獲得」の両方に否定がかかる。本研究では、各 scope での否定の意味表示を可能にするために、表 1 の unary rules を追加する。ただし、narrow scope での否定の意味表示を与えるための unary rule は先行研究 [6] で提案されたものである。提案した unary rule を使用した wide scope での否定による意味表示の合成過程を図 1 に示す。

「歩行未獲得」にそれぞれの unary rule を適用すると、以下の通りの意味表示となる。narrow scope での否定の意味表示では、「獲得する」というイベントのみが否定されている。一方、wide scope での否定の意味表示では、「歩行し、かつ獲得する」というイベントが否定されている。

(1) narrow scope での否定の意味表示

$$\exists x (\text{患者}(x) \wedge \exists y (\exists e (\text{歩行}(e) \wedge \text{する}(e) \wedge (\text{Nom}(e) = x) \wedge (\text{asEntity}(e) = y)) \wedge \neg \exists e (\text{獲得}(e) \wedge \text{する}(e) \wedge (\text{Nom}(e) = x) \wedge (\text{Acc}(e) = y))))$$

(2) wide scope での否定の意味表示

$$\exists x (\text{患者}(x) \wedge \neg \exists y (\exists e (\text{歩行}(e) \wedge \text{する}(e) \wedge (\text{Nom}(e) = x) \wedge (\text{asEntity}(e) = y)) \wedge \exists e (\text{獲得}(e) \wedge \text{する}(e) \wedge (\text{Nom}(e) = x) \wedge (\text{Acc}(e) = y))))$$

この wide scope の意味表示は「歩行イベントに対応していて、かつ、獲得イベントにおいて獲得されるような対象は存在しない」ことを表している。しかし、上記の意味表示だけでは wide scope の複合語を適切に分析できない。wide scope での否定の意味表示は、「歩行する」というイベントを A、「獲得する」というイベントを B とおくと、 $\neg(A \wedge B)$ という論理式とみなすことができる。wide scope での否定を含む複合語を適切に分析するためには、上記の論理式から $\neg A$ が導出されなければならないが、この論理式だけでは導出できない。したがって、この意味表示からは、「患者は歩行した」が意図通り帰結しない一方、「患者は歩行しなかった」も帰結しない。この問題を解決するために、次の公理を導入する。

	未 _{NEG}	獲得 _{S_EV}	
	$\frac{S_{neg,X2,ev}/S_{ev,X2,ev}}{\lambda S.\lambda K.S(K)}$	$\frac{S_{ev,X2,ev}}{\lambda K.\exists e.K(\lambda e_2.(獲得(e_2) \wedge する(e_2)), e)}$	>
歩行 _{EV}	$\frac{S_{neg,X2,ev}}{\lambda K.\exists e.K(\lambda e_2.(獲得(e_2) \wedge する(e_2)), e)}$		< un >
$\frac{S_{ev,X2,ev}}{\lambda K.\exists e.K(\lambda e_1.(歩行(e_1) \wedge する(e_1)), e)}$	$\frac{S_{wo,X2,ev} \setminus S_{X1,X2,ev}}{\lambda S.\lambda K.\neg \exists y.(S(\lambda J.\lambda e_1.(K(J, e_1) \wedge asEntity(e_1) = y)) \wedge \exists e.K(\lambda e_2.(獲得(e_2) \wedge する(e_2)), e) \wedge Acc(e) = y)}$		<
$\frac{S_{wo,X2,ev}}{\lambda K.\neg \exists y.(\exists e.(K(\lambda e_1.(歩行(e_1) \wedge する(e_1)), e) \wedge (asEntity(e) = y)) \wedge \exists e.K(\lambda e_2.(獲得(e_2) \wedge する(e_2)), e) \wedge Acc(e) = y)}$			

図 1: 提案した unary rule を使用した wide scope での否定の意味表示の合成過程

	適用前	適用後	unary rule
narrow scope	S[ev]	S[wo] \ S[ev]	$\lambda S1.\lambda S2.\lambda K.\exists y.(S2(\lambda J2.\lambda e2.(K(J2, e2) \wedge asEntity(e2) = y)) \wedge S1(\lambda J1.\lambda e1.(K(J1, e1) \wedge Acc(e1) = y)))$
wide scope	S[neg]	S[wo] \ S[ev]	$\lambda S1.\lambda S2.\lambda K.\neg \exists y.(S2(\lambda J2.\lambda e2.(K(J2, e2) \wedge asEntity(e2) = y)) \wedge S1(\lambda J1.\lambda e1.(K(J1, e1) \wedge Acc(e1) = y)))$

表 1: 否定を含む複合語を解析するための unary rule

3.2 公理の導入

- (3) $\forall x \forall y (\exists e (歩行(e) \wedge する(e) \wedge (Nom(e) = x) \wedge (asEntity(e) = y)) \rightarrow \exists e (獲得(e) \wedge する(e) \wedge (Nom(e) = x) \wedge (Acc(e) = y)))$

上記の公理は「歩行しているのならば、その歩行イベントは同じ主体によって獲得されていなければならない」ことを表すものであるが、これが wide scope の否定辞が関わる推論を媒介する。なお、このような公理は、「歩行」と「獲得」というイベント間に成立する関係についての常識の一部であり、現在の wide scope での意味表示のみでは、否定を含む複合語に対して適切な推論ができない問題を解決するためだけに追加する adhoc な知識とは異なる。自然演繹に基づく定理証明支援系である Coq に上記で提案した公理を追加した。Coq を用いることで、前提文の意味表示と仮説文の意味表示間の含意関係を判定し、判定結果が「含意」の場合は計算機上で判定の正しさを検証することが可能となる。

4 評価

Coq に下記の意味表示を入力し、前提文の意味表示が仮説文の意味表示を含意することを確認した。

- (4) 前提文の意味表示

$\exists x (患者(x) \wedge \neg \exists y (\exists e (歩行(e) \wedge する(e) \wedge (Nom(e) = x) \wedge (asEntity(e) = y)) \wedge \exists e (獲得(e) \wedge する(e) \wedge (Nom(e) = x) \wedge (Acc(e) = y))))$

- (5) 仮説文の意味表示

$\exists x (患者(x) \wedge \neg \exists y \exists e (歩行(e) \wedge する(e) \wedge (Nom(e) = x) \wedge (asEntity(e) = y)))$

これにより、追加した unary rule による wide scope の分析に基づいた「患者は歩行未獲得だった」が「患者は歩行しなかった」を含意することが正しく計算された。

5 おわりに

本稿では、Medc21 に残された課題である否定のスコープに関する問題を解決するために wide scope での分析を提案し、そのための unary rule と公理を Medc21

に追加した。今後は、Medc21 の環境構築を行い、考案した分析を元に評価実験を行う。また、評価のために症例報告コーパス MedTxt-CR¹ を用いて否定を含む複合語のデータセットを作成する。

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] Pascual Martínez-Gómez, Koji Mineshima, Yusuke Miyao, and Daisuke Bekki. ccg2lambda: A compositional semantics system. In Sameer Pradhan and Marianna Apidianaki, editors, *Proceedings of ACL-2016 System Demonstrations*, pp. 85–90, Berlin, Germany, August 2016. Association for Computational Linguistics.
- [2] Natsuki Murakami, Mana Ishida, Yuta Takahashi, Hitomi Yanaka, and Daisuke Bekki. Knowledge injection for disease names in logical inference between Japanese clinical texts. In Tristan Naumann, Asma Ben Abacha, Steven Bethard, Kirk Roberts, and Anna Rumshisky, editors, *Proceedings of the 5th Clinical Natural Language Processing Workshop*, pp. 108–117, Toronto, Canada, July 2023. Association for Computational Linguistics.
- [3] Mark Steedman. *The Syntactic Process*. MIT Press, 2000.
- [4] The Coq Development Team. *The Coq Proof Assistant: Reference Manual: Version 8.9.0*. INRIA, 2019.
- [5] 戸次大介. 日本語文法の形式理論. くろしお出版, 東京, 2010.
- [6] 石田真捺, 谷中瞳, 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.

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