

厚生労働科学研究費補助金
(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))
分担研究報告書

課題名 : 「医療テキストのための表現計算モデルの構築」に関する研究

研究分担者名 : 戸次 大介

国立大学法人お茶の水女子大学・基幹研究院・自然科学系・情報科学専攻・准教授

研究要旨

高階論理に基づく自然言語推論システム ccg2lambda による診療テキストの高度な意味解析を実現した。特に、診療テキストに頻出する複合語を解析し推論を行うモジュール、および外部知識と定理自動証明器を連動させるモジュールを開発することで、現在の標準的な言語モデルが不得手とするような意味理解の機構と、それに基づく検索システムの実現を目指す。

A. 研究目的

自然言語意味解析システム ccg2lambda に複合語解析モジュール Medc21 を追加することで、症例報告のテキストから正しい意味表示を導出することを目指した。

B. 研究方法

ccg2lambda の複合語解析モジュール Medc21 は、以下の要素から成り立つ。

- (1) 症例テキストに含まれる複合語を、意味合成への意味的貢献に応じて分類する「意味現象タグ」体系
- (2) (1) に基づいた症例テキスト複合語アノテーションデータセット
- (3) (2) を教師データとして学習した意味現象タグ付与のための系列ラベリング深層学習モデル
- (4) 意味現象タグに基づいた複合語内部構造のための CFG 構文解析器
- (5) 複合語 CFG 構文木から CCG 統語構造への変換アルゴリズム
- (6) 症例テキストの正しい意味表示を導出する意味合成過程

これらを接続することによって、症例テキスト中の複合語内部の意味合成過程が導かれ、それ以外の部分の CCG 統語解析結果と相まって医療テキストの意味表示が導出される。

(倫理面への配慮)

本研究は、医薬基盤・健康・栄養研究所において倫理審査、承認を得た後、人を対象とする医学系研究に関する倫理指針に従って実施している。

C. 研究結果

評価用に用意したデータセットに対する精度は、(2023 年 3 月の研究期間終了時には) 意味現象タグの平均予測精度が 91.3 % (BiLSTM) / 88.6 % (日本語 BERT) に達した。

D. 考察

(4)(5)(6) を踏まえた(1)のタグ体系設計は、理論言語学においても自然言語処理においても前例のない取り組みであったため、理論面・技術面の双方において難しさがあるが、その設計はアノテーションの一致度にも影響し、ひいては(3)における意味現象タグ予測モデルの精度にも影響する。

この成果を踏まえた(5)の統語変換および(6)の意味合成は、それによって導出された意味表示を用いた自動推論を含めて、本研究項目における R4 年度の課題となった。

E. 結論

この成果は、人工知能学会第 35 回全国大会 (査読なし)、および LENLS18 国際学会 (査読有り) に論文が採択され、それぞれ登壇発表を行ったが、コロナ禍のため、同年度にはそれ以上の成果発表・情報収集の機会を得ることができなかった。本研究内容は 2022 年 11 月までの期間延長を認めて頂き、当該の機会

は ESSLLI2022 において得ることができた。

F. 研究発表

1. 論文発表

- 1) Ishida, Mana; Yanaka, Hitomi; Bekki, Daisuke; "Compositional Semantics for Multiword Expression in Medical Case Retrieval", In Proceedings of the 18th International Workshop on Logic and Engineering of Natural Language Semantics (LENLS18), pp.231-239. (2021). [査読付き]

2. 学会発表

- 1) 石田真捺, 谷中瞳, 馬目華奈, 戸次大介, 「論理推論による症例検索に向けた日本語症例テキストの複合語解析の試案」, 第34回人工知能学会全国大会論文集, オンライン, 2021/6/8-11. (2021).
- 2) 石田真捺, 谷中瞳, 戸次大介, (2022). 日本語症例テキストの複合語解析と論理推論, 第36回人工知能学会全国大会論文集, 国立京都国際会館, 2022/6/14-17.

G. 知的財産権の出願・登録状況 (予定を含む。)

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし