

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))
分担研究年度終了報告書

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した。

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

本研究プロジェクトが提案する意味解析システムに残された課題は、複合語を含む診療テキストから合成的に導出された意味表示を用いて、実際に症例テキストと検索クエリ間の推論を行うことであった。

本年度の研究では、第一に、複合語推論モジュールを完成させた。これは、定理証明支援系 Coq を用いて、前提文である症例テキストが、仮説文である検索クエリを含意しているか判定を行うものである。並行して、複合語を含む症例テキストのための 212 件の (検索クエリ) 推論テストセットを構築して、Medc21 の評価を行った。

第二に、複合語推論モジュールにおいて外部知識を利用する研究を開始した。外部知識としては、荒牧グループが開発した「万病辞書」の活用を試みた。本研究の特徴は、予め大量の知識を投入する代わりに、統語解析および推論の結果 (証明できずに残存した論理式) を見てから on-the-y で万病辞書を参照し、追加する公理を自動生成する、という点にあり、定理自動証明器の負荷を大幅に低減する。

(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

構築した推論テストセットを用いて提案システムと日本語 BERT の比較評価を行った結果、日本語 BERT モデルは JSICK で事前学習したモデルが 64.0 %、JSNLI で事前学習したモデルが 69.4 % であるのに対し、提案システムは BiLSTM に基づくモデルが 69.7 % と上回った。

D. 考察

日本語 BERT は含意・非含意いずれの問題でも含意と予測する傾向があるのに対して、提案する推論システムは非含意のケースを正しく予測する傾向が見られた。

外部知識の利用においては、テキストに現れる文字列と、外部知識の辞書に登録されている専門用語のズレが課題となっていたが、この問題は Medc21 で設計・実装した複合語用の意味現象タグを利用することで解決しつつある。

全体としては、形式統語論・意味論、汎用言語モデル、複合語モジュール、外部知識、証明支援系 Coq を融合した診療テキスト検索の仕組みを提案することに成功し、当初目標としていた高度な診療テキスト検索が実現しつつある。

E. 広報

本研究の成果は、言語処理学会第 29 回年次大会（査読なし）および第 37 回人工知能学会全国大会（査読なし）に論文が採択済みであり、登壇発表を行った。現在、これらの成果をまとめた内容を論文誌（査読有り）に投稿中である。

また、2022 年 11 月には京都において「Kyoto Work-shop of Computational Linguistics」および東京において LENLS19 国際学会を主催した。招待講演として国際的に著名な計算言語学者を 2 名招聘し、同分野の最新の研究動向について情報収集をするとともに、本プロジェクトとその成果の周知に努めた。

F. 研究発表

1. 論文発表

- 1) 石田真捺, 谷中瞳, 戸次大介, (forthcoming). 日本語症例テキストの複合語解析・推論システム Medc21, (投稿中) .

2. 学会発表

- 1) 村上夏輝, 石田真捺, 高橋優太, 谷中瞳, 戸次大介, (2023). 病名知識の公理補完を用いた症例テキスト間の論理推論, 第 37 回人工知能学会全国大会論文集, 2L5-GS-3-05, 熊本城ホール, 2023/6/6-9.
- 2) 村上夏輝, 石田真捺, 谷中瞳, 戸次大介, (2023). 症例テキスト間の論理推論における病名知識補完の試み, 言語処理学会第 29 回年次大会, B6-3, 沖縄コンベンションセンター, 2023/3/13-17.

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