

令和 6 年 6 月 28 日現在

機関番号：33401

研究種目：基盤研究(C)（一般）

研究期間：2020～2023

課題番号：20K11833

研究課題名（和文）意味表現学習の性能向上と病名推定への応用に関する研究

研究課題名（英文）A Study on Enhancing Semantic Representation Learning and Its Application to Disease Name Estimation

研究代表者

芥子 育雄（Keshi, Ikuo）

福井工業大学・工学部・教授

研究者番号：40815867

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：本研究は、電子カルテの経過要約や主訴からの自動病名推定技術を開発し、医療現場への診断支援の可能性を探った。具体的には、1) 病名シソーラスを用いた意味表現学習による高精度かつ解釈性の高い病名推定、2) 大規模言語モデルBERTによる経過要約からの更なる高精度な病名推定、3) 主訴からの病名推定における意味表現学習による最適な主訴条件の特定と生成AIのGPT-4を用いたRAG（検索拡張生成）による高精度な病名推定を達成した。これにより、RAGの参照データ範囲の影響も確認した。

研究成果の学術的意義や社会的意義

本研究は、電子カルテの経過要約や主訴から病名を自動推定する技術を開発することで、医療分野における診断の透明性と効率性を向上させる点で意義を持つ。学術的には、病名シソーラスを活用した意味表現学習により、高精度な推定と解釈可能性を両立できることを示した。さらに、大規模言語モデルBERTやGPT-4を用いることで、精度が飛躍的に向上することを実証した。社会的意義としては、診療情報管理士や医師の負担軽減、医療費削減、医療サービスの質向上への貢献が期待される。本研究の成果は、医療現場の業務効率化を推進し、より質の高い医療提供体制の構築に貢献するものである。

研究成果の概要（英文）：This study developed a technology for automatically estimating disease names from electronic medical records (discharge summaries and chief complaints) for potential diagnostic support in clinical settings. Using a disease name thesaurus, semantic representation learning achieved highly accurate and interpretable estimations. Further accuracy improvements were attained with BERT for discharge summaries. For chief complaints, semantic representation learning identified optimal conditions, and the use of the generative AI GPT-4 with retrieval augmented generation (RAG) achieved high accuracy, demonstrating the influence of reference data scope.

研究分野：自然言語処理、データ工学

キーワード：電子カルテ 病名推定 意味表現学習 大規模言語モデル BERT GPT-4 病名シソーラス 機械学習

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

1. 研究開始当初の背景

医師の働き方改革はその特殊性から 2024 年以降に導入される予定であった(その後、2024 年 4 月から開始された)。厚生労働省は診療報酬制度で医師の負担軽減を図っているが、総合病院の勤務医は依然として激務である。地域医療ネットワークの活用が推奨されているが、登録患者数は少なく、診断結果の有効活用が進んでいない。

米国では CAC (Computer Assisted Coding) と呼ばれる病名の自動コーディング技術が進んでいるが、日本では診療情報管理士による手作業で行われている[1]。例えば、電子カルテから得られた学習済ニューラルネットワークモデルを利用し、患者の主訴から病名を推測し、地域医療ネットワークに登録されている医療機関の中から患者の病名に最適な医療機関を推薦する機能が考えられる。しかし、主訴のデータ量が少ないため、データのスパース性が問題となり、各総合病院単独での病名推定は限定的である。特にレアな病気の病名推定精度が高まれば、日本の診療情報管理士の生産性を大幅に向上させることができる。

研究代表者らは、ニューラルネットワークに専門家が構築した辞書を反映させる手法を提案している[2]。この手法により、機械学習の分類結果の解釈性が高まり、小規模データでも高精度な分類が可能となる。これにより、診療情報管理士だけでなく、医師の生産性も向上し、日本の医療現場全体の効率化が期待される。

2. 研究の目的

本研究の目的を以下に示す：

- 1) 新分野への辞書の対応方法の確立：ニューラルネットワークにより獲得される分散表現に専門家が構築した辞書を反映する手法を提案し、退院サマ리를元に、患者の経過要約を対象とした病名推定を行う。病名には ICD-10 (国際疾病分類) を用いる。
- 2) 病名推定の性能向上：文書の特徴量抽出に深層学習を用いることで、病名推定の性能を格段に向上させる。解釈性を維持しながら、メジャーな病名に対する病名推定の精度を改善する。
- 3) 病名推定のプラットフォーム化：主訴から病名を推定できるかを評価し、主訴から病名推定の有用性を検証するためのプラットフォームを構築する。

これにより、診療情報管理士および医師の生産性を向上させ、日本の医療現場全体の効率化を実現することを目指す。

3. 研究の方法

3-1. 新分野への辞書の対応方法の確立

本研究では、医療分野における診断の透明性と説明責任を向上させるために、意味表現学習を応用した。これは、結果を定量的に示すことができるため、医療分野に非常に適している。具体的には、病名シソーラスから自動的に構築された医療用単語意味ベクトル辞書を導入した。図 1 に提案手法の全体像を示す。

- 1) 共通の課題：電子カルテの機械学習を用いた病名推定モデルの開発において、一般的な課題は、ある病院のデータで訓練したモデルを他の病院に適用すると精度が低下することである。本研究の目的は、異なるデータ分布を持つ二種類の電子カルテ間で高い推定精度を達成するモデルを開発することである。

- 2) 経過要約データセット：富山大学附属病院から提供された退院サマ리를使用する。訓練データセットとして使用するのは、2004 年から 2014 年の旧電子カルテ (以降、旧 EMR と呼ぶ) で、94,083 件の症例と 3,204 種類の病名が含まれている。評価データセットとして使用するのは、2015 年から 2019 年の新電子カルテ (以降、新 EMR と呼ぶ) で、61,772 件の症例と 2,849 種類の病名が含まれている。データクレンジングの条件として、以下の症例や各症例のフィールドを除外する：
 - 欠損値がある症例
 - 説明変数として使用しないフィールド
 - 全記録の 0.02%未満のレアな病名の症例
 - 経過要約が 50 文字未満の症例
- 3) データ準備：退院サマリから経過要約を抽出し、医療

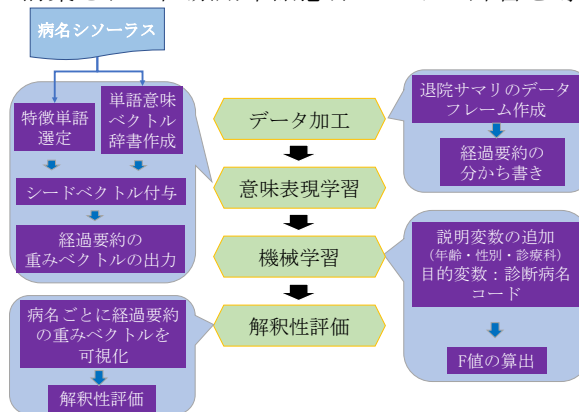


図 1. 提案手法の全体像

S111111111111112 単純ヘルペス脳幹脳炎

S11	11	11	11	11	11	11	12
区分1	区分2	区分3	区分4	区分5	区分6	区分7	

S11 神経系障害
 S1111 脳障害
 S111111 脳及び周辺組織の炎症性疾患
 S11111111 脳炎
 S1111111111 ウイルス性脳炎
 S111111111111 単純ヘルペス脳炎
 S11111111111112 単純ヘルペス脳幹脳炎

図 2. T 辞書の病名の階層例



図 3. 医療用単語意味ベクトル辞書の構成と意味表現学習のプロセス

辞書を用いて形態素解析を行う。医療辞書を用いることで、通常は分解される専門用語等を原形のままとどめることができる。

- 4) 意味表現学習: 芥子等[2]が確立した意味表現学習手法を用いてシードベクトルを生成する。T辞書[3]から選定した特徴単語を使用し、医療用単語意味ベクトル辞書を作成する。T辞書とは、病名の同義語と7階層からなる上位下位概念を記述した病名シソーラスであり、その例を図2に示す。ここから多くの病名と関連がある7階層までで5文字以下の264種類、6階層までで5文字以下の458種類の上位概念の病名および汎用的な概念から与えられる264種類の単語[1]を設定した。医療用単語意味ベクトル辞書にはT辞書より36768種類の基本病名を抽出し設定した。この医療用単語意味ベクトル辞書は、基本病名と458種類および264種類の病名特徴単語との関係を記述したものである。ICD-10の標準形に正規化することで、医療用単語意味ベクトル辞書の基本病名が31033種類に減少した。
- 5) 重み学習: 退院サマリから抽出された全単語を入力層とし、病名特徴単語を中間層として周辺語を予測できるように重み学習を行う。図3のように、辞書内の各病名特徴単語をニューラルネットワークの隠れノードに割り当てる。そして、ニューラルネットワークの重みは、辞書の再帰的拡張によって初期化される。
- 6) 評価: 新旧EMRのデータを用いて意味表現学習を行い、新EMRの上位20病名に対して旧EMRでモデルを作成し、新EMRの病名推定を評価する。線形SVMと非線形SVMを使用し、SMOTEでデータの補完を行った場合も検証する。評価結果のマクロ平均F値(以降、F値と呼ぶ)を比較し、最も高かったものの意味表現学習の解釈性を評価する。統計解析ソフトStatFlexを用いて各病名コードの最上位概念となる病名特徴単語の重みの分散や標準偏差、平均値を算出し、グラフ化して評価を行う。

3-2. 病名推定の性能向上

本研究では、病名推定の性能向上を目指し、意味表現学習と大規模言語モデル(BERT[4])を用いたアプローチを比較検討する。以下に提案手法の全体像を示す。

- 1) データ準備: 富山大学附属病院から提供された退院サマリを使用する。退院サマリから経過要約を抽出し、医療辞書を用いて形態素解析を行う。
- 2) 意味表現学習: 医療辞書を反映したニューラルネットワークを構築し、264種類の病名特徴単語と35種類の主な症状を追加した合計299次元の重みベクトルを出力する。
- 3) モデル構築: 旧EMRのデータを用いて、事前学習された様々なBERTモデルを微調整し、20病名のモデルを構築する。新EMRのデータで評価を行う。
- 4) 精度比較: 意味表現学習の精度と微調整した大規模言語モデル(BERT)の精度を比較する。意味表現学習では299次元のベクトルに年齢や性別の情報を追加し、線形SVMで分類を行う。評価指標はF値と意味ベクトルの解釈性とする。意味ベクトルの解釈性では、重みが最も高い病名特徴単語が推定対象の病名の上位概念であるかを確認し、正答率を算出する。大規模言語モデル(BERT)の微調整では、上位20病名に対する教師あり学習を実施し、F値を比較する。

3-3. 病名推定のプラットフォーム化

本研究では、主訴を基にした病名推定のためのプラットフォーム化を目指し、高速な意味表現学習を用いて病名推定に適した主訴条件を求めた後、評価に時間のかかるBERTやコストのかかるGPT-4の評価を行う。以下に提案手法の全体像を示す。

- 1) 主訴データセット: 旧EMRと新EMRから主訴が記載されている症例を抽出するためのプログラムを作成する。主訴データセットの症例には、診断病名コード(ICD-10)、主訴、診療科名、性別、年齢が含まれる。
- 2) 意味表現学習の適用: 主訴データを用いて意味表現学習を行い、最も病名推定の精度が高くなる主訴条件を選定する。
- 3) モデル開発: 主訴データを用いて、主訴から病名を推定するモデルを開発する。GPT-4を使用し、主訴データセットを訓練および評価に使用する。尚、本実験では、強力な生成能力とMicrosoft Azureによる堅牢なセキュリティを備えたAzure OpenAIのGPT-4(モデルバージョン: 1106-Preview)を使用した。

- 4) 評価：選定した主訴条件を用いて、意味表現学習とGPT-4（ゼロショット学習、フューショット学習、RAG（検索拡張生成））の精度を比較評価する。200件の主訴データセットを使用して病名推定の精度を検証する。
- これにより、診療情報管理士および医師の生産性を向上させ、日本の医療現場全体の効率化を実現することを目指す。

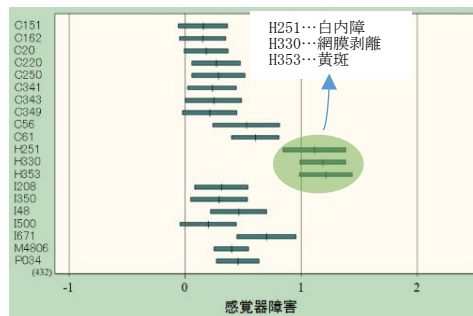


図 4. 病名特徴単語「感覚器障害」に関連する ICD-10 コード別の重み分布の可視化

4. 研究成果

4-1. 新分野への辞書の対応方法の確立

本研究では、病名シソーラスをシードベクトルとして用いた意味表現学習を導入し、以下の成果を得た。

- 1) 意味表現学習の特徴単語の選定と機械学習アルゴリズムの評価：百科事典から選出した従来の汎用的な特徴単語[1]よりも病名シソーラスから選出した病名特徴単語とした方がF値が10ポイント高かった。264種類の病名特徴単語を使用した意味表現学習を行った場合、機械学習アルゴリズムはSMOTEを使用した線形SVMが最も高い精度を示し、病名推定モデルのF値は72.4であった[5]。これにより、病名シソーラスを用いた方法が一般的な単語意味ベクトル辞書を用いるよりも高い精度を示すことが確認された。
- 2) 解釈性評価：意味表現学習により、電子カルテの経過要約から病名の上位概念を提供できることが示された。評価では、病名コードに対して重みが高い特徴単語の正答率は70%であった[5]。不正解の原因は、病名特徴単語選定時に5文字以下の病名に制限したため、推定病名の上位概念の単語が含まれていなかったことに起因する。例として図4に感覚器障害における解析結果を示すが、重みが高くなっている3病名とも感覚器障害で合っているため上手く分類できていることが分かる。
- 3) 比較実験：doc2vecを用いた比較実験では、分散表現学習が特定のコーパスに適応できることが示されたが、異なるデータ分布を持つ学習モデルでは病名推定の精度が大幅に低下することが確認された[5]。提案手法はこの問題を解決できたが、実用化にはさらにF値の向上が必要である。

本研究では、病名シソーラスを用いた意味表現学習が病名推定の精度と解釈性の向上に有効であることが示された。

4-2. 病名推定の性能向上

本研究では、意味表現学習を用いて病名推定の性能向上を目指した。医療辞書を反映したニューラルネットワークを構築し、264種類の病名特徴単語に加えて35種類の主な症状を追加し、合計299次元の重みベクトルを出力するモデルを開発した。この方法の結果を以下に示す。

- 1) モデルの比較：表1に示す通り、意味表現学習を用いた線形SVMモデルのF値は72.1であった。一方、UTH-BERTを使用したモデルはF値85.3を記録し、意味表現学習+線形SVMのモデルを大きく上回る精度を示した。UTH-BERTは1億2000万件の臨床テキストで事前学習されたモデルであり、日本語Wikipediaで事前学習されたTU-BERT(80.0)や、医療系テキストも含むMedBERTjp(80.4)よりも高い精度を示した。
- 2) 解釈性の向上：意味表現学習の解釈性評価では、表2に示す通り、推定病名に対して重みが高い病名特徴単語の正答率が70%であった。これに対して、病名シソーラスの最上位の概念について5文字以上の病名を全て含めるように改善した結果、解釈性が90%に向上した。これは、最上位概念の単語が全て病名特徴単語に含まれるようになったためである。
- 3) 病名推定モデルの精度向上：UTH-BERTを用いたモデルが最も高いF値を記録したが、意味表現学習を用いたモデルも、特に解釈性の面で有用であることが示された。これは、医療現場での実用性を考えると重要な要素であり、モデルの予測過程を理解しやすいことが評価される。

本研究では、大規模言語モデルを用いることで、意味表現学習と比較して病名推定の精度を大幅に向上させることができた。特に、臨床テキストで学習させたUTH-BERTは、他の大規模言語モデルよりも優れた結果を示した。さらに、意味表現学習における解釈性も、改善策を取り入れることで大幅に向上した。

表 1. 各モデルでのマクロ平均F値

モデル	マクロ平均F値
LinearSVC(意味表現学習)	72.1
UTH-BERT*1(東京大学大学院医学研究科BERT)	85.3

TU-BERT*2(東北大学 BERT)	80.0
MedBERT jp*3(大阪大学大学院医学系研究科 BERT)	80.4

*1 <https://ai-health.m.u-tokyo.ac.jp/home/research/uth-bert>

*2 <https://github.com/cl-tohoku/bert-japanese>

*3 <https://github.com/ou-medinfo/medbert.jp>

表 2. 病名コードに対しての重み上位の病名特徴単語の正答率（赤文字は正解病名を示す）

病名コード	重み 1 位	重み 2 位
C151	消化管障害	食道障害
C162	消化管障害	呼吸器系疾患
C20	消化管障害	腫瘍
:	:	:
正解率	90%	60%

4-3. 病名推定のプラットフォーム化

本研究では、主訴を基にした病名推定のためのプラットフォーム化を目指し、以下の成果を得た。

1) 主訴の条件設定：

- 意味表現学習と機械学習を組み合わせた評価では、主訴が 10 文字以上で診断病名コード（ICD-10）が中分類の場合に最も高い精度が得られた。
- ロジスティック回帰を用いた場合、全体の精度は 62.0%、F 値は 51.7 であった。
- 線形 SVM も同様の結果を示し、主訴が 10 文字以上で診断病名コード（ICD-10）の中分類の条件が最適であることが確認された。

2) BERT の評価：主訴を基にした評価では、UTH-BERT が BERT モデルの中で最も高い精度と F 値を示したが、精度では意味表現学習とロジスティック回帰の組み合わせが若干上回る結果となった。

3) GPT-4 の評価：

- GPT-4 を使用した評価では、Top-5 精度がゼロショット学習で 52.5%、フューショット学習で 61.0%、RAG1（新 EMR の上位 20 の診断病名コードに対応する主訴すなわち正解は除いた全主訴、参照文書 15 件）で 65.5%、RAG2（新 EMR の上位 20 の診断病名コードに対応する旧 EMR の主訴、参照文書 5 件）で 82.5%、RAG3（正解を含む全主訴、参照文書 15 件）で 84.5% であった。
- GPT-4 は広範囲の病名を推定できる特性を持ち、複数の選択肢を提供するため、希少または未知のケースに対応するための柔軟性を持つ。

本研究では、主訴を基にした病名推定のためのプラットフォーム化を実現するために、意味表現学習と機械学習、BERT、GPT-4 を比較し、それぞれの精度を評価した。意味表現学習と機械学習の組み合わせは、特定の条件下で BERT を上回る精度を示し、高い解釈性も確認された。GPT-4 を用いた評価では、RAG が最も高い Top-5 精度を示し、参照データの範囲によって精度が変動することが確認された。今後は、ICD-10 の全中分類を含むようにベンチマークを拡大して客観評価を行うと共に、経験の豊富な医師による主観評価を行い、実用的な診断支援ツールとしての実現を目指す。

参考文献

- [1] 辻岡和孝, 芥子育雄, 中川肇, 林篤志, 自然言語処理を利用した本邦版 Computer Assisted Coding 構築手法の研究. 診療情報管理, Vol. 34, No1, pp. 56-64, April 2022.
- [2] I. Keshi, Y. Suzuki, K. Yoshino, and S. Nakamura, Semantically Readable Distributed Representation Learning and Its Expandability Using a Word Semantic Vector Dictionary, IEICE Trans. On Information and Systems, Vol. E101-D, No. 4, pp. 1066-1078, April 2018.
- [3] 田代朋子, 医薬系シソーラス辞書「t 辞書」の構造と特徴, オンライン検索, Vol. 25, No3&4, pp. 197, 2004.
- [4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, arXiv:1810.04805.
- [5] I. Keshi, R. Daimon, and A. Hayashi, Interpretable Disease Name Estimation based on Learned Models using Semantic Representation Learning of Medical Terms, IC3K2022 Volume1: KDIR, pages265-272.

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 1件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 Ikuo Keshi, Ryota Daimon, Atsushi Hayashi	4. 巻 1
2. 論文標題 Interpretable Disease Name Estimation based on Learned Models using Semantic Representation Learning of Medical Terms.	5. 発行年 2022年
3. 雑誌名 Proceedings of the 14th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management	6. 最初と最後の頁 265-272
掲載論文のDOI（デジタルオブジェクト識別子） 10.5220/0011548900003335	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 辻岡和孝, 芥子育雄, 中川肇, 林篤志	4. 巻 Vol.34. No.1.
2. 論文標題 自然言語処理を利用した本邦版Computer Assisted Coding構築手法の研究	5. 発行年 2022年
3. 雑誌名 日本診療情報管理学会誌「診療情報管理」	6. 最初と最後の頁 56-64
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計4件（うち招待講演 0件 / うち国際学会 1件）

1. 発表者名 大門諒大, 芥子育雄, 林篤志
2. 発表標題 意味表現学習による解釈可能な病名推定と大規模言語モデルによる高精度な病名推定
3. 学会等名 2023年度電気・情報関係学会北陸支部連合大会
4. 発表年 2023年

1. 発表者名 Ikuo Keshi, Ryota Daimon, Atsushi Hayashi
2. 発表標題 Interpretable Disease Name Estimation based on Learned Models using Semantic Representation Learning of Medical Terms.
3. 学会等名 KDIR 2022 : 14th International Conference on Knowledge Discovery and Information Retrieval (国際学会)
4. 発表年 2022年

1. 発表者名 大門 諒大, 芥子 育雄, 林 篤志
2. 発表標題 医療辞書を用いた意味表現学習による解釈性のある病名推定
3. 学会等名 2022年度電気・情報関係学会北陸支部連合大会
4. 発表年 2022年

1. 発表者名 辻岡和孝, 芥子育雄, 中川肇, 林篤志
2. 発表標題 意味表現学習とサポートベクタマシンを用いたCACシステムの評価
3. 学会等名 第17回医療情報学会中部支部会学術大会
4. 発表年 2022年

〔図書〕 計0件

〔出願〕 計1件

産業財産権の名称 医療用単語意味表現学習方法を用いた病名学習済みモデルとこれを用いた解釈性のある病名推定システム及びその推定方法	発明者 芥子育雄	権利者 金井学園
産業財産権の種類、番号 特許、特願2022-057301	出願年 2022年	国内・外国の別 国内

〔取得〕 計0件

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	中川 肇 (Nakagawa Hajime) (30135256)	富山大学・学術研究部医学系・教授 (13201)	
研究分担者	林 篤志 (Hayashi Atsushi) (20283773)	富山大学・学術研究部医学系・教授 (13201)	

6. 研究組織（つづき）

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分 担 者	辻岡 和孝 (Tsujioka Kazutaka) (50724960)	金城大学・社会福祉学部・講師 (33306)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関