

診療録情報と疾病に特有な症状の紐づけによる非専門医の治療支援

Assisting Nonspecialists in Treating Patients through Linking
Medical Record Information to Disease Specific Symptoms

見附拓馬† 島川博光‡ 原田史子‡

Takuma Mitsuke Hiromitsu Shimakawa Fumiko Harada

1. はじめに

医療における患者の診断に関する記録は、一般的にカルテが用いられる。カルテを用いた診療記録は従来の紙の媒体から電子カルテを利用したデジタル方式に移行している。電子カルテは、従来の方式と比較して情報の活用や共有が容易となるため医療従事者は多くの恩恵を受ける。

しかし、カルテをデジタル化しただけでは、従来のカルテと同様に最適な情報選択が難しいという課題を解消できない。救急外来のような診療では、専門医が患者を診療するまでに、非専門医である当直の医師が事前の診療を実施する。患者が急性的な病態の場合、非専門医はカルテの情報と少ない所見から、迅速かつ適切な処置の意思決定を迫られる。また、病歴の多い患者は、診療録に記されている情報の量が膨大となるため非専門医の情報選択をさらに困難にする。これに加えて、診療録の主観的情報は曖昧性が強いことがあり、客観的情報と統括して診療に活用することが難しい[1]。このような診療録の課題が診断漏れを引き起こす可能性もあるため、診療録の情報を医学的な根拠に基づいて整理し非専門医の意思決定を支援するシステムが求められる。

電子カルテを用いた診断支援に関する研究が進んでいる。先行研究[2][3]は、電子カルテから構築した大規模言語モデルで診断予測を検証し、医学のドメイン知識をテキスト埋め込みで表現した。先行研究[4]は電子カルテに含まれる臨床テキストの意味的類似性をTransformerにより予測している。

しかし、これらの研究は、電子カルテの情報構造を考慮せずにテキストを扱っている。電子カルテの診療録に含まれる検査結果や医師の所見等の客観的情報と、患者の訴える主観的情報を効果的に組み合わせることで、より強力な大規模言語モデルを構築することができると考える。加えてこれらの研究は診断予測や医学知識の表現に成功しているが、実際に開発したモデルを利用する臨床医の意思決定を補助するような診断支援については言及していない。

本研究では、非専門医の意思決定を支援するために電子カルテの診療録情報を整理する。提案手法は診療録に含まれる主観的情報と客観的情報を各疾病特有の症状を示した医療辞典のテキストと紐づける。さらに紐付いている情報から、疾病ごとに所属確率のスコアを算出することで、患者が罹患している可能性のある疾病候補を非専門医に提示する。

2. 診療録の解析

2.1. SOAP

現在の診療録は医療従事者間で共有するため、一般的にSOAPという構造化された記録形式で記録される。SOAPは主観的情報、客観的情報、評価、計画から構成されており、約50年前にラリーウィードによって理論化された[5]。

主観的情報は主に主訴、病歴、経歴から構成され、患者または患者に近い人からの報告内容である。客観的情報は主に検査結果、他の臨床医の評価から構成され、医師が患者との診断から得た情報である。評価は、これらの主観的情報と客観的情報を統合した問題の分析結果に基づいた記録である。計画は治療計画や他の臨床医との相談の必要性に言及した記録である[6]。

本研究では、診療録における主観的情報と客観的情報と、疾病やそれに対する医療処置等の疾病特有症状をまとめた医療辞典の情報を紐づける。

2.2. テキスト埋め込み

本研究では診療録に記載されている主観的情報と客観的情報のテキストに対する埋め込み表現によってテキストをベクトル化する。これにより単語や文章などのデータの意味や関係性を数値的に表現し、解釈することができる。また、文書内のテキストを多次元空間に存在するベクトルで表すことで、似た意味を持つ文章が近くに存在することを確認できるため、これらをクラスター化して分析することができる。

埋め込みには単語をベクトル表現する方法と文章をベクトル表現する方法がある。どちらもベクトル表現したテキスト同士の意味的類似度を計算することが可能である[7]。本研究においては診療録の文章と医療辞典内の文章との類似度を算出するために最適な埋め込み法を選択する。

3. 診療録の紐付けと情報整理

本研究では、診療録に記録される主観的情報と客観的情報を含んだテキストと医療辞典内の疾病特有の症状を示すテキストとの多次元空間表現による意味的類似度から、各疾病に対して患者が罹患している可能性を示す所属確率を算出する。さらに患者が罹患している可能性のある疾病を、所属確率を基準にソートして提示することで、非専門医による最適な評価と計画を支援する。図1に提案手法の概要図を示す。

† 立命館大学 Ritsumeikan University

‡ 立命館大学 Ritsumeikan University

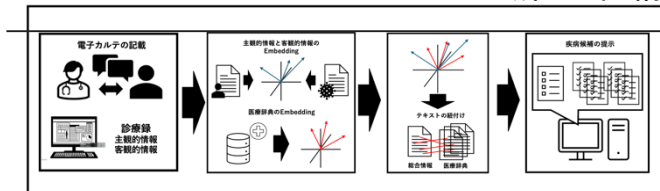


図1：手法概要図

3.1. 主観的情報と客観的情報

本研究では、診療録に記される主観的情報と客観的情報のテキストを利用する。診療録の基本的な順番として、主観的情報の後に客観的情報を記すことが一般的だが、現状のすべての診療録は必ずしもこの順番が守られているとは限らない。そのため診療録内の記載項目名から主観的情報と客観的情報のテキストを分類抽出する必要がある。分類された主観的情報と客観的情報を文章埋め込みによって多次元空間のベクトルに変換する。

主観的情報は患者自身の表現や自己報告であるため曖昧性が高いが、疾病を推定するために有益な情報が含まれている可能性がある。そのため本研究では主観的情報と客観的情報を適切な手法で組み合わせて特徴量テキストとして利用する。

3.2. 医療辞典

本研究では、主観的情報と客観的情報に含まれる情報から罹患している可能性のある疾病の候補を推定するために医療辞典を用いる。医療辞典は各疾病特有の症状、検査値、治療方法等を詳細にまとめているテキスト形式のデータセットである。本研究では、医療辞典を正確なルールベースとして利用する。

3.3. 診療録と医療辞典の紐付け

自然言語処理を用いて主観的情報と客観的情報に現れるテキストと類似度の高い医療辞典に含まれるテキストを紐づける。意味が似ているテキストは同じ疾病を表現するテキストである可能性が高いため、埋め込みにより高次元空間でテキストの意味をベクトルとして表現する。文書内のテキストを比較することで、似た意味を持つテキストは近くの方および位置座標に存在することを確認できる。またベクトル同士のコサイン類似度を求めることで、数値によるテキストの紐付けが可能となる。

本研究では3.1節で述べた主観的情報と客観的情報の適切な組み合わせを基にテキストを紐づける。客観的情報のみでは当てはまる疾病の候補が多く、いずれの候補も同等の所属確率を持つことが考えられる。そのため主観的もしくは客観的いずれかの情報を、類似度を計算する際の重みとして用いることで、より詳細に所属確率を算出することができる。と考える。

3.4. 診療録の整理

3.3節では主観的情報と客観的情報のテキストと病名辞書内のテキストを紐づけた結果が得られる。医療辞典内の各疾患特有の症状を示すテキストに紐づけられた、診療録内のテキストの個数や信頼度を基準として、患者が罹患している可能性の高い疾病の候補を推定する。医療においては疾病を検出する適合率よりも再現率が高い診断が求められるため、診断漏れが発生しないように、僅かでも可能性のある疾

病を候補として非専門医に提示することで診断を支援する。

また、候補となった疾病に対する必要な治療方法や検査方法を医療辞典から抽出し、疾病の候補とともに提示することで非専門医の意思決定を補助する。

4. おわりに

本研究では非専門医の診断を補助するために、診療録を用いた診断支援手法を提案した。提案手法は診療録の主観的情報と客観的情報のテキストと医療辞典の疾病特有の症状を示したテキストを埋め込みにより多次元空間上にベクトルとして表現する。これらのテキストベクトルに対して類似度の高いテキストのペアを紐づける。また紐づけられたテキストを、患者が罹患している可能性のある疾病の候補として所属確率を基準に整理する。

本研究で提案した手法は診断漏れを防ぎ、非専門医が迅速でかつ適切な治療計画を患者に施すことができると期待される。今後は、診療録のデータと最適な病名辞書を調査し、実験を通じて本手法の有効性を検証していく。

参考文献

- [1] Singhal, K., Azizi, S., Tu, T. et al. Large language models encode clinical knowledge. *Nature* 620, 172–180 (2023). <https://doi.org/10.1038/s41586-023-06291-2>
- [2] Charlotte A Nelson, Riley Bove, Atul J Butte, Sergio E Baranzini, Embedding electronic health records onto a knowledge network recognizes prodromal features of multiple sclerosis and predicts diagnosis, *Journal of the American Medical Informatics Association*, Volume 29, Issue 3, March 2022, Pages 424–434, <https://doi.org/10.1093/jamia/ocab270>.
- [3] Roberta A. Sinoara, Jose Camacho-Collados, Rafael G. Rossi, Roberto Navigli, Solange O. Rezende, Knowledge-enhanced document embeddings for text classification, *Knowledge-Based Systems*, Volume 163, 2019, Pages 955–971, ISSN 0950-7051, <https://doi.org/10.1016/j.knosys.2018.10.026>.
- [4] Ormerod M, Martínez Del Rincón J, Devereux B. Predicting Semantic Similarity Between Clinical Sentence Pairs Using Transformer Models: Evaluation and Representational Analysis. *JMIR Med Inform.* 2021 May 26;9(5):e23099. doi: 10.2196/23099. PMID: 34037527; PMCID: PMC8190645.
- [5] Podder V, Lew V, Ghassemzadeh S. SOAP Notes. 2023 Aug 28. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan–. PMID: 29489268.
- [6] Sando KR, Skoy E, Bradley C, Frenzel J, Kirwin J, Urteaga E. Assessment of SOAP note evaluation tools in colleges and schools of pharmacy. *Curr Pharm Teach Learn.* 2017 Jul;9(4):576-584.
- [7] Edward Choi, Mohammad Taha Bahadori, Elizabeth Searles, Catherine Coffey, Michael Thompson, James Bost, Javier Tejedor-Sojo, and Jimeng Sun. 2016. Multi-layer Representation Learning for Medical Concepts. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. Association for Computing Machinery, New York, NY, USA, 1495–1504. <https://doi.org/10.1145/2939672.2939823>.