１Ａ０２

生成 AI と自然言語処理の医療への適用：診療記録からの患者 QOL 測定

○新村和久，重松愛里（株式会社ユカリア），藤倉将平（株式会社サイシキ）
kazuhisa.shimmura@eucalia.jp

1. はじめに

近年、AI 技術の急速な進展により、生成 AI や自然言語処理技術が医療分野におけるイノベーションを牽引している。特に、Generative Pre-trained Transformer (GPT) などの先進的なモデルが、テキスト生成や自然言語処理のタスクにおいて顕著な成果を上げている[1]。これにより、従来困難であった診療記録からの患者の生活の質（QOL）の測定やテキスト情報の構造化、活用が可能となってきた。

この有用性としては、2019 年 4 月より、医薬品・医療機器を対象に費用対効果評価制度が導入されるなど、治療介入上の患者 QOL 測定は重要であるが[2]、全ての患者の経時的な QOL 測定は困難であり、診療記録から自然言語処理により精度よく QOL 測定が可能であれば、治療の効果検証に役立つことが期待される。一方で、この技術的進展は、医療分野におけるセンシティブな情報の取り扱いや個人情報保護法との関連性など、多くの社会的・倫理的課題をもたらしている。

本発表では、匿名化電子カルテ情報の医師所見、看護記録からの回顧的な QOL 測定方法の検証結果を共有し、その活用の具体的事例と問題点について議論につなげる。

2. 先行研究

①生成 AI の医療分野への適用

近年、AI 技術の進展に伴い、生成 AI や自然言語処理技術は、医療テキストの解析や情報抽出の分野で注目されている。AI 歴史は長いが近年の急速な転換の転換になったのは Transformer モデル[3]の誕生であり、このアーキテクチャは後の BERT（Bidirectional Encoder Representations from Transformers）、GPT などの有名なモデルの基盤となっている。Transformer は、従来の RNN（Recurrent Neural Network）や CNN（Convolutional Neural Network）を使用せずにアテンションメカニズムのみを使用してシーケンスデータを処理し、翻訳のようなシーケンス対シーケンスのタスクにおいて、従来の RNN ベースのモデルよりも優れた性能を示し、かつ長い依存関係を効果的に学習することができるようになった[3]。

ChatGPT で注目を集めた GPT は、自然言語処理の分野でのテキスト生成や解析において、高い性能を持つことが示されている[1]。特に、医療文献のテキスト生成やマイニングにおいて、GPT は有望な結果を示している[1]。また、最新の GPT-4 による性能比較では、UBE（米国統一司法試験）の上位 10%、GRE（米国大学院入学試験）でも好成績を残すことが報告されている[4]。国内での検証では、GPT-3.5 と GPT-4 は、日本の医師国家試験でのパフォーマンスを比較する研究が行われており、GPT-4 は GPT-3.5 よりも高い精度を示し、非英語圏での臨床的推論と医学的知識の信頼性を示している[5]。

②自然言語処理技術の医療情報解析

ヘルスケアなど一部の領域では、情報を正確に保存する必要があるため、完全に自動化されたアプローチは使用できない。その代わりに、オートコンプリートを用いた半自動化されたアプローチを使用することで、より迅速かつ高品質にテキストを簡略化することが検討されている。

具体的には、BERT, RoBERTa（Robustly optimized BERT approach），XLNet: Generalized Autoregressive Pretraining for Language Understanding （XLNet），および GPT-2 の 4 つのモデルを比較し、文を単純化するためのオートコンプリートをどのように組み込むことができるかを検証している[6]。自然言語処理を用いた患者 QOL に関する研究では、甲状腺がん患者を対象に QOL 指標の軌跡を予測したところ、ロジスティック回帰分析、サポートベクターマシンを用いた機械学習では約 60%の精度であったが、BERT モデルを用いた場合には AUC（area-under-curve）76.3%まで向上したこと、および GPT-2 を用いたデータの拡張も、分類性能を向上させたことが報告されている[7]。

③AI の医療分野での社会的・倫理的課題に関する研究

医療分野においても、GPT の技術が健康ケアの提供における可能性や問題点についての議論が行われている[8]。これらの技術の進展に伴い、AI の説明可能性（explainable AI(XAI)）や社会的・倫理的課題が浮上してきており、これらの課題に対する取り組みも進められている[9] [10]。

特に倫理面では日本の個人情報保護法上、予め利用目的についての同意が必要であり[11]、医療情報を学習用データに用いるという点について、取得段階で具体的に利用目的に定めておくべきか否かという論点もある[10]。なお、医療データ利活用の推進が現在進められているが、必ずしも本人の同意ではない形で、本人の権利利益を保護しながら、利用可能な制度・運用を整備する必要性を現在議論しているように[12]、現時点でデータ取得時の本人同意を得た上で二次利用可能な医療データへのアクセスのハードルは高い。

3. 目的

文脈を考慮した分散表現が得られる BERT の日本語事前学習モデル（cl-tohoku/bert-base-japanese-whole-word-masking）に診療記録の活用によるファインチューニングを行うことで、EQ-5D-5L[1]の判定を自動で行うアルゴリズムを作成し、医療技術の経済評価に資する回顧的 QOL 値の算出を試行する。

4. 方法

①プレ検討

株式会社ユカリアが所有する匿名化電子カルテデータベースより、パーキンソン病患者 34 名の医師所見、看護記録について KH Coder を用いて抽出後リストを作成し、EQ-5D-5L（日本語版）の質問紙を踏まえた医療従事者による目視判定で 66 個のキーワードを設定し、患者毎に 1 月平均のキーワード出現頻度を算出した。

②自然言語処理を用いた判定

プレ検討を踏まえ、29 個のキーワードから教師データの作成対象となる診療情報を抽出した。医療従事者の確認を含めた人手によるアノテーションにより 226 件の教師データ、及び ChatGPT（GPT-3.5）を活用したデータ拡張により 129 件の教師データを作成し、機械学習による EQ-5D-5L 表現の抽出、及びスコア化を実施した。EQ-5D-5L のスコア化は池田ら（2015）[13]の報告に準じた。結果は下記 A,B の 2 パターンで判定を実施した。

A. EQ-5D-5L の尺度を加えない（1～5 のいずれかの判定）の抽出精度

B. EQ-5D-5L の尺度を一部加えた場合（1 と 2～5）の抽出精度

図 1: EQ-5D-5L の判定

項目	設問	スコア
移動の程度	歩き回るのに問題はない	1
	歩き回るのに少し問題はある	2
	歩き回るのに中程度の問題がある	3
	歩き回るのにかなり問題がある	4
	歩き回ることができない	5
身の回りの管理	自分で体を洗ったり着替えをするのに問題はない	1
	自分で体を洗ったり着替えをするのに少し問題がある	2
	自分で体を洗ったり着替えをするのに中程度の問題がある	3
	自分で体を洗ったり着替えをするのにかなり問題がある	4
	自分で体を洗ったり着替えをすることができない	5
ふだんの活動	ふだんの活動を行うのに問題はない	1
	ふだんの活動を行うのに少し問題がある	2
	ふだんの活動を行うのに中程度の問題がある	3
	ふだんの活動を行うのにかなり問題がある	4
	ふだんの活動を行うことができない	5
痛み/不快感	痛みや不快感はない	1
	少し痛みや不快感がある	2
	中程度の痛みや不快感がある	3
	かなりの痛みや不快感がある	4
	極度の痛みや不快感がある	5
不安/ふさぎ込み	不安でもふさぎ込んでもいない	1
	少し不安あるいはふさぎ込んでいる	2
	中程度に不安あるいはふさぎ込んでいる	3
	かなり不安あるいはふさぎ込んでいる	4
	極度に不安あるいはふさぎ込んでいる	5

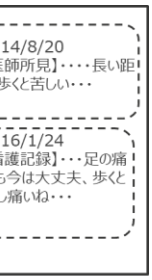
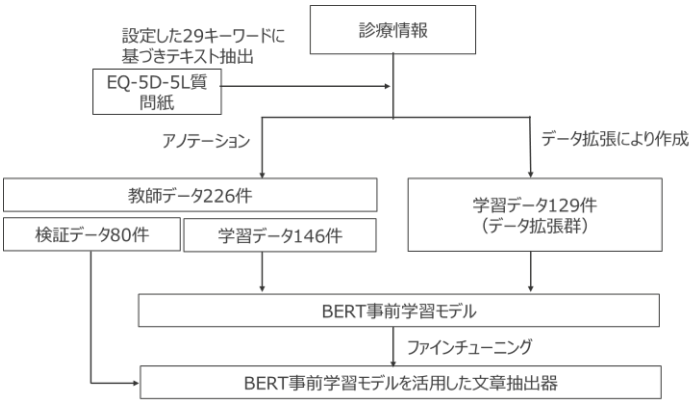


図 2:構築したモデル



5. 結果

A のスコアを考慮しない場合の f1-score は 0.344～0.707、B のスコアを一部考慮した場合の f1-score はスコア 1 の判定で 0.543～0.692（不安/ふさぎ込みを除く）、スコア 2～5 の判定で 0.325～0.650 であった。

[1] EQ-5D-5L(EuroQol 5-dimensions 5-levels)：5 つのドメイン(移動の程度、身の回りの管理、ふだんの生活、痛み/不快感、および不安/ふさぎ込み)に基づく患者報告による QOL の評価指標

表 1：A の抽出精度

	precision	recall	f1-score	形態素数
O	0.963	0.978	0.970	16807
移動の程度	0.881	0.590	0.707	629
身の回りの管理	0.400	0.703	0.510	148
ふだんの活動	0.583	0.438	0.500	281
痛み/不快感	0.629	0.695	0.660	453
不安/ふさぎ込み	0.732	0.225	0.344	231
accuracy	0.939	0.939	0.939	
weighted avg	0.939	0.939	0.939	18549

表 2：B の抽出精度

	Score	precision	recall	f1-score	形態素数
O		0.963	0.978	0.970	16807
移動の程度	1	0.993	0.531	0.692	275
	2-5	0.745	0.576	0.650	354
身の回りの管理	1	1.000	0.357	0.526	28
	2-5	0.376	0.783	0.508	120
ふだんの活動	1	0.577	0.513	0.543	154
	2-5	0.595	0.346	0.438	127
痛み/不快感	1	0.557	0.899	0.688	168
	2-5	0.643	0.519	0.575	285
不安/ふさぎ込み	1	0.000	0.000	0.000	8
	2-5	0.712	0.211	0.325	223
accuracy		0.936	0.936	0.936	
weighted avg		0.937	0.936	0.933	18549

6. 議論

技術面では、診療記録からの EQ-5D-5L 表現の抽出において、一定精度での抽出可能性が示唆された。特に表現の有無だけに絞れば、条件 A,B ともに O（EQ-5D-5L の判定に設定したキーワード以外の判定精度）の f1-score は 0.970 であり、何らかの QOL に関する表現が出現しているか否かは高精度で判定できる。ただし、各項目間に焦点を当てるとデータの不均衡性、及びスコア化の課題が残存している。この精度自体は、教師データの数を増やすこと等にて一定程度改善が期待されるが、重要な点はヘルスケア分野において 100%の正解率にはならない中で活用することが許容されるか否かという点にある。

一方で、AI の活用により一定精度の EQ-5D-5L スコアを過去の診療記録から回顧的に算出することは今までは実現できなかったため有用と考えられる。例えば、用途としては、同一人の EQ-5D-5L スコアの経時変化を観測しての特定薬剤介入の前後比較、心身症状の予兆の観測によるモニタリングシステムなど、連続的な評価や他の事項との総合評価に用いるなどが想定される。このように、自然言語処理等の AI の活用においては実用技術ごとの課題と有用性を踏まえながら許容される活用用途を議論する

ことが重要と考えられる。

社会・倫理面では、医療情報を活用するための患者同意取得の在り方が問題となる。現行法では、院内掲示等により公表 患者から明示的に留保の意思表示がなければ、患者の黙示による同意があるとみなせるとされているが、それ以外では原則二次利用の際に患者同意が必要となる（内閣府, 2023）。今回の解析に用いた株式会社ユカリアのデータは提携先病院の協力を得て患者の黙示の同意を得た電子カルテデータを匿名化、および構造化して二次利用可能な整備している。その他のわが国で利用可能なデータベースの種類では、次世代医療基盤法の認定匿名加工医療情報作成事業者（認定事業者）のデータベースやメディカル・データ・ ビジョン株式会社、リアルワールドデータ株式会社、株式会社日本医療データセンターなどがあるが、データベース毎に含まれているデータが異なり、研究目的に適したデータベースを見極め、選択する必要があるのが現状である。特に今回用いた電子カルテの医師所見・看護記録について薬剤、検査値情報等と連結した上で二次利用可能なデータベースは日本ではほとんど存在していないと考えられる。これらを踏まえると、AI を活用した医療技術を普及するためには、患者に不利益がなくかつデータを活用しやすい形とするための制度・運用を整備することが重要となる。

参考文献

[1] Luo, R., Sun, L., Xia, Y., Qin, T., Zhang, S., Poon, H., & Liu, T.-Y. (2022), “BioGPT: Generative Pre-trained Transformer for Biomedical Text Generation and Mining”, Bioinformatics, Vol. 23, 6. https://dx.doi.org/10.1093/bib/bbac409

[2] 保健医療経済評価研究センター（2022），中央社会保険医療協議会における費用対効果評価の分析ガイドライン（第 3 版）

[3] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., Polosukhin, I. (2017). “Attention is all you need” 31st Annual Conference on Neural Information Processing Systems (NIPS), vol. 30, pp.5998-6008.

[4] OpenAI, (2023), “GPT-4 Technical Report” arXiv preprint arXiv:2303.08774. https://arxiv.org/abs/2303.08774.

[5] Takagi, S., Watari, T., Erabi, A., Sakaguchi, K., (2023), “Performance of GPT-3.5 and GPT-4 on the Japanese Medical Licensing Examination: Comparison Study”, Vol. 9, e48002. https://dx.doi.org/10.2196/48002

[6] Van, H., Kauchak, D., Leroy, G., (2020), “AutoMeTS: The Autocomplete for Medical Text Simplification”, In Proceedings of the 28th International Conference on Computational Linguistics, pp.1424-1434. http://dx.doi.org/10.18653/v1/2020.coling-main.122

[7] Lian, R., Hsiao, V., Hwang, J., Ou, Y., Robbins, S.E., Connor, N.P., Macdonald, C.L., Sippel, R.S., Sethares, W.A., Schneider, D.F. (2023). “Predicting health-related quality of life change using natural language processing in thyroid cancer”, Intelligence-Based Medicine, Vol.7, 100097. pp.1-6. https://doi.org/10.1016/j.ibmed.2023.100097

[8] Korngiebel, D. M., & Mooney, S. D. (2021), “Considering the possibilities and pitfalls of Generative Pre-trained Transformer 3 (GPT-3) in healthcare delivery”, npj Digital Medicine, Vol. 4, 93. https://doi.org/10.1038/s41746-021-00464-x

[9] Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F. (2020). “Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI”, Information Fusion, Vol. 58, pp.82-115. https://doi.org/10.1016/j.inffus.2019.12.012

[10] 日本医師会生命倫理懇談会（2022），「医療 AI の加速度的な進展をふまえた生命倫理の問題」について

[11] 個人情報保護委員会「「個人情報の保護に関する法律についてのガイドライン」に関する Q&A」(平成 29 年 2 月 16 日(令和 5 年 5 月 25 日更新)

[12] 内閣府規制改革推進会議（2023）「医療データ利活用について（論点整理（骨子））」，第 12 回 医療・介護・感染症対策ワーキング・グループ

[13] 池田，白岩，五十嵐，能登，福田，齋藤，下妻（2015），日本語版 EQ-5D-5L におけるスコアリング法の開発、保健医療科学, Vol. 64, 1, pp.47-55.

[14] 内閣府 HP, 次世代医療基盤法制度の概要 https://www8.cao.go.jp/iryou/gaiyou/gaiyou.html（最終アクセス 2023/9/20）