

研究テーマ

救急初療現場の混乱を軽減する言語データ構造化 GPT モデルの開発

研究代表者

施設名 : 東京科学大学
整形外傷外科治療開発講座

氏名 : 山田 賢太郎

研究テーマ：救急初療現場の混乱を軽減する言語データ構造化GPTモデルの開発

所属先：東京科学大学（前 東京医科歯科大学）整形外科

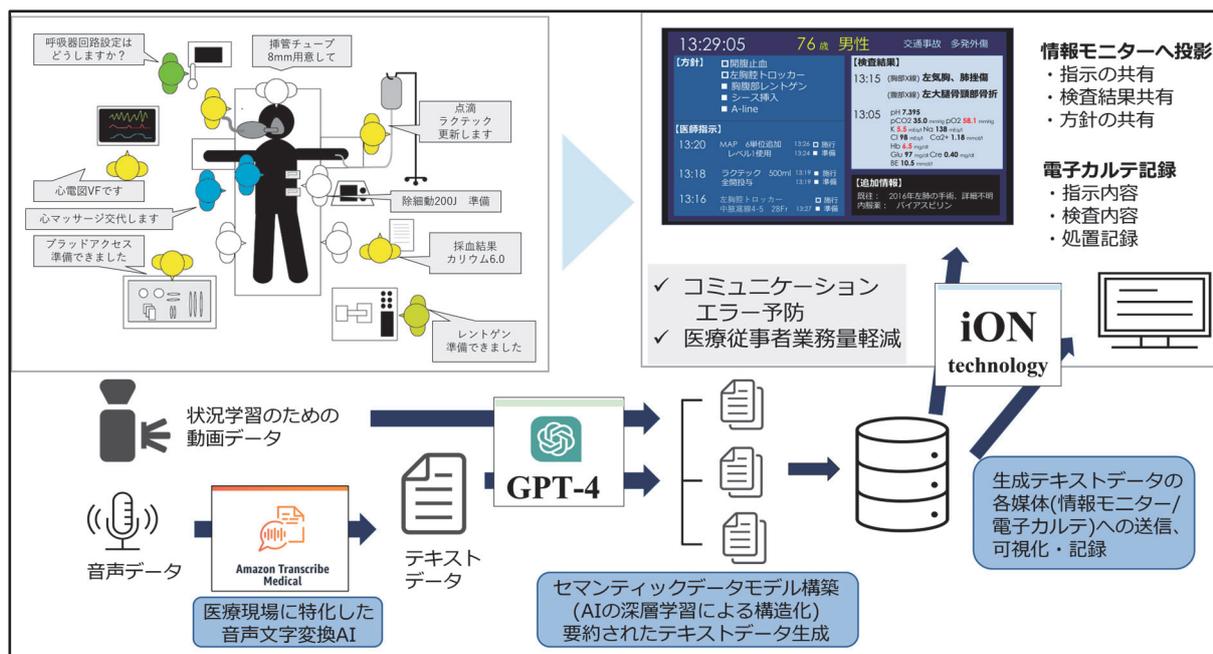
氏名：山田 賢太郎

【目的と方法】

医療現場において、医療従事者間のコミュニケーションエラーが誤投薬や誤処置などインシデントに繋がっている¹⁻³⁾。これを予防するためにTeam STEPPS⁴⁾に代表されるようなノンテクニカルスキルと呼ばれるコミュニケーションの向上や口頭指示を禁止するようなシステム作りが対応されている。しかしながら特に救急医療や集中治療、あるいは医師が手術など他の業務に集中している状況では医療従事者間で特にコミュニケーションエラーが起こりやすい状況は日常的に存在し、既存の対応ではインシデントを撲滅する事はできない。

特に重症外傷診療は、患者の病状が急速かつ流動的に変化するため、リアルタイムでカルテでのオーダーやカルテ記載も困難なことが多く、患者に行う処置などは口頭で行われる指示が多い。エラーを減少させるべく確認手順を増やすこと⁵⁾は医療従事者の業務を増加させることになり、時間との戦いでもある重症外傷診療にとっては利益より不利益の方が多い。

従って、医療従事者の業務を減少させつつ、コミュニケーションエラーを防止するシステムの開発、実用化は、特に救急初療/医療安全において喫緊の課題である。しかしながら、これを実現したシステムは過去に例がない。



本研究では文字情報から診療系統(指示/記録/検査結果)を自動で分類し、要約する事ができるGPTモデル(Generative Pretrained Transformer)を人工知能(AI)を用いて開発する事を目的とした。最終的に救急初療現場での医療従事者間の会話から、救急初療での状況をリアルタイムでモニターに表示させると同時に、電子カルテに記録可能な形式に出力させることのできるシステムの完成【上図】を目指す。

【システム開発方法】 システムの開発は以下の4段階で行う。

(1)第一段階：音声情報を文字情報への変換。

本技術には既存の医療音声文字変換技術を基に、実際の医療現場で収集した音声/動画データ(研究分担者 森下幸治)からAIを用いて精度の向上を目指す。音声/動画データの解析は東京科学大学 医療工学研究部門 情報医工学分野(研究分断者 中島義和)で行う。

(2)第二段階：文字情報の出力。

近年発展が著しいChat GPTに代表されるGPTモデル(大規模言語モデル：LLM)を用い、収集したデータの内容を把握し、過去のデータと対比を行いながら出力する。すなわち、単に音声情報を正確に文字化するのではなく、音声情報は要約され(セマンティックデータモデル)、かつエラーが疑われる際はその注意喚起が自動で促されるシステムを目指す。

(3)第三段階：タブレット/スマートフォンを用いた可視化/電子カルテへの送信。

研究分担者 中島義和が基本原理を確立した医療情報連携システムのiON技術(特開2017-224158)を用い、医療従事者が確認しやすいように大型モニターあるいはスマートフォンに情報が表示、確認可能なインターフェンスを確立する。確認後は簡便に電子カルテへ送送信できるよう無線通信システムを用いて電子カルテへ送信可能なシステムを確立する。各種カンファレンスや患者説明の電子カルテへの記載を簡潔にまとめる事が可能な音声文字変換技術を確立する。

(4) 第四段階：開発したシステムの精度検証。

システム導入前/後のそれぞれ500例において指示出しから実施までの時間およびインシデントの発生率を評価する。なお対象症例数については100例実施後にサンプルサイズに関する統計解析評価を行い、適宜追加する(研究分担者 工藤篤)。

【結果】

本研究計画は東京医科歯科大学医学部倫理審査委員会に2023年9月1日承認された(M2023-057)。その後研究分担者で月1回の研究ミーティングを軸に、研究を開始した。

(1)第一段階：音声情報を文字情報への変換

①ビデオカメラに録音された音声の文字変換

まず過去に教育/研究目的に東京医科歯科大学 救急初療部で撮影/保存された音声・動画を、Amazon transcribe medical®を用いて、文字情報に変換した。しかし、

ビデオカメラに付属するマイクでは、機械音を含むあらゆる音源を拾っていたため、既存の音声文字変換APIを単純に適用するだけでは、読解可能な文章に変換されず精度不良であった。周波数による音声信号処理による解決を図ったが、救急現場では不特定多数の人間が複数人で同時に会話が行われており、これらの音声を自動的に分離する事は現在の技術の開発には時間を要し現実的ではないと判断された。

そのため、音声文字変換をリアルタイムで間違いなく行うためには、音声収集のシステムを確立する必要があるとの結論に至った。

②音声文字変換を確実にを行うための収集システムの確立

i) 個別マイクシステムと3Dカメラを用いた記録システム

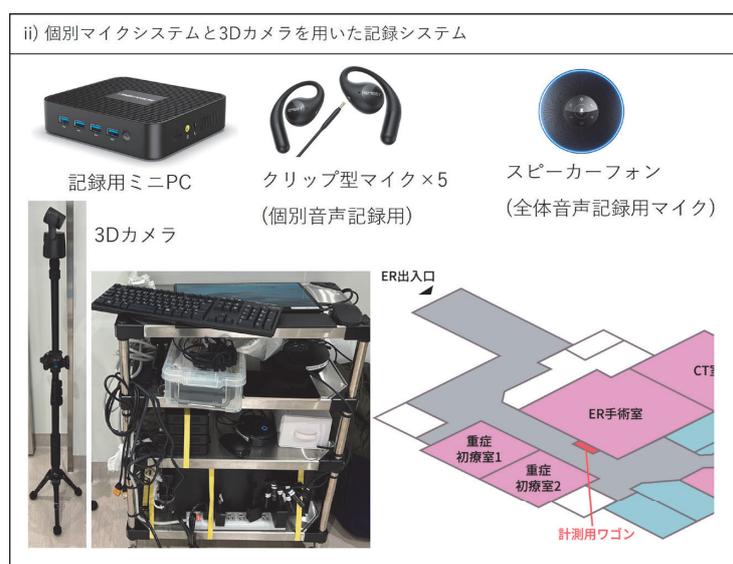


まず複数のクリップマイクを有するマイクシステムとGo Proによる動画記録を組み合わせたシステムを試作し、実際の救急初療を模して実験を行った(左図)。

しかしクリップマイクにより音声は鮮明に録音できるものの、音声は1つのファイルに統合されるため、複数人数の会話においてAmazon transcribe medical®による文字変換は十分な精度を実現する事ができなかった。

目的達成のためには一人の発言を個別の音声ファイルとして保存し、すべての音声ファイル/動画ファイルを同期するシステムの構築が必要であると結論付けた。

ii) 個別マイクシステムと3Dカメラを用いた記録システム



次に個別に音声記録が可能なように片耳に付けるクリップマイク5台を用意し、それぞれでミニPCを用いて録音を行った。また音声の同期用に全体音声記録用マイクを用意し、3Dカメラと共に同期して録音/録画できる記録システムを構築した(左図)。実際の救急初療での記録を試したが、複数の個別音声記録の同期が不十分であった。

現在、Githubを用いて同期記録のプログラムを構築中である。

(2)第二段階：文字情報の出力。

①音声文字情報のセマンティックデータモデル構築

第一段階で救急現場での音声/動画の取得に難渋しているため、第二段階のセマンティックデータモデルの構築のため、過去の救急初療部で撮影/保存された音声・

動画の音声部分(左表)を研究チームが人力で文字起こしを行った。テキスト化した文章に対して研究代表者 山田および分担者 森下が専門用語および重要箇所のマーキングを行い、セマンティックデータモデルの学習用教師データを作成した。

ファイル名	内容	時間
A00010	来院時心肺停止	24分
A00011	来院時心肺停止	7分
A00020	来院時心肺停止	180分
A00050	狭心症	29分
A00060	腹部外傷	180分

作成した学習用教師データで重要情報の構造化(Table化)を種々の大規模言語モデル(LLM)を用いて行った。試用したLLMはGPT-3, Llama2, Swallowである。このうちLlama2は日本語の対応が不十分であり意味のある出力がなされなかった。一方でGPT-3およびSwallowは意図した構造化がほぼ達成されており、モデル構築の候補LLMとして選定された(下図)。

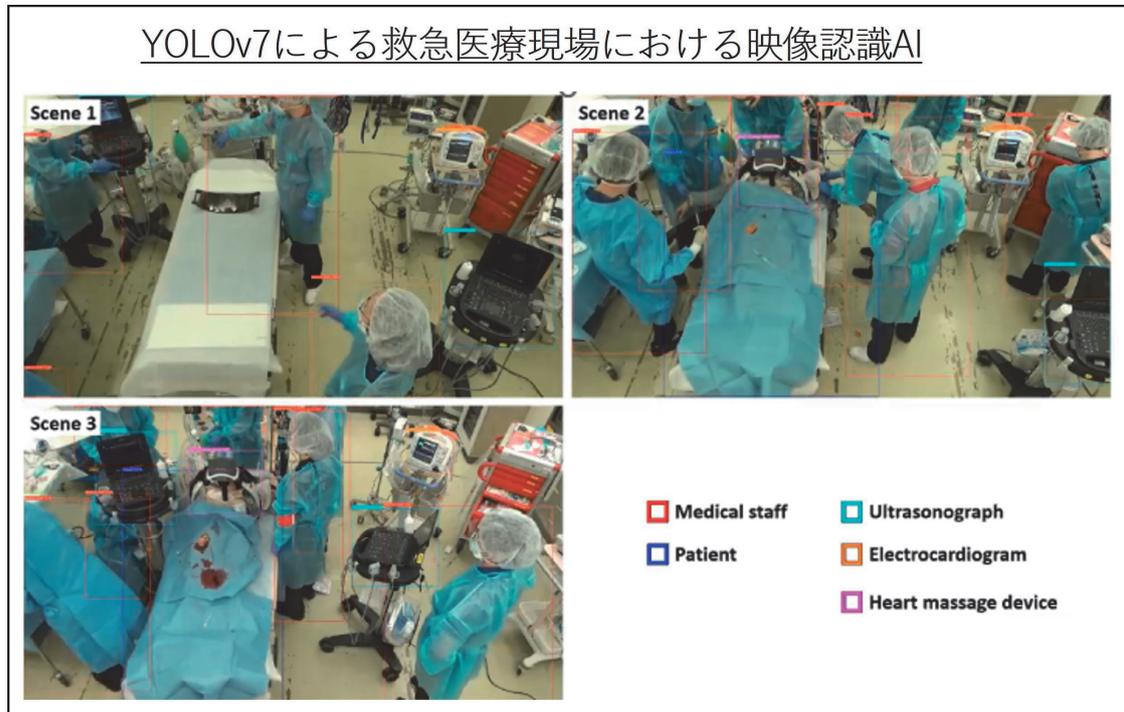
GPT-3による構造化の結果					Swallowによる構造化の結果
完了	種類	内容	指示者	報告者	### 応答: 完了 種類 内容 指示者 報告者 指示 8ミリ、22センチ A <input checked="" type="checkbox"/> 報告 7、8単位 B A <input checked="" type="checkbox"/> 指示 FFPの解を A B <input checked="" type="checkbox"/> 指示 Aライン A C
0	指示	挿管: 8ミリ、22センチ	D	-	
x	質問	輸血単位確認	A	-	
0	報告	輸血: 7、8単位	-	B	
0	指示	Aライン設置	A	-	
x	指示	FFP解凍確認	B	-	
0	質問	Aライン設置可能者確認	A	-	
0	報告	Aライン設置可能	-	C	

②映像認識AIの導入

(2)①において救急現場の会話を解析するにあたり、現場の会話は視覚による情報を前提として成り立っている事が判明した。そのため、会話から変換したテキストデータのみではAI解析における情報源として不十分であると判断した。

その解決のため、撮影された動画から映像認識AIを導入する事で、どのような場面において発生している会話であるのかをAIに組み込む事で精度向上を目指せると考えた。映像認識AIの試用として(1)②ii)のシステムにおける3Dカメラを用いた映像を用いて、撮影された映像の人物特性・医療機器の検出が可能か検証した。

深層学習による物体検出技術はYOLOv7⁶⁾を使用し、17分12秒の救急初療動画のうち、10秒おきに抽出した静止画像104枚を使用した。Vott⁷⁾を用いて教師データを作成した。その結果、医療従事者/患者/エコー機器/心電図/心臓マッサージデバイスを正常に認識する事に成功した(下図)。



【考察】

本研究は重症救急初療現場の会話を文字情報に変換し、診療系統(指示/記録/検査結果)を分類/要約するGPTモデルの確立とモニターなどに出力するシステムの開発する事を目指し研究を開始した。

研究開発は4段階に分割し、それぞれ相互的に進行させるよう計画した。本研究助成期間では、まず会話を文字情報に変換する手法の確立(第一段階)に取り組んだが、難渋した。現在では音声文字変換技術は一般商用されるレベルまで発展しているため、当初本段階は容易に達成可能であろうと予測していた。しかし、予測に反していくつかの音声文字認識ソフトウェアを試用したが、その精度は極めて低いものであった。スマートフォンに搭載される音声認識機能(Apple Siriなど)や会議の議事録の自動文字起こし商品では、収集される音源は一人の人間の声のみであることが前提に開発されている。一方で重症救急初療現場では多くの医療従事者が小チームを構成し、同時に診察/治療を行うために、多くの会話が同時になされる。加えて医療機器から様々な機械音やアラームが発生される。そのため一般的なマイクでは様々な音源が混合され、これらを分離する必要があった。周波数による分離を試みたが、機械音は分離できても不特定多数の医療従事者の音声を正確に分離することは、現在の技術では困難であ

ろうと判断した。

次に個人ごとにヘッドセットを着用することで、初めから独立して録音した音声データを文字変換、それぞれの文字変換データを統合して小チームごとの会話としてデータを作成する手法を検討した。いくつかのデバイスを検討し試用したが、録音は可能であるが、複数の音声データの時間的同期を行うことが、どのデバイスでも困難であった。複数の音声データの同期は本研究目的遂行のため必須の工程であるため、独自で音声データを同期させるプログラムを開発した。現在(1)②ii)で示した記録システムの試作機が完成している。今後、実際の救急初療現場でデータを採取し精度検証を行いつつ、データの蓄積を計画している。

一方で音声のみのデータでは、どのような状況下での会話であるのかAIが学習する事が困難である点も懸念された。この点に関して映像認識AIを応用し状況認識させ、音声データと合わせて学習することで、解決可能ではないかと考えた。映像認識AIの開発には(2)②で示したように3Dカメラによる映像を一般利用可能なYOLOv7を用いて解析が可能であることを確認した。今後(1)②ii)に記載したシステムで実際の救急初療の映像を収集し、診療状況判断を行うVision AIの開発を計画している。

第二段階のセマンティックデータモデルの構築は本研究の核をなす部分である。第一段階で救急初療の音声データを十分に収集できていないため、LLMの検証のための予備実験として人力で音声をテキスト変換し、医師監修の元で要約化すべきデータを作成した。LLMはGPT-3/(4Open-AI), Palm2(Google), Llama-2/3(Meta)など現在多くの種類が活用されている。Swallowは東京工業大学 情報理工学院と国立研究開発法人 産業技術総合研究所が共同開発したLLMである。Llama2 (Meta) を基として、日本語の文字や単語などの語彙を追加したうえで、新たに開発した日本語データを用いてモデルの構築を継続的に行う継続事前学習を行い、2023年にパラメータ数が70億パラメータ (7B)、130億パラメータ (13B)、700億パラメータ (70B) であるモデルを公開した。予備実験ではSwallowはGPT-3と同等の日本語のテキストからの構造化が可能であることが示された。一方で医療用語や救急専門用語に関しては学習が不足していると考えられた。今後は(1)のシステムで収集した情報でSwallowのファインチューニングを行い、精度の検証と向上を計画している。

第三段階(セマンティックデータモデルの可視化)、第四段階(開発したシステムの精度検証)については、第一および第二段階がある程度完成した時点で開発に取り掛かる予定である。

【結論】

本研究助成期間から研究を開始した。4段階に分割した研究開発で予測外の頓挫も経験したが、研究計画を修正し研究を進めている。重症外傷例の救急初療現場における会話を系統に応じて情報モニターに表示し、電子カルテに記録を行うための言語データ構造化を可能とするLLMを用いたAIシステムの開発を継続していく。

【参考文献】

- (1) 日本医療機能評価機構 医療事故情報収集等事業 医療安全情報. No.27, 口頭指示による薬剤量間違い. https://www.med-safe.jp/pdf/med-safe_27.pdf (accessed 2024. Oct.17)
- (2) 日本医療機能評価機構 医療事故情報収集等事業 医療安全情報. No.84, 誤った処方の不十分な確認. https://www.med-safe.jp/pdf/med-safe_84.pdf (accessed 2024. Oct.17)
- (3) 日本医療機能評価機構 医療事故情報収集等事業 医療安全情報. No.102, 口頭指示の解釈間違い. https://www.med-safe.jp/pdf/med-safe_102.pdf (accessed 2024. Oct.17)
- (4) Team STEPPS[®]: National Implementation. Agency for Healthcare Research and Quality. <https://www.ahrq.gov/teamstepps/index.html> (accessed 2024.Oct.17)
- (5) 石川雅彦. 第84回口頭指示の“解釈間違い”に関わるアクシデント事例の未然防止! 月刊地域医学 2020;34(10):844-50.
- (6) CY Wang, A Bochkovskiy, and HY Mark Liao. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. CVPR 2023. https://openaccess.thecvf.com/content/CVPR2023/papers/Wang_YOLOv7_Trainable_Bag-of-Freebies_Sets_New_State-of-the-Art_for_Real-Time_Object_Detectors_CVPR_2023_paper.pdf (accessed 2024.Oct.17)
- (7) Microsoft Vott. <https://github.com/microsoft/vott/releases> (accessed 2024.Oct.17)