ニュースリリース

令和 8 年 11 月 26 日

報道機関　各位

国立大学法人電気通信大学
学校法人順天堂 順天堂大学

「何気ない会話」から HIV 感染リスクの兆候を捉える AI へ
― 生成 AI でデータ不足を克服、新たな早期スクリーニング技術を開発 ―

【ポイント】
＊会話から HIV 感染リスクの兆候を捉える AI を開発
＊HIV 関連の対話データ不足を補う、新しい「ハイブリッドデータ拡張手法」を開発
＊「日常会話らしさ」と「HIV 感染リスクに関わる兆候」を両立した合成対話を生成
＊合成対話データによる学習で、陽性例を捉える性能が大幅に向上

【概要】

国立大学法人電気通信大学大学院情報理工学研究科情報学専攻の坂本真樹教授と順天堂大学大学院医学研究科総合診療科学の内藤俊夫教授らの共同研究グループは、日常会話に含まれる健康リスクに関連する表現から HIV 感染リスクを推定することを目指し、大規模言語モデル（LLM）を活用した新たな AI 手法を開発しました。実際の HIV 関連対話データが限られるという課題に対し、患者の語りや日常会話をもとに自然な合成対話データ（※1）を生成するハイブリッドデータ拡張手法を提案し、モデルの学習に活用しました。その結果、HIV 陽性例を検出する再現率（Recall）が 0.710 から 0.980 へ向上し、実際の患者―医療者対話を用いた検証でも、陽性例を捉える性能向上の可能性が示されました。本論文は Springer Nature の AI in Healthcare のオンライン版に 2026 年 8 月 12 日付で掲載されました。

【背景】

HIV 感染症は、抗レトロウイルス療法の進歩により、現在では適切な治療を継続することで長期的に管理可能な疾患となっています。一方で、早期の HIV 感染では、発熱、咽頭痛、発疹、倦怠感、リンパ節腫脹など、一般的なウイルス感染症と区別しにくい非特異的な症状が現れることがあり、感染に気づかないまま診断が遅れることがあります。また、HIV に対するスティグマも、検査や受診をためらわせる要因となり得ます。

近年、大規模言語モデル（LLM）の発展により、医療対話の理解や健康に関する言語情報の解析が進み、日常会話の中に含まれる微細な健康リスクの兆候を AI が捉え、早期スクリーニングを支援できる可能性が注目されています。しかし、こうした AI を開発するには、実際の患者の言語表現や心理状態を反映した高品質な対話データが必要です。特に HIV に関する情報は極めてセンシティブな医療情報であるため、プライバシー保護の観点から実際の対話データを大量に収集・公開することが難しく、AI 開発に必要な学習データの不足が大きな課題となっています。

そこで本研究では、実際の患者対話を大量に収集することなく、HIV に関する実際の語りを参考にして現実的な合成対話データを生成し、それを用いて会話から HIV 感染リスクを捉える AI を学習させるというアプローチを検討しました。公開されている HIV 当事者のインタビュー、ドキュメ

ンタリー、Vlog などから症状や心理状態、生活背景に関する情報を抽出し、それらを基に合成対話を生成することで、データ不足という課題の克服を目指しています。

【手法】

本研究では、日常会話の中に表れる体調の変化や不安の表現から、HIV 感染リスクの兆候を AI が捉えることができるかを検討しました。ただし、HIV に関する実際の会話データは、プライバシーの問題や情報のセンシティブさから大量に集めることが難しいという課題があります。そこでまず、公開されている HIV 当事者のインタビュー、ドキュメンタリー、Vlog などを収集し、そこから年齢や職業、感染段階、心理状態、症状の表現などを整理して、500 件の患者シナリオを構築しました。

次に、それらのシナリオを使って、AI に学習させるための合成対話データを作りました。方法は 2 つあります。1 つ目は、既存の日常会話の流れや言いよどみ、相槌などをできるだけ残しながら、そこに HIV 感染に関連する潜在的な兆候を織り込む方法です。2 つ目は、「体調の変化」「健康情報の検索」「将来への不安」などの話題を設定し、人物の属性や心理状態に応じて会話を一から生成する方法です。どちらの方法でも、「HIV」「AIDS」といった直接的な言葉は使わず、曖昧な体調不良や健康への不安などとして表現させています。

【成果】

こうして作成した合成対話データ 1,798 件を用いて、日本語に適応した大規模言語モデル Llama-3.1-Swallow-8B をファインチューニング（※2）しました。学習前のモデルと比較したところ、HIV 陽性例を正しく捉える再現率（Recall）が 0.710 から 0.980 に向上し、Accuracy や F1 スコア（※3）も改善しました。特にスクリーニングでは「見逃しを減らす」ことが重要であるため、再現率の大幅な向上は本研究の重要な成果です。

さらに、順天堂大学医学部附属順天堂医院で得られた実際の HIV 陽性患者 7 例の患者―医療者対話でも検証を行いました。その結果、7 例中 5 例でファインチューニング後のモデルの陽性予測確率が上昇しました。これは、合成データで学習したモデルが、実際の会話に含まれるリスク関連の微妙な表現にも反応できる可能性を示しています。

【今後の期待】

本研究では、生成 AI を活用した合成対話データによって、実際の患者対話データが限られる HIV に関する研究においても、会話からリスクの兆候を捉える AI の性能を向上できる可能性を示しました。一方、実患者データによる検証は現時点では 7 例に限られており、今後はより多くの陽性例・陰性例を含む多様なデータを用いた検証が必要です。また、長時間の会話を扱うための技術改善や、偽陽性・偽陰性を含めた臨床的な安全性の評価も重要となります。将来的には、HIV に限らず、プライバシーや希少性のため実データの収集が困難なさまざまな医療領域において、生成 AI による合成データを活用した早期スクリーニング支援への展開が期待されます。

【研究者のコメント】

HIV のようにセンシティブな医療分野では、AI の学習に必要な実データを十分に集めることが難しいという課題があります。本研究では、生成 AI を活用して現実に近い合成対話データを作成することで、この課題に取り組みました。今後は、実患者データによる検証を重ねながら、会話から健康リスクの兆候を捉え、早期スクリーニングを支援する安全で信頼できる AI 技術へ発展させたいと考えています。

（学会発表・掲載情報）

本研究の成果は国際学会で発表され、その内容が Springer Nature 刊行の『AI in Healthcare』に収録されました。

タイトル： Bridging the Data Gap: A Hybrid Data Augmentation Approach for Early HIV Screening via Dialogue Classification

タイトル(日本語訳)：データ不足を埋める：対話分類による HIV 早期スクリーニングのためのハイブリッドデータ拡張手法

著者：Zhiyang Qi, Haruto Saito, Ryuma Koseki, Mai Suzuki, Sanae Kubota, Toshio Naito, Michimasa Inaba, Chiyoko Kameue, Haruka Matsukura, Maki Sakamoto

著者（日本語表記）：斉志揚 1)2)、齋藤晴渡 1)、小関竜真 1)、鈴木麻衣 3)、久保田早苗 3)、内藤俊夫 3)、稲葉通将 1)、亀上知世子 1)、松倉悠 1)、坂本真樹 1)

著者所属：1)電気通信大学大学院情報理工学研究科、2)東京大学大学院工学系研究科、3)順天堂大学大学院医学研究科

eBook ISBN 978-3-032-35387-0
Print ISBN 978-3-032-35386-3
DOI：https://doi.org/10.1007/978-3-032-35387-0_33

（外部資金情報）

本研究は、厚生労働科学研究費補助金[エイズ対策政策研究事業][JPMH24HB1002]の支援を受け、多施設との共同研究の基に実施されました。なお、本研究にご協力いただいた皆様には深謝いたします。

（用語説明）

※1：合成対話データ

人同士の会話を収集したデータではなく、生成 AI で人工的に作成した会話データ

※2：ファインチューニング

大量のデータで学習済みの AI モデルに、目的に特化した追加学習を行うこと

※3：F1 スコア

再現率と適合率の両方を考慮して、AI の分類性能を総合的に評価する指標。再現率とは、実際に陽性であるもののうち、AI が正しく陽性と判定できた割合。適合率とは、AI が「陽性」と判定したもののうち、実際に陽性だった割合。誤って陽性と判定する「偽陽性」の多さと関係する指標

【連絡先】

＜研究内容に関すること＞

電気通信大学大学院情報理工学研究科情報学専攻
【職名】教授
【氏名】坂本　真樹
Tel：042-443-5535　　E-Mail：maki.sakamoto@uec.ac.jp

順天堂大学大学院医学研究科総合診療科学
【職名】教授
【氏名】内藤　俊夫
Tel：03-3813-3111　　E-Mail：naito@juntendo.ac.jp

＜報道に関すること＞
電気通信大学総務部総務企画課広報係
Tel：042-443-5019　Fax：042-443-5887
E-Mail：kouhou-k@office.uec.ac.jp

順天堂大学　総務局 総務部 文書・広報課
Tel：03-5802-1006　E-mail：pr@juntendo.ac.jp