研　究　報　告　書

「医療ビッグデータからの病態進行のシミュレーションによる先制医療に向けた研究開発」

研究タイプ：通常型
研究期間：　2016 年 12 月～2020 年 3 月
研 究 者：　河添　悦昌

1．　研究のねらい

- 医療への AI 応用が本格化しているなか、画像や動画を対象とした AI は撮影モダリティの数だけ開発され、多くの報告がなされている。その一方で、投薬オーダや保険登録病名、検査結果といった種類のデータは、多くの系列からなる時系列としての性質を持つデータであるが、全系列が同時に測定されず、また測定間隔が不規則であることが解析上の問題となる。本研究では、このような多系列からなる医療データを利用する手法の開発を目指す。
- このような技術開発が必要な一方、医療分野における予測モデルを臨床現場で活用するためには、技術開発だけでは十分ではない。諸外国と比べ医療が普及・成熟している本邦において、AI による予測モデルの臨床現場への導入を進めるためには、予測モデルが臨床のアウトカムに与える具体的なインパクトを示す必要がある。開発した予測モデルを利用したと仮定した場合に、このようなアウトカムに与える影響が同時に提供できれば、導入を進めるための一つの手がかりになると考えられる。

2．　研究成果

（1）概要

- 電子カルテデータは longitudinal なデータを多く含み、これは 1 つの対象について経過を追い複数の時点で観察結果を記録したものであるが、このことが予測モデルの開発を困難としている。例えば、血液や尿などの検体検査結果は典型的な longitudinal データであり、極めて多くの項目からなる多変数の時系列データであるが、全項目が同時に測定されず、また測定間隔が不規則であることにより時系列データとして扱う際に多くの欠損を生じ、解析の上での問題となる。また、検体検査結果以外にも、投薬オーダや保険登録病名などの情報を利用する上でも同様の問題が生じる。本研究の狙いの一つは、複数の医療データを時系列データとして活用する手法を開発することである。
- 本研究もう一つの狙いである、予測モデルが臨床のアウトカムに与える効果の見積もりについて、これを行った研究を提示する。この研究ではまず、高齢の入院患者における転倒の発生により延長する入院日数を、因果推論の手法により推定する。次に、診療テキストを利用して入院患者における転倒の発生を予測するモデルを開発し、このモデルの予測に従った転倒予防介入を行ったと仮定した場合に、どの程度の入院日数が短縮されるかの見積もりを行う。このことにより、どの程度のコストをかけてまで予測モデルを導入するべきかを判断する一つの手がかりを提供できると考えられる。

（2）詳細

研究テーマ 1：多系列医療データの表現手法と深部静脈血栓の発生予測モデルの開発

- これまで、不等間隔で測定される血液検査の欠損値の補間に畳み込みオートエンコーダを使った手法を適用し、線形補間に比べ精度の向上を認めること、ならびに、糖尿病検査項目の系列から、欠損を考慮したエンコーダ・デコーダーモデルを用いて低次元特徴量を取り出し、この特徴量を元に予後分類を行うモデルの効果を報告してきた。
- 最終年度はこれを発展させ、多系列の共起関係から求めた分散表現によって医療データを表現する手法、ならびに深部静脈血栓の発生を予測するタスクにおける表現の有効性を検討した。
- 提案手法による特徴量を説明する。1症例における多系列を週の単位で量子化するとともに、値を 2 値に離散化する。同じ週において同時に値が 1 となる系列を共起している系列とし、全系列の共起頻度をカウントした行列 M から正の相互情報量を算出した行列 M_{PPMI} を得る。M_{PPMI} を特異値分解し、特異値の小さなものを一定の基準で切り捨てることで次元を削減する。削減された行列において系列に対応する行をその系列の分散表現とする。同じ週に異なる系列の値が1となる場合には、分散表現ベクトルの各次元の最大値を採用することで複数の系列を集約する。提案手法による特徴量とベースライン（代表値）による特徴量を入力とするロジスティック回帰により DVT の有無を予測するタスクの精度を比較した。また、提案手法の特徴量を系列データとして扱い、bi-LSTM により DVT 有無を予測するモデルの精度の評価も行った。
- 提案手法による特徴量とベースラインの特徴量を入力とするロジスティック回帰モデルの ROC-AUC は、それぞれ 0.62 と 0.61 であり有意な差は見られなかった。また、提案手法による特徴量を Bi-LSTM に入力するモデルの精度は 0.61 であり、同様に有意な差は見られなかった。
- 提案手法による特徴量とベースラインの特徴量を入力とするロジスティック回帰モデルの精度に差が見られなかった理由として、実験に利用したデータセットは血液検査結果のみで DVT を推定するに十分な情報が含まれていたと考えられ、他カテゴリの系列との共起関係も埋め込むことのできる、提案手法の優位性を示すための適切なデータセットではなかった可能性が考えられた。

研究テーマ 2：診療テキストからの機械学習モデルによる高齢患者の転倒予測と予測がもたらす効果の検討

- 医療機関において、高齢の入院患者の転倒は寝たきりや要介護状態をきたすことから、これを予防することが重要課題である。多くの病院では、入院時のリスクアセスメントツールにより転倒評価が行われるが、手作業による労力の問題や結果の再現性の問題から、電子カルテに記録される情報を 2 次的に活用してより高い精度の転倒予測を行うことが期待される。また、転倒の発生によりどの程度入院日数が延びるかを具体的に見積もった研究は知られていない。この日数が見積もられることにより、どの程度のコストを掛けてでも転倒を予防するべきかの目安になる可能性がある。

3. 今後の展開

- 研究テーマ 1 の多系列医療データの表現手法と深部静脈血栓の発生予測モデルについて、

疎な時系列情報として観測される医療データを表現する方法として、同じく離散的な情報を扱う自然言語処理の分散表現の手法に着想を得て、これを適用する手法を提案した。提案手法を評価するために用いたデータセットは実診療から得られた深部静脈血栓の発生有無がラベル付けされたものであるが、様々なバイアスの影響で提案手法を評価するために適切ではなかった可能性があるため、他のデータセットでの評価が必要であると考えている。

- 研究テーマ 2 の診療テキストからの機械学習モデルによる高齢患者の転倒予測と予測がもたらす効果の検討について、本研究ではまず因果推論の手法により、転倒が発生することにより延長する入院日数を見積もった。次に、診療テキストから転倒の予測を行う機械学習モデルの精度を評価し、従来のリスクアセスメントツールに用いられるロジスティック回帰の精度と比較を行った。最後に、モデルの予測に従ったと仮定した場合に入院期間がどの程度短縮するかとその損益分岐点を見積もった。研究の結果、診療テキストには、転倒に関するリスクアセスメントツール以上の情報が含まれていることがわかり、機械学習モデルがこれを利用して一定の精度で予測が可能になったと考えられた。また、これまで、高齢の入院患者が転倒することにより延長する入院日数の具体的な値は知られていなかったため、本研究で示す結果が病院内の転倒予防対策における一つの指標になる可能性がある。更に、研究の過程で作成した事前学習済みの汎用的言語モデルを一般に公開することにより、医療分野での自然言語処理の精度向上に貢献する可能性がある。

4. 自己評価

- 採択時の提案研究は、リアルワールドの医療データを利用するものであるため、電子カルテからこれを抽出し、大規模なデータセットを用意することの困難さとデータセットに含まれるバイアスの影響もあり、当初想定していた目的に十分に達成できなかった。一方で、医療の分野では、機械学習による予測モデルの精度を示すだけでは臨床現場への普及は難しく、予測モデルがもたらす臨床アウトカムの見積もりが必要であると考えた。研究期間の途中から追加した研究テーマ 2 ではこのことに取り組み、これまで知られてこなかったことを明らかにするとともに、予測モデルがもたらす具体的な臨床アウトカムが提示できたと考えている。

5. 主な研究成果リスト

（1）論文（原著論文）発表

1. Hayakawa M, Imai T, Kawazoe Y, Kozaki K, Ohe K., Auto-Generated Physiological Chain Data for an Ontological Framework for Pharmacology and Mechanism of Action to Determine Suspected Drugs in Cases of Dysuria., Drug Safety. 2019,42:1055-1069

（2）特許出願

研究期間累積件数： 0 件（公開前の出願件名については件数のみ記載）

（2）その他の成果（主要な学会発表、受賞、著作物、プレスリリース等）

- 河添 悦昌, 嶋本 公徳, 篠原 恵美子, 山口 亮平. 事象の共起関係から求めた分散表現を用いた多次元医療データによる深部静脈血栓症の発症予測. 第 2 回日本メディカ

ル AI 学会学術集会（2020/01/31）
- 河添悦昌，倉沢央，岩井聡，香川璃奈，大江和彦，状態空間モデルと深層ニューラルネットワークによる検体検査結果の欠損値推定精度の比較，医療情報学 37(Suppl.), pp.820-824, 2017.