

患者状態把握を目的とした機械学習と共起有向グラフによる診療プロセス解析

山下貴範^{*1}, 若田好史^{*2}, 中熊英貴^{*3}, 野原康伸^{*4}, 岡田美保子^{*5}, 中島直樹^{*1},
副島秀久^{*3}

^{*1} 九州大学病院, ^{*2} 徳島大学病院, ^{*3} 済生会熊本病院,

^{*4} 熊本大学大学院先端科学研究部, ^{*5} 一般社団法人医療データ活用基盤整備機構

Clinical Process Analysis for Grasping Inpatient Condition with Machine Learning and Directed Graph

Takanori Yamashita^{*1}, Yoshifumi Wakata^{*2}, Hideki Nakaguma^{*3},
Mihoko Okada^{*4}, Naoki Nakashima^{*2}, Hidehisa Soejima^{*3}

^{*1} Kyushu University Hospital, ^{*2} Tokushima University Hospital, ^{*3} Saiseikai Kumamoto Hospital, ^{*4} Faculty of Advanced Science and Technology, Kumamoto University, ^{*5} Institute of Health Data Infrastructure for All

抄録: 施設間の診療データ解析のためには、データの標準化と構造化が必要である。AMED ePath 事業では、診療データを標準出力するインターフェースとリポジトリを開発し、施設間のデータ統合作業の簡便化を達成した。本稿では、ePath 事業の複数施設の DPC とクリニカルパスデータに対して、機械学習と共起有向グラフを用いて、患者状態の時系列解析を行った。長期在院では AUC=0.913、寝返り、衣服着脱、呼吸状態、循環状態が抽出され、退院先では AUC=0.773、口腔清潔、食事が抽出された。時系列的関連が把握でき、診療プロセス解析に有用であることが確認された。

キーワード ePath、クリニカルパス、DPC、機械学習、共起有向グラフ

1. はじめに

医療施設間のベンチマークや統合解析には、データ収集や加工に多くの人手と時間を費やし、データの標準化と構造化が必要である。クリニカルパス(以下、パス)の Basic Outcome Master が 2019 年に HELICS 標準に採択された。2015 年に日本医療情報学会と日本クリニカルパス学会が合同委員会を設立し、パスシステムやデータモデルの標準化について検討を重ね、2018 年度に AMED 事業「クリニカルパス標準データモデルの開発および利活用」(以下、ePath)に採択された。ePath の目的は、パスの仕組みを活用して相互運用性のある標準パスシステムを構築し、診療データを効率よく収集し、多施設間の統合解析を行い、医療の質改善に役立てることである。ePath は、各実証病院の電子カルテから診療データを出力するインターフェースとリポジトリを開発実装し、施設間のデータ統合作業の簡便化を達成している。

パス解析の先行研究では、患者状態や合併症予防などに対する統計解析モデルや Learning health system が構築され、重要な患者状態や医療行為が抽出された。しかし、それらの作用順序や時系列的解釈、複数施設データでの統合解析が課題となっている。また、医療データへの機械学習の応用も進んでおり、その結果がブラックボックスになる場合が多いが解釈の研究も進んでいる[1]。本稿では、複数施設のパスデータから、入院期間中の患者状態について、重要度の順位付けと時系列の可視化を行い、臨床的介入の優先度を判断するための手法について提案する。

2. 方法

ePath 実証病院 2 施設の 2019 年の 8 つのパス適用手術症例(人工股関節手術、胸腔鏡視下肺切除術、腹腔鏡補助下大腸切除術、経尿道的膀胱腫瘍切除術、経皮的心筋焼灼術、ロボット支援前立腺切除術、内視鏡的胃粘膜下層剥離術、経皮的冠動脈形成術)の 265 症例を対象とし、データは DPC(様式 1、H ファイル)とパスのアウトカム評価を用いる。アウトカム評価は患者状態により、達成または未達成が判断され、未達成時にはバリエーションが記録される。H ファイルとアウトカムは、術前 1 日目(POD-1)～術後 5 日目(POD5)の時系列データに変換し、患者属性を合わせて 401 の説明変数とした。目的変数は、術後在院日数 8 日以上の“長期在院(89 例)”と様式 1 の“退院先が自宅以外(32 例)”を設定した。

機械学習は、野原ら[1]の手法を用いる。Gradient Boosting Decision Tree(GBDT)で予測し、そのモデルについて SHAP(SHapley Additive exPlanation)で解釈する。さらに、共起有向グラフから変数の関連性を評価する。グラフのノードを $N_k = (\text{POD}_k, V_k)$ (H ファイルまたはアウトカム評価)と定義する。目的変数の症例群(Q)と各ノードの共起尺度(jaccard 係数)を求める。次に POD が小さく、共起尺度の大きい順にノードをソートする。Q を根として N_k は $N_1 \dots N_{k-1}$ のソート順で近いノードを選び、 N_{k-1} に向かう枝を引くように表現する。

3. 結果

1) 機械学習による予測

SHAP 総合プロット(Fig.1, 2)は、縦軸に予測に対す

る寄与度が高い変数が上から並ぶ(POD、H:H ファイル、O:アウトカム)。赤色が当該変数の値が大きい、青色は小さいことを示す。横軸は SHAP 値であり、今回のデータは、性別は 0:男性、1:女性、アウトカム評価は 0:達成、1:バリエーション、H ファイルは 0~2 で値が大きいほど介助が必要である。赤点が右に、青点が左に集まる場合、値が大きいほどリスクが高いことを意味する。

長期在院の予測器の性能は、AUC=0.913 であった。SHAP 法からは、POD1~4 の寝返りや衣服着脱、口腔清潔、移乗の介助が抽出され、創部状態、循環状態、バイタルサイン安定のバリエーションがリスクに寄与していた。体温、呼吸状態、疼痛コントロール、排泄のバリエーションはリスクが小さい結果となった(Fig.1)。退院先が自宅以外の予測性能は、AUC=0.773 であり、精度としては十分であった。SHAP 法からは、口腔清潔、移乗の介助が抽出され、患者属性からは年齢と性別(女性)、食事摂取のバリエーションが抽出された(Fig.2)。

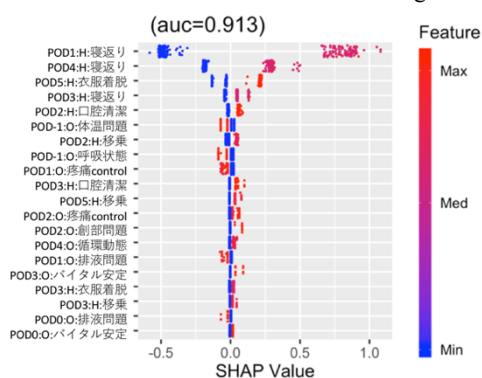


Fig.1 SHAP(長期在院)

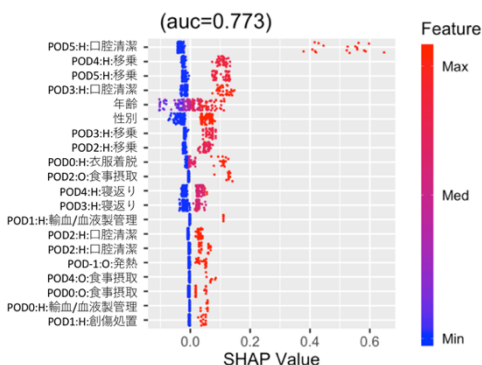


Fig.2 SHAP(退院先が自宅以外)

2) 共起有向グラフによる可視化

長期在院のグラフからは、POD3 の寝返りを中心として、上位方向に POD3~4 の“衣服着脱&体温問題&移乗&循環動態安定&疼痛コントロール”の介助、またはバリエーション発生が 21 例確認できた(Fig.3)。退院先が自宅以外のグラフは、POD1~2 の H ファイルの衣服着脱や寝返り、移乗の介助が抽出された(Fig.4)

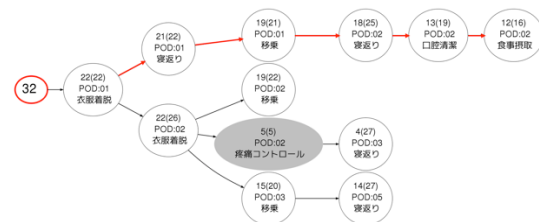


Fig.4 共起有向グラフ(退院先が自宅以外)

4. 考察

ePath 手術症例の長期在院では、“寝返り、衣服着脱、呼吸状態、循環状態”が、長期在院に寄与しており、“移乗、寝返り、食事摂取”が、退院先が自宅以外では、口腔清潔、食事摂取が確認できた。共起有向グラフでは術後の早い段階で、介助が必要な事象であることが示された。術後の観察項目は、既に医療現場でも周知されているが、順位付けすることで介入の優先度が決まり、最適な医療資源の投入や医療の効率化が可能となる。変数間の時系列的関連が把握でき、診療プロセス解析に有用であることが確認された。

5. 結語

ePath では、現在も多くのデータが蓄積され続けている。今回提案した解析手法は、機械学習による大規模なデータ解析とその時系列的解釈が可能となり、ePath 基盤を用いて多くの診療プロセス解析が実現できる。多施設の患者状態を表す品質の高いデータから、診療プロセスのエビデンス創出が期待される。

参考文献

- [4] 野原康伸, 松本晃太郎, 副島秀久, 中島直樹: Shapley Additive Explanation を用いた機械学習モデルの解釈と医療実データへの応用, 第 23 回日本医療情報学会春季学術大会抄録集,60-61, 2019.

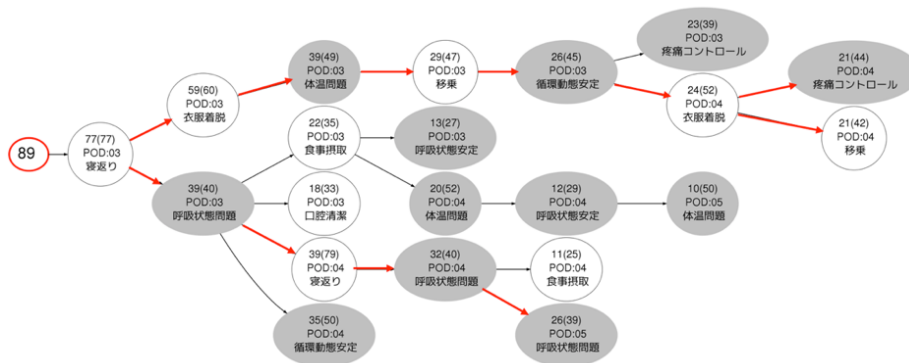


Fig.3 共起有向グラフ(長期在院): 背景色白色は H ファイル、灰色はアウトカム、赤線は強い相関を表す。ノード内: (左上から) 左側ノードを満たす症例数、ノード条件を満たす症例数、POD、H ファイルまたはアウトカム。