
一般口演**一般口演10****データマイニング・テキストマイニング**2017年11月21日(火) 16:00 ～ 17:30 F会場 (10F 会議室1004-1005)

[2-F-3-OP10-5] 急性期病院における転倒転落事故の探索的リスク要因解析野原 康伸¹, 鴨打 正浩², 中島 直樹¹ (1.九州大学病院, 2.九州大学大学院医学研究院)

【はじめに】病院における転倒転落事故は外傷や骨折につながり、患者 QoLに大きな影響を及ぼす深刻な問題の一つである。古くから防止対策が取られているものの、患者の高齢化に伴い事故は増加傾向にあり、より一層の対策が求められている。本研究では、入院患者に関する様々なデータを収集し、その中から探索的なアプローチにより転倒転落事故に関連するリスク要因を抽出することで、今後の転倒転落対策に生かすことを目的とする。

【方法】2016年2月～3月に急性期病院 Aで起きた転倒転落事故事例、および同時期に入院した20歳以上の全患者を対象として、DPCの様式1と EFファイル(各日の薬効分類毎の薬の処方)、看護必要度、転倒転落アセスメント結果を収集した。期間中の延べ入院人数は51886人日(患者数3447名)であり、918のデータ項目が抽出された。転倒転落が起きたか否かを目的変数(アウトカム)、残りを説明変数として予測器を作成した。予測器には決定木を用いた機械学習アルゴリズムの一つである Gradient Boosting Decision Treeを用いた。各説明変数が予測にどの程度寄与したかを示す変数重要度を計算し、重要度の高い説明変数に対して、当該変数がアウトカムにどのような影響を及ぼすかを Partial Dependence Plotにより調査した。

【結果と考察】作成された予測器の汎化性能は0.763であった。変数重要度が高い変数として、転倒転落リスク評価結果(スコアが高いほど事故リスクが高い)のような当然の変数だけでなく、喫煙指数(喫煙する人ほど事故リスクが高い)も重要変数として抽出された。患者がこっそり喫煙しようとして転倒した場合は、患者が正直に報告しづらく事故報告書には直接上がってこない。探索的な本手法でなければ抽出できなかった事例の可能性があり、交絡調整等の詳細解析を実施する計画である。

急性期病院における転倒転落事故の探索的リスク要因解析

野原 康伸^{*1}、鴨打 正浩^{*2}、中島 直樹^{*1}

*1 九州大学病院、*2 九州大学大学院医学研究院

Exploratory Factor Analysis of Falling Accidents in Acute Hospital

Yasunobu Nohara ^{*1}, Masahiro Kamouchi ^{*2}, Naoki Nakashima ^{*1}

*1 Kyushu University Hospital, *2 Graduate School of Medical Sciences, Kyushu University

Abstract

A falling accident in hospitals is one of the import problems that affect a quality of life of patients and the number of the accidents is increasing as aging of patients. Identifying important predicative indicators for the accident helps account for the accidents and may lead to decrease the accidents. We extracted important factors for falling accidents in an acute hospital from EHR data. We analyzed structured EHR data of 51,886 patient-days with 918 variables using gradient boosting decision tree, one of the machine learning techniques. Extracted factors include not only well-known factors such as risk scores of falling accidents but also new factors such as the number of cigarettes smoked. This method may be useful to identify previously unknown but significant factors related to the accidents.

Keywords: falling accident / exploratory factor analysis / machine learning

1. はじめに

病院における転倒転落事故は外傷や骨折につながり、患者の QoL(Quality of Life)に大きな影響を及ぼす深刻な問題の一つである。古くから防止対策が取られているものの、患者の高齢化に伴い事故は増加傾向にあり、より一層の対策が求められている。これまでの転倒転落対策は、医療者の経験と勘に基づくものが多かったが、経験と勘だけでは複雑な要因の中から、重要な関連因子を見落とすことが多いと考えられる。

2. 目的

転倒転落と関連しそうかといった主観は排し、入院患者に関連する入手可能な全てのデータを EHR などから収集する。それらのデータと転倒転落結果を用いて解析を行う探索的なアプローチにより転倒転落事故に関連するリスク要因を抽出する。抽出されたリスク要因の因果関係等について議論することなどを通じて、今後の転倒転落対策に生かす。

3. 方法

2016 年 2 月～3 月の 2 か月間に急性期病院 A に入院した 20 歳以上の全患者を対象とした探索的なリスク要因解析を行う。期間中の延べ入院人数は 51886 人日(患者数 3447 名)であった。

3.1 目的変数

転倒転落事故の報告書から、ある日に患者が転倒転落事故を起こしたか否か(True or False)を抽出し、目的変数とする。入院中に転倒転落事故を起こした患者であっても、事故日以外の目的変数は False となる。

3.2 説明変数

対象期間内の全患者に対して、各日ごとに以下のデータを抽出し、説明変数として使用する。918 のデータ項目が抽出された。

3.2.1 DPC 様式 1 情報

急性期病院 A は、DPC 対象病院の一つであり、様式 1 とよばれる退院時サマリを作成している。様式1には、病院情報や患者情報(性別、生年月日など)、入院退院情報(入院日、退院日など)、診断情報(発症病名や併存症など)、手術情報(手術名や手術日など)、そしてケア情報(入院退院時における ADL

や Japan Coma Scale など)の情報が含まれている。解析では、入院初日からの経過日数等の情報を除いて、同一患者には入院期間の全てにおいて、同一の様式 1 から抽出した情報を用いる。

3.2.2 薬効分類毎の処方(DPC E/F ファイル)

DPC E/F ファイルから、薬効分類毎の処方情報を抽出する。ある日の前後 3 日間(前日・当日・翌日)の薬の処方には 8 つの投与の組み合わせが考えられるが、各組合せは以下のような 3 種類の投与パターンに分けられる。

- 薬が事故原因となる可能性があるパターン(前日もしくは当日投与)
 - 100,101,110,010,011,111
- 薬が事故結果となる可能性があるパターン(前日未投与で、当日もしくは翌日に投与)
 - 001,010,011
- 薬が無関係と考えられるパターン(3 日間で変化なし)
 - 000,111

ここで、3 桁の数値は、各日(前日・当日・翌日)に投与(1)もしくは未投与(0)を示す。太字の数値は複数の可能性が考えられるパターンである。

各患者の薬効分類毎に、薬が原因パターンにあたるか否か、結果パターンに当たるか否かを抽出し、説明変数とする。例えば、ある患者に対する薬効分類 3229(その他の無機質製剤)の薬の投与パターンが 110 だった場合は、med3229_cause が 1,med3229_result が 0 と抽出される。

3.2.3 看護必要度

看護必要度は、患者の疾患・病態によって異なる看護サービスの量を定量的に評価するための指標であり、医学的な処置の必要性を示す A 項目と、患者の日常生活機能を示す B 項目により構成される。看護必要度は、全入院患者について毎日評価されている。本稿では、A 項目および B 項目の全ての評価項目に関して、説明変数として用いる。

3.2.4 転倒転落アセスメント

患者の転倒転落リスクを把握するため、患者のリスクスコア評価が随時実施される。アセスメントスコアシートにしたがって、年齢や既往などの評価項目ごとにスコアを求め、その合計点

によりリスク評価がなされる。評価が行われるのは、1)入院から 3 日以内、2)患者状態が大きく変化した場合(例:手術後や転倒転落事故発生後)、3)前回の評価日から一定期間経過後(概ね1週間程度)である。

転倒転落リスク評価実施日であれば、当該リスクスコアを当日の説明変数として用いることができるが、評価は毎日実施されるわけではない。当日に評価未実施の場合は直近の評価時点におけるリスクスコアを、当日のリスクスコアとする。

3.3 解析手法

1)医療データには非線形性が存在すること、2)多数の説明変数があってもうまく働くこと、3)変数重要度が解析結果を解釈するのに有用なことから、本論文では、決定木をベースとした機械学習手法である Gradient Boosting Decision Tree (GBDT)[1]を解析に用いる。GBDT を用いて、患者が当日転倒転落事故を起こすか否かを、918 個の説明変数で予測し、事故に強く関連しているのはどの説明変数か、その説明変数がどう予測に影響を与えているかを解析する。

3.3.1 ブートストラッピング

大きさ n のデータセット D が与えられたとき、ここから重複ありでランダムに n 個のデータをサンプリングし、訓練データセット D_t とする。 D_t における n 個のデータ中、63.2%のデータがユニークであることが期待される。訓練データセット D_t として一度も選ばれなかったデータ $D-D_t$ を検証用データセット D_v として用いる。 D_t を用いて学習し、 D_v を用いて、AUC (area under curve)を評価することで、予測モデルの正確性を検証する。この際、各説明変数の変数重要度の計算も併せて行う。

上記手順を 20 回測定し、AUC と各変数の重要度の平均を計算する。なお、GBDT のハイパーパラメータ(ラウンド数や決定木の最大深さ、学習率など)は、10 分割交差検証による AUC の平均を最大化するようにチューニングして決定した。

3.3.2 Partial Dependence Plot

決定木をベースとした機械学習手法では、各説明変数が目的変数の予測にどれだけ寄与しているかを示す変数重要度を計算することができる。しかし、変数重要度は説明変数と目的変数の関連性の強さを示すのみで、説明変数の値が変化するとき、目的変数がどのように変化するかは分からない。

PDP(Partial Dependence Plot; 部分従属グラフ)は、Friedman[1]が提案した、高次元関数を可視化するための手法の一つであり、目的変数と説明変数の関係を示すのに有用である。 $f(x_1, x_2, \dots, x_p)$ を p 個の説明変数を用いてアウトカムを予測する予測器とし、 i 番目($1 \leq i \leq p$)の説明変数 x_i の効果を可視化することを考える。このとき PDP を表す $F_i(x)$ は、次式で与えられる。

$$F_i(X) = \frac{1}{N} \sum_{j=1}^N f(x_1^j, x_2^j, \dots, x_{i-1}^j, X, x_{i+1}^j, \dots, x_p^j)$$

ここで、 x_i^j は、患者 j ($1 \leq j \leq N$) の i 番目の説明変数の値を表すものとする。この式は、 N 人全ての患者の、効果を見たい説明変数 x_i を全て X に変化させたと仮定したときのアウトカム予測値の平均であると見ることができる。 $F_i(x_a)$ と $F_i(x_b)$ では、 i 番目の説明変数をそれぞれ x_a と x_b に変えただけで、その他の説明変数は全て同じものを用いている。したがって、 $F_i(x_b) - F_i(x_a)$ は、他の説明変数の影響を除去した、 i 番目の説明変数を x_a から x_b に変えた時の純粋な変化を表していると解釈でき[2,3]、交絡調整もなされているものとみなせる。

4. 結果

作成された予測器の予測(汎化)性能を AUC (Area Under Curve) で評価したところ 0.763 であり、良い予測性能が得られた。変数重要度が高い上位 20 変数に関する変数重要度のグラフを図 1 に示す。Gain が大きいほど、予測への寄与度が高いことを示し、転倒転落リスクスコア(重要度 1 位)や入院日からの経過日数(同 2 位)、入院時 ADL 評価[移乗] (同 8 位)のような当然に予測される因子に交じって、喫煙指数(同 3 位)という意外な説明変数が重要因子として抽出された。

重要度上位の説明変数に関して、その変数がアウトカムとどのように関連しているかを調べるために、PDP を作成した。図 2 は、重要度 1 位の転倒転落リスク評価に関する PDP である。横軸が説明変数の値であり、縦軸が転倒転落の起こしやすさ(値が大きいほどリスクが高い)を示す。一般的に危険度 2 と判定される 6 点以上から、転倒転落リスクの予測スコアが高まるのが見て取れ、GBDT を用いて作成した予測器が非線形な関係をうまく扱えていることが分かる。

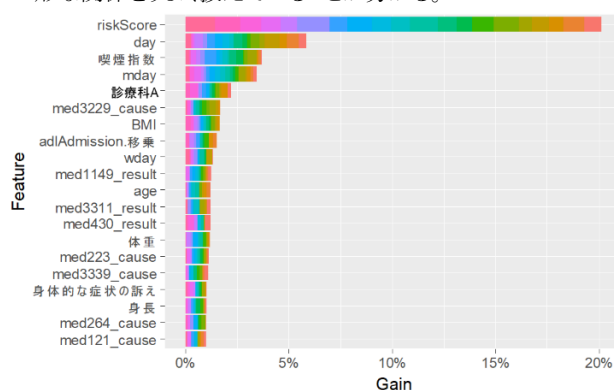


図 1 転倒転落予測への寄与が大きい説明変数 Top20

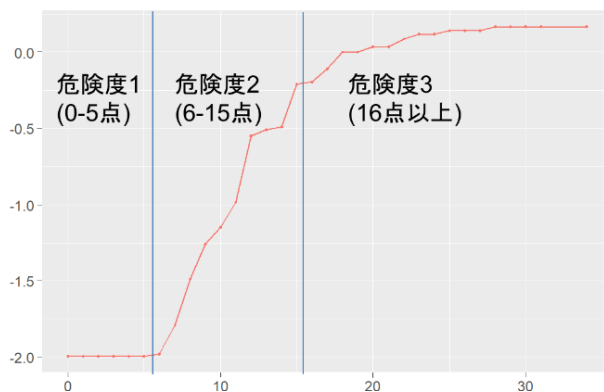


図 2 転倒転落リスク評価に関する PDP

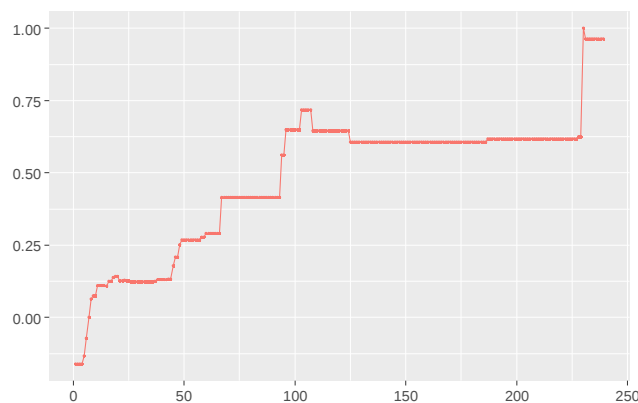


図 3 入院からの経過日数に関する PDP

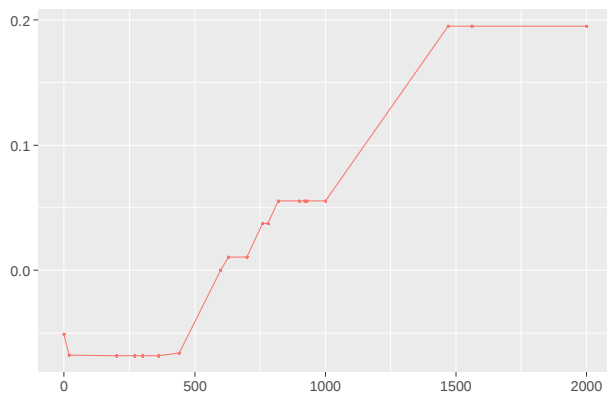


図 4 喫煙指数に関する PDP

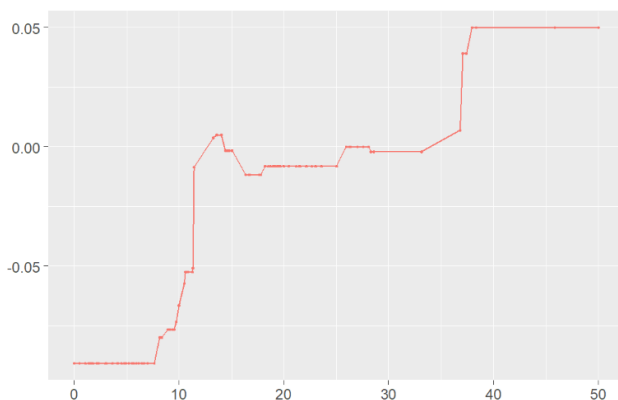


図 5 1日の推定喫煙本数に関する PDP

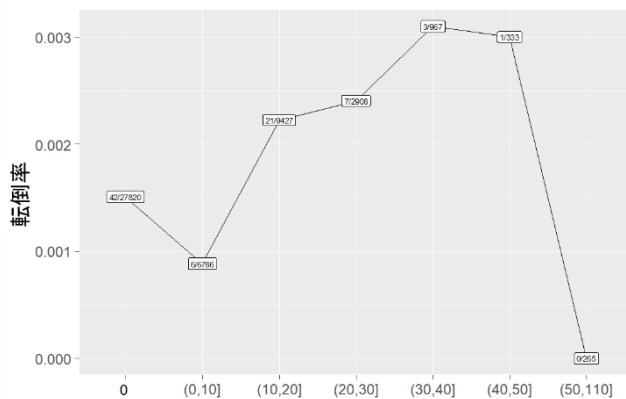


図 6 1日の推定喫煙本数と実際の転倒率の関係
(交絡調整なし)

図 3 は、入院からの経過日数に関する PDP である。入院日数が経過している患者ほど転倒転落を起こすリスクが高いことが分かる。入院日数が経過している患者は、比較的移動距離が長く、事故を起こしやすいことなどを反映しているものと考えられる。

図 4 は、重要度 3 位の喫煙指数に関する PDP である。喫煙指数は、DPC 様式 1 に記載されている値で、(1 日の喫煙本数)×(喫煙年数)で計算される。PDP によると喫煙指数が高い患者ほど転倒転落を起こしやすいとの結果が得られた。

喫煙年数は、年齢による影響を受けるため、年齢による交絡の可能性が考えられる。そこで、喫煙開始年齢を20歳と仮定して、喫煙年数を推測し、そこから計算された1日当たりの

推定喫煙本数を説明変数として再解析を実施した。喫煙指数を1日の推定喫煙本数に置き換えても、予測精度は変わらず、変数重要度も依然として3位のままであった。図 5 は、1日の推定喫煙本数に関する PDP である。喫煙本数が多いほどリスクが高い関係が得られた。

使用した解析方法による誤検知の可能性を考え、GBDT や PDP を使わずに、1日の推定喫煙本数と実際の転倒率の関係を直接調べた結果を図 6 に示す。1日の推定喫煙本数が10本～50本の群では、他の群と比べて転倒率が高いとの結果が依然として得られた。

5. 考察

今回の解析では、リスクスコアとは別に喫煙指数(1日の推定喫煙本数)が重要なリスク要因(重要度 3 位)として抽出され、喫煙指数と転倒転落事故の間に強い相関が認められた。今回の解析結果からは、喫煙本数を転倒転落リスク評価の評価指標の一つとして考慮すべき可能性が示唆される。なお、喫煙が事故と関連しうるとの事前知識があったわけではなく、喫煙指数が DPC 様式 1 に記載され、説明変数として解析した結果、関連性が見出されたものである。

相関関係は、必ずしも因果関係を意味するわけではないが、喫煙習慣が原因だとすると、喫煙者の行動パターンや患者背景などについて、さらに精査をする必要があると思われる。一方、今回の実験対象期間を含む 2014 年 4 月から 2016 年 3 月までの 2 年間の転倒転落の事故調査報告の事故原因を「喫煙」でキーワード検索したところ、該当するのは 2 件のみであり、関連性を積極的に支持する結果は得られなかった。

喫煙と転倒転落事故の双方の原因となる潜在変数(交絡因子)の存在による疑似相関の可能性は依然として捨てきれない。しかし、喫煙は調査報告書を読んでも分からない、探索的な本手法だからこそ抽出できたリスク因子である可能性がある。今後対象期間を拡大するなどの詳細な解析が必要であると考えられる。

また、現状の解析では、診療科間の差をあまり考慮せずに、病院全体のデータによる解析にとどまっている。各疾患、病棟、職員に特有の因子に関しても調査を進める必要があると考えられる。

6. おわりに

急性期病院 A において、2 か月間の全患者データ(延べ入院人数 51886 人日)を対象とした転倒転落の予測器を機械学習手法の一つである GBDT によって作成し、転倒転落リスク要因の抽出を行った。転倒転落リスクスコアなどの既知因子だけでなく、喫煙指数のような未知のリスク因子が抽出された。探索的な本手法を用いることにより、既知の因子のみならず、要因として重要であるにも関わらず関連性を疑いにくい未知の因子も抽出できた可能性がある。今後、本手法を用いてさらに詳細な解析を実施することで、原因を可視化し特定していくことが医療の質向上に有用であると考えられる。

参考文献

- 1) Friedman JH. Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics Vol. 29, No. 5, pp. 1189-1232, Oct. 2001
- 2) Yasunobu Nohara, Yoshifuru Wakata and Naoki Nakashima. Interpreting Medical Information Using Machine Learning and Individual Conditional Expectation, Proceedings of the 15th World Congress on Medical and Health Informatics (MedInfo2015), p.1073, Aug. 2015

- 3) 野原 康伸, 若田 好史, 中島 直樹, “機械学習による医療情報の解釈方法の提案”, 第 1 回医用人工知能研究会, Vol. 001, No.12, 2015 年 9 月