一般口演
# 一般口演21
# 医療データ分析９（機械学習・テキストマイニング２）
2018年11月25日(日) 10:40 〜 12:10 D会場 (4F 413+414)

[4-D-2-1] 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

○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】近年、深層学習をはじめとする機械学習手法の医療分野への応用が進んでいる。従来、医療統計でよく用いられてきた線形モデルと異なり、機械学習モデルでは中身がブラックボックスの場合が多く、なぜそのような結果が出たのかが説明できない点が大きな課題とされてきた。 SHAP　(SHapley Additive exPlanation)は、機械学習の結果を解釈する手法の一つである。経済学者シャプレーが考案した「複数人が協力して仕事をした場合の利益の公平分配方法」を応用し、機械学習の出力を各説明変数の貢献度に応じて決まるSHAP値の線形和の形で表現する。本発表では、 SHAPをロジスティック回帰分析と対比させることでその意義を議論するとともに、医療実データを基に構築した機械学習モデルに適用して解釈を行い、従来解釈手法との比較を行う。【方法】急性期病院 Aで2年間に起きた全転倒事故事例、および同時期に入院した20歳以上の全患者データを基に、機械学習を用いて転倒に関する予測器を作成する。 Gainを用いた変数重要度および Partial Dependence Plotを用いた解析結果（2018年春季学術大会で発表、従来手法）と SHAPを用いて新たに解釈した結果を比較する。【結果と考察】アウトカムへの影響が大きな説明変数 Top20を、従来手法と SHAPを用いて求めたところ、両者で概ね一致したが、一部異なるものが見られた。 SHAPでは、体重・身長・ BMIのように関連性の高い値をグループ化して、その合成変数の重要度を求めることができ、解析結果をより分かりやすい形にすることが可能となった。また交互作用項があると、同じ説明変数の値に対して、 SHAP値がばらつくことを利用し、交互作用項を容易に探索することができた。 SHAPは、従来の解釈方法よりも解釈性などの面で優れていた。

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

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

Explanation of Machine Learning Models Using Shapley Additive Explanation and Application for Real Data in Hospital

Yasunobu Nohara [*1], Masahiro Kamouchi [*2], Naoki Nakashima[*1]
*1 Kyushu University Hospital, *2 Faculty of Medical Sciences, Kyushu University

For using machine learning techniques in decision making process, explainability and interpretability of models is important. In this paper, we adopt Shapley additive explanation (SHAP), which is based on fair allocation for profit among many people in economics, for interpreting gradient boosting decision tree model. We compare the explanation results between SHAP and existing gain method.

Keywords: explainable machine learning, Shapley additive explanation, gradient boosting decision tree

1. はじめに

近年、深層学習をはじめとする機械学習手法の医療分野への応用が進んでいる。従来、医療統計でよく用いられてきた線形モデルと異なり、機械学習モデルでは中身がブラックボックスの場合が多く、なぜそのような結果が出たのかが説明できない点が大きな課題とされてきた。SHAP (SHapley Additive exPlanation)[1]は、機械学習の結果を解釈する手法の一つである。経済学者シャプレーが考案した「複数人が協力して仕事をした場合の利益の公平分配方法」を応用し、機械学習の出力を各説明変数の貢献度に応じて決まる SHAP 値の線形和の形で表現する。本発表では、SHAP をロジスティック回帰分析と対比させることでその意義を議論するとともに、医療実データを基に構築した機械学習モデルに適用して解釈を行い、従来解釈手法との比較を行う。

2. 機械学習モデルの解釈に関する先行研究

2.1 一般化線形モデル

一般化線形モデルは、予測値をリンク関数によって変換したものが、説明変数の線形結合となるようなモデルである。医療統計解析でもよく使われるロジスティック回帰モデルは、確率 p の線形オッズ(ロジット)をとると、

$$\log\frac{p}{1-p} = \sum_{i=1}^{k} a_i x_i + \mathrm{b}$$

と表されるから、一般化線形モデルの一種である。ここで、a_i は偏回帰係数、x_i は説明変数、b は定数、k は説明変数の個数である。

ロジスティック回帰を例に、各説明変数の値が変化すると、どのように予測値が変化するかを考える。説明変数 x_i を 1 単位だけ変化させたとき、対数オッズは a_i だけ増加することになる。したがって偏回帰係数 a_i の符号を調べることによって、説明変数 i がアウトカムに対して正の効果を持つのか、負の効果を持つのかが分かる。また x_i が平均 0、分散 1 となるように正規化した値である場合は、a_i の絶対値を比較することで説明変数同士の間でアウトカムへ与える影響の大きさ(変数重要度)を比較することができる。

このように一般化線形モデルでは、説明変数とアウトカムの関係が単純明快に示されていることから、人間による解釈が可能である。しかしながら、説明変数とアウトカムの間に非線形の関係がある場合や、説明変数同士の間で交互作用がある場合は、これらの関係をモデルに取り込むことが難しく、予測精度を上げることが困難である。

2.2 決定木/アンサンブル木モデル

決定木モデルとは、条件分岐を繰り替えして、対象を分割していくことで、予測を行うモデルである。予測値が、if-then ルールで記述でき、説明変数とアウトカムの関係が単純明快に示せるが、決定木単体では高い予測精度を得ることは難しい。

アンサンブル木は、決定木を弱学習器として多数組み合わせたアンサンブル学習により、予測精度を高めた手法であり、Random Forest や Gradient Boosting Decision Tree(GBDT)[2]などがある。アンサンブル木は、説明変数とアウトカムの間にある非線形な関係や、高次元の交互作用項を扱えることができる。大規模高次元データにおいて経験的に良い予測精度が得られることから、アンサンブル木は機械学習の分野では広く使われている。ただし、決定木を多数組み合わせたことで予測精度は向上するが、直接解釈することは困難となる。

アンサンブル木において、各説明変数がアウトカムに与える影響を調べる方法として Partial Dependence Plot(PDP)[3,4]がある。また、決定木ではゲインと呼ばれる値を用いることで、各説明変数がアウトカムに与える影響の大きさ(変数重要度)を比較することが良く行われている。

2.3 Shaplay Additive Explanation

アンサンブル木において、変数重要度を表すものとしてゲインが広く使われているが、一貫性がないという問題が Lundberg らによって指摘されている[1]。これは、ある変数が予測に実際に与えている影響と比べて、過大もしくは過小評価されてしまうことがある問題であり、アウトカムに関連する重要因子の抽出に Gain を用いる際に問題が生じうる。

Gain に一貫性がないという問題に対して、Lundberg らは SHapley Additive exPlanation (SHAP)を使うことを提案している[1]。経済学者 Shaplay が考案した「複数人が協力して仕事をした場合の利益の公平分配方法」であるシャプレー値を応用し、機械学習モデルのアウトカムを各説明変数の貢献度に応じて決まる SHAP 値の線形和の形で表現する。シャプレー値は、利益分配方法として最適な性質を持つことが経済学で証明されており、SHAP 値を変数重要度として用いた際に一貫性を持つことも保証される。

SHAP 値は、原理的には深層学習を含む全ての機械学習アルゴリズムに適用することが可能であるが、計算量の観点から実用的なアルゴリズムは現状、アンサンブル木によるもの[5]等に限られている。

3. 目的

これまでに GBDT 等のアンサンブル木で作成した予測器に対して、ゲインを用いた解釈を行う、筆者らもいくつか論文を発表している[6]。ゲインを用いた従来手法では、一貫性の点で問題が生じうることから、SHAP 値を用いた解釈方法で再解釈を行い、どのように解析結果が変わるかを検証する。また、SHAP をロジスティック回帰分析と対比させることでその意義を議論する

4. 方法

急性期病院 A で 2 年間に起きた全転倒事故事例、および同時期に入院した 20 歳以上の全患者データを基に、機械学習を用いて転倒に関する予測器を作成する。この予測器に関してゲインを用いた変数重要度および Partial Dependence Plot を用いた解析結果に関しては、2018 年春季学術大会で発表済み(以下、従来手法とよぶ)である[6]。同じ予測器に対して、SHAP を用いた再解析を行い、両者の結果を比較する。

以下は、予測器の作成に関する内容で、既発表の内容であるが再掲する。

4.1 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。入院期間中に転倒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。

4.2 説明変数

転倒転落後に取得できる値ではなく、それ以前に取得可能な 632 変数を用いる。

(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)、入退院情報(入院日など)、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(入退院時における ADL や Japan Coma Scale など)の情報が含まれている。

(2) 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、対象期間では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。全入院患者について毎日評価されており、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。

看護必要度はその日の夜に評価されることが多いため、事故後の評価がその日の評価値となる可能性がある。そこで、当日の看護必要度情報は、翌日の説明変数として用いる。

(3) 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(1週間程度)経過後である。

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。そこで、当日に評価未実施の場合は直近の評価時点のリスクスコアを、当日のリスクスコアとする。なお、事故当日に評価を行っている場合は、評価時刻が事故時刻より前の場合のみ、当該評価を用いるものとし、そうでない場合は直近の評価を用いるものとする。

4.3 予測器の作成方法

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)を解析に用いる。GBDT を用いて、患者が当日転倒事故を起こすか否かを、632 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後に影響を与えているかを解析する。

5. 結果と考察

5.1 一般化線形モデルと SHAP の比較

交互作用項のない一般化線形モデルにおいて、患者 i に対する説明変数 j の SHAP 値を計算すると、

(説明変数 j の偏回帰係数) ×[(患者 i の説明変数 j の値)-(説明変数 j の平均)]

となる。各患者のリスクが、各説明変数の SHAP 値の線形和で表されるので、線形回帰の同様の手法で解釈すればよいことが分かる。したがって、ロジスティック解析分析で言えば、SHAP 値は対数オッズのように扱えばよい。

なお、交互作用項がある場合は、その値を関係する説明変数の間で、公平に分配して、上記の値に足し合わせたものになる

5.2 SHAP による説明変数度の導出と SHAP summary plot

作成した予測器とそれを用いた各患者のリスク推定値を基に患者毎に SHAP 値を導出し、その値を図示したもの(SHAP 総合プロット)を図 1 に示す。

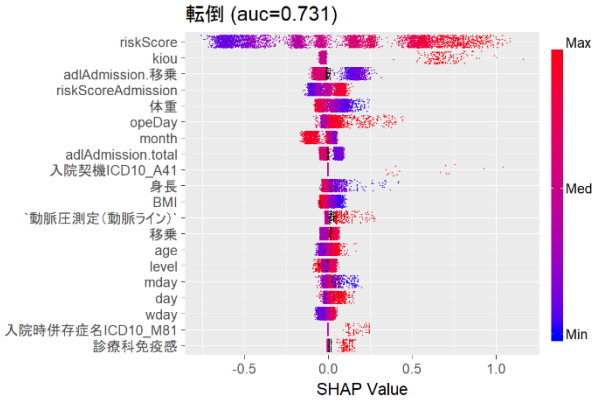

図 1 SHAP 総合プロット

SHAP 総合プロットでは、予測に対する寄与度が高い説明変数が上から順に並ぶ。各行は説明変数に対応し、その中の各点が 1 患者を表す。赤色が患者の当該変数の値(SHAP 値ではない)が大きいことを、青色は小さいことを、黒色は欠損値を示す。

一番上位の riskScore(当日の転倒転落リスクスコア)に注目すると、 SHAP 値が小さい左側の方(最小値: -0.75)に青点が、

SHAP 値が大きい右側の方に赤点(最大値 1.0)が分布していることが分かる。これは、riskScore が大きいほど転倒リスクが高いことを示しており、オッズ比が最大 5.8(=exp(1.0+0.75))倍程度となることが分かる。既存手法では説明変数毎に PDP を描画しないと各説明変数とアウトカムの関係は分からなかった。SHAP 値を用いた手法により、変数の重要度の関係だけでなく、各変数のアウトカムに対する影響を含めて全体的に俯瞰することが可能となった。

なお、Lundberg らの元々の手法は、説明変数毎に全ユーザの SHAP 値の絶対値の和を求め、変数重要度の順序付けに用いている。しかし、本論文では説明変数毎の SHAP 値の分散(標準偏差)で順序付けを行った。この順位付けの方法は、次節で述べる説明変数のグループ化を考えると、自然な考えである。詳しくは、後述する。

5.2 ゲインと SHAP による各説明変数の重要度の違い

アウトカムへの影響が大きな説明変数 Top20 を、従来手法と SHAP を用いて求めた。表 1 に両者の順位を比較したものを示す。

表 1　新旧手法での変数重要度順位の比較

	既存手法での順位	新手法で順位
当日の転倒転落スコア	1	1
転倒経験あり	2	2
入院時 ADL 移乗	3	3
手術日からの経過日数	4	6
当日の年月日の日	5	16
体重	6	5
入院日からの経過日数	7	17
入院日の転倒転落スコア	8	4
当日の年月日の月	9	7
年齢	10	14
BMI	11	11
1 日当たりの喫煙本数	12	
身長	13	10
入院契機病名:敗血症	14	9
曜日	15	18
入院契機病名・大動脈瘤	16	
入院時 ADL 総合	17	8
看護必要度レベル	18	15
移乗	19	13
肺炎の重症度分類	20	
動脈圧測定		12
入院併発症名・骨粗鬆症		19
診療科 A		20

両者の順位を比較すると、当日の転倒転落スコア、転倒経験あり、入院時 ADL 移乗という Top3 において順位は完全に一致した。それ以下では、若干順位が異なる場合がみられるが、Top20 のうち、17 の変数が共通していることが分かった。既存手法では Top20 の重要変数に挙がっていた、1 日当たりの喫煙本数、入院契機病名・大動脈瘤、肺炎の重症度分類が 21 位以下に下がり、代わりに、同脈圧測定、入院併発症名・骨粗鬆症、診療科 A が重要変数として挙がってきた。

5.3 SHAP を用いた説明変数のグループ化

変数重要度 Top20 を見てみると、体重・身長・BMI のように関連性の高い値が入っていることが分かる。似たような変数が入っている場合、各変数のアウトカムへの影響が分かりづらくなるため、代表的な値だけを残して残りの値を除去し、再解析を行うということが良く行われる。ただし、この場合は再解析に対する余計な計算時間が必要となるし、説明変数を除去したことにより予測性能も低下してしまう。そこで、直接解釈することを考える。

体重・身長・BMI のアウトカムに対する貢献度に応じて決まる SHAP 値を合計したものは、体重・身長・BMI という体格に関する説明変数が協同した場合におけるアウトカムへの貢献度を表しているといえる。複数の変数を合計した値(合成変数)の分散は、(合成変数の分散)=(変数 1 の分散)+(変数 2 の分散)+2*(変数 1 と変数 2 の共分散)という関係になる。したがって、変数重要度の順位付けの方法として分散を用いる方法は、合成変数を考えた場合に極めて自然な考えであるといえる。

図 2 に、体重・身長・BMI の 3 変数を体格に関する合成変数としてまとめて SHAP 総合プロットを描画したものを示す。

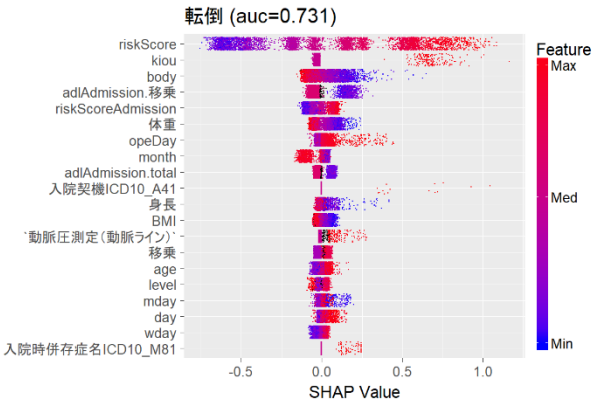

図2　SHAP 総合プロット(体格関連値のグループ化)

3 位の body が 3 変数の合成変数であり、元の変数群の順位(各 5,10,11 位)よりも変数重要度が上昇していることが分かる。body の色は体重の値に基づいて決定しているが、赤点の SHAP 値が低く、青点は大きいことから、低体重ほど転倒リスクが高いことが分かる。

体重単体でも上記の関係は分かるが、body の分散の方がより大きく、関係がより明確となった。

参考文献

1) Scott M Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In Advances in Neural Information Processing Systems 30. Curran Associates, Inc., 4768–4777. http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

2) Friedman JH. Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics Vol. 29, No. 5, pp. 1189-1232, Oct. 2001

3) Friedman JH. Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics Vol. 29, No. 5, pp. 1189-1232, Oct. 2001

4) Yasunobu Nohara, Yoshifuru Wakata and Naoki Nakashima. Interpreting Medical Information Using Ma-chine Learning and Individual Conditional Expectation, Proceedings of the 15th World Congress on Medical and Health Informatics (MedInfo2015), p.1073, Aug. 2015

5) Scott M. Lundberg, Gabriel G. Erion, Su-In Lee. Consistent Individualized Feature Attribution for Tree Ensemble. arXiv,

http://arxiv.org/abs/1802.03888, 2018
6) 野原 康伸, 鴨打 正浩, 中島 直樹, “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”, 第 22 回日本医療情報学会春季学術大会抄録集, pp.78-79, 2018 年 6 月