一般口演 | 医療データ解析

一般口演17
データマイニング

2021年11月20日(土) 16:30 〜 18:00 F会場 (2号館2階224)

[3-F-3-04] 在院日数の二段階予測とその要因分析

*桑原 惇[1]、松田 敦義[2]、荒木 賢二[3]、中田 和秀[1] （1. 東京工業大学 工学院, 2. 株式会社ログビー, 3. 宮崎大学 医学部附属病院）

*Shun Kuwabara[1], Atsuyoshi Matsuda[2], Kenji Araki[3], Kazuhide Nakata[1] （1. 東京工業大学 工学院, 2. 株式会社ログビー, 3. 宮崎大学 医学部附属病院）

キーワード：Length of Stay, Electronic Health Record, Data Mining

現在の日本の医療における問題点として、国民医療費の増加が問題となっている。そのため、国民医療費の増加を抑える為に医療資源の有効活用が求められている。さらに DPCの導入により、入院時の診療報酬が出来高制から包括評価制度に移行し、病院経営の観点からも患者を長く入院させることは避けるべきことになっている。このような背景から、本研究では医療資源の効率的な活用と患者の在院日数短縮の為に、在院日数予測モデルの提案と在院日数が増減する要因分析を行った。この中で特に三つの課題が存在した。一つ目は入院日数のばらつきが大きいことである。データの中には様々な状態の患者が含まれており、同じ傷病の他の患者に比べて極端に在院日数が長い患者も存在する。二つ目は多くの種類の病気を扱う必要があるという点である。これらをカテゴリカルデータのまま使用すると、要因分析が困難になることが予想される。三つ目は予測精度と解釈性の両立の必要性である。一般的に予測精度の高いモデルはブラックボックスになりがちで、逆に解釈性の高いモデルは予測精度が低くなりがちである。それらの課題に対して、本研究では二段階での在院日数予測、 LDAを用いた特徴量変換による予測に対する解釈、 SHAPを用いた分析を組み合わせて行うことで解決した。数値実験では、提案モデルは他の手法で在院日数を予測するよりも、予測値と実際の入院日数の平均絶対誤差において改善が見られた。また、 LDAを用いて傷病の特徴量を変換したことにより予測精度の改善と SHAPを用いた解釈が行いやすくなることが確認できた。 SHAPを用いた入院日数の要因分析では、医療の知見に即した特徴量と SHAP値の関係を確認することに成功した。これにより、医療資源の効率的な活用と、在院日数の長期化を防ぐ為の取り組みが進むことが期待される。

在院日数の二段階予測とその要因分析

桑原惇[*1]、松田敦義[*2]、
荒木賢二[*3]、中田和秀[*1]
*1 東京工業大学 工学院、*2 株式会社ログビー、
*3 宮崎大学 医学部附属病院

Two-step prediction of length of stay and factor analysis

Shun Kuwabara[*1], Atsuyoshi Matsuda[*2]
Kenji Araki[*3], Kazuhide Nakata[*1]
*1 Tokyo Institute of Technology, School of Engineering, *2 Logbii, Inc.,
*3 University of Miyazaki Hospital

With the increase in national health care expenditures and the introduction of DPC, it has become necessary to avoid keeping patients in the hospital longer. With this background, this study proposes a prediction model of hospital length of stay and analyzes factors that increase or decrease hospital length of stay to utilize medical resources and shorten the length of stay of patients, using data from the DPC and data warehouse (DWH) of the University of Miyazaki Hospital. In dealing with this data, there are some problems such as the large variation in the length of hospital stays and the huge number of types of injuries and diseases. To solve this problem, we proposed a framework that can be used for the prediction and factor analysis of hospital stay. In numerical experiments, we confirmed that the proposed model can predict hospital stay more accurately than other methods and that it can perform factor analysis based on medical knowledge and the current situation in the medical field.

Keywords: Length of Stay, Electronic Health Record, Data Mining

1. 諸論

近年、情報技術の発達に伴い医療の現場での電子カルテの導入が進んでいる。そして電子カルテは、その一次利用だけでなく研究、連携医療などの幅広い二次利用が促進されることが期待されている [1]。現在の医療における問題点として、国民医療費の増加があげられる [2]。そのため、この傾向を抑える為に医療資源を有効活用することが求められている。さらに DPC の導入により、入院における診療報酬が出来高制から包括評価制度に移行し、効率的な病院経営の観点からも患者を長く入院させることは避けるべきことになっている [3]。このような背景から、医療資源の効率的な活用と、在院日数の長期化を防ぐことが求められている。

在院日数の予測や、在院日数が長くなる患者の特徴を分析する研究は行われているものの、予測結果の精度と、患者の特徴量と在院日数の関係性を詳しく分析することを両立できているものは私の調べた限り報告されていない。

そこで本研究では LightGBM、LDA、SHAP といった手法を組み合わせることにより、在院日数の予測と、患者の特徴量と在院日数の関係性について分析を行うことの出来るフレームワークの提案を行う。

本稿は以下のように構成されている。2 節では本研究の目的を述べる。3 節では 2 節で述べた目的に関連する研究についてと、提案分析フレームワークに用いた手法の説明を行い、提案分析フレームワークの提案を行う。4 節では提案分析フレームワークの有効性とそれを用いて行った分析の結果について述べる。最後に 5 節で考察と結論を述べる。

2. 目的

医療資源の効率的な活用と患者の在院日数短縮の為に、在院日数予測モデルの提案と在院日数が増減する要因分析が役立つと考えられる。各入院患者の退院する日の予測が可能になることで病床管理が行いやすくなり、医療資源の活用の効率化に繋がることが期待されるからだ。また、在院日数が増減する要因がわかることで、在院日数を短くする為にはどのような治療方針、医療現場の仕組みを作ればいいのかといった改善を行う助けになると考えられる。そこで本研究では在院日数予測モデルを作成し学習・予測を行い、その後に学習済みモデルを使用して予測結果に対する各特徴量の与えた影響を分析することを目的とする。これを実際の電子カルテのデータを利用して行う為に、いくつかの問題が存在する。1 つ目が、在院日数が他の患者に比べて極度に長くなる患者の存在である。これによりモデルが在院日数予測の為の学習を行うことが困難になる。2 つ目が膨大な種類の傷病を扱う必要があることである。電子カルテには膨大な種類の傷病が含まれる為、そのまま特徴量として使用するとデータがスパースになりやすい。これにより、予測精度や解釈性が低下する恐れがある。3 つ目が予測精度と解釈性の両立である。一般的な機械学習モデルでは、予測精度と解釈性はトレードオフの関係にある [4]。本研究ではこれらの課題を克服して、在院日数の予測から要因分析までを行うことが可能なフレームワークの提案を行う。

3. 方法

3.1 関連研究

本研究では在院日数予測と在院日数が増減する要因分析を行うことの出来る分析フレームワークの提案を行う。この問題の関連研究として、DPC データを用いて在院日数予測を行った研究 [5]や循環器疾患を抱える患者に対する在院日数予測と予測時に重要とされた特徴量の分析を行った研究 [6]がある。仲濱らの研究 [5]では、予測が簡単な胃体部癌のみを対象の疾患として在院日数予測を行った。この時の対象データの入院期間の分布は尖度が 4.64、歪度が 1.77 であった。このデータからさらに、入院期間のばらつきが大きい、手術が行われていたデータを抜いたものを用いて実験が行われていた。この研究では、自己組織化マップを用いて入院日数の予測及び在院日数の長い患者の特徴を分析していた。しかし入院期間のばらつきの少ない、予測が簡単なデータに絞っ

たにもかかわらず予測結果の絶対平均誤差は 5.14 日であった。そのため、データを正しく学習できていない可能性があり、要因分析結果の信頼性が低いという問題点がある。また、Daghistani らの研究[6]では情報ゲインに基づいて重要な特徴量選択を行った後に Random Forest 等の手法を用いて在院日数を短期、中期、長期の三クラスに分割して予測を行っていた。この研究では Random Forest を用いて予測を行った場合は精度良く、在院日数クラスの予測が行われていた。一方でこの研究では在院日数予測に重要な特徴量がなにかは示されていたが、それらの特徴量が在院日数を増加させる方向に影響を与えていたのか、減少させる方向に影響を与えていたのかは示されていない。そこで本研究ではこれらの先行研究の良い点を引き継ぎつつ、問題点を改善できるように LightGBM、 LDA、 SHAP といった手法を組み合わせた分析フレームワークの提案を行った。

3.2 本研究で使用する手法

3.2.1 LightGBM

本研究では在院日数の予測のために Light GBM を使用した。LightGBM[7]は kaggle などのデータサイエンスのコンペティションにおいてランキング上位のモデルでも頻繁に使用される手法の一つである。LightGBM は勾配ブースティング回帰木のアルゴリズム(GBDT)であり、XGBoost や pGBRT といったその他の GBDT のアルゴリズムよりも高速に学習を行うことが可能である。GBDT はアンサンブルな手法で、複数の弱学習器を組み合わせることで回帰を行う。同じように複数の弱学習器を組み合わせた手法である Random Forest との違いは、弱学習器の作成の仕方にある。Random Forest はブーツストラップサンプルを用いることで学習に使うデータや特徴量を変えて複数の弱学習木を作成するが、GBDT では前の弱学習木の予測値の誤差を引き継いでそれを小さくするような学習器を作成することで複数の弱学習器を作成する。LightGBM 以前の GBDT を用いたアルゴリズムでは全ての可能な分割点の information gain の推定の為に全てのデータを使用する必要があった。その為、次元数が大きかったり、データサイズが大きいものに対しては計算に時間がかかるという問題点があった。そこで LightGBM では学習に用いるデータ量を減らす為の工夫として GOSS(Gradient-based One-side Sampling)。特徴量を減らすための工夫として EFB(Exclusive Feature Bundling)を用いることにより、従来の GBDT よりも 20 倍以上の高速化を可能にした。

3.2.2 LDA

Latent Dirichlet Allocation(LDA)[8]は文章中の各単語の出現頻度から、単語同士の共起性に注目しそれをモデル化したモデルである。本研究では、傷病間の共起関係に着目して特徴量を抽出するために LDA を使用した。LDA では文章は複数の潜在トピックから生成されることを仮定している。LDA では Dirichlet 分布のパラメータ α によって、各文章の潜在トピック分布が生成される。さらに単語の種類数を V として、単語のインデックス集合を $V = \{1, 2. . . , V\}$ とすると、Dirichlet 分布のパラメータ β によって各潜在トピックにおける単語の出現分布 $\phi_k = (\phi_{k,1}, \dots, \phi_{k,V})$ が生成される。これは V 上の離散確率分布である。また、LDA では文章中の単語がどの潜在トピックによって生成されたのかを示す潜在変数を導入する。具体的に、文章 d の i 番目の単語 $w_{d,i}$ に対応する潜在変数は $z_{d,i}$ と定義される。潜在トピックが K であるとすると、$z_{d,i} \in \{1,2, \dots, K\}$ となる。このように LDA では、潜在変数の値が同じ単語は、同じ単語出現分布に従って生成されるようにモデル化がされている[9]。LDA を利用することにより、潜在変数のもつ隠れた情報の利用が可能となることで、データの意味の解釈に役立たせることも可能である。このように LDA を使用して共起関係から、データの意味を解釈、利用した研究は多くあり、例えばヘアサロンの POS データを使用して、顧客を文章、スタイリストを単語と見立てて LDA を適用し、潜在トピックを計算しその類似度を利用することで、顧客とスタイリストの相性値を計算した例もある[10]。本研究では、患者を文章、傷病を単語として見立てて LDA を適用することで、各傷病がどのトピックに出現するかを表す確率分布

$$\phi_{k,V} = p(k|v) \tag{1}$$

を計算し、これを元に傷病の特徴量を作成する。

3.2.3 SHAP

医療分野では予測の可否だけでなく、薬の影響や、性別、年齢などの特徴を解釈可能な方法で考慮することも重要である[11]。また、解釈が可能となることで医療事象におけるリスク要因や、医学における新しい知見を発見できる可能性もある。線形モデルなどのシンプルな手法であれば、そのモデル自身から解釈を行うことが可能である。しかしそれらのシンプルな手法ではタスクにおいて高い精度を出すことは難しい。一方で複雑なモデルを用いると、そのモデル自身を解釈に用いることは困難で、よりシンプルな解釈のためのモデルが必要となる。本研究でも、患者の在院日数が長くなるリスクとなる要因や、逆に早期に退院が可能となる要因を見つけるために SHapley Additive exPlanation(SHAP) を用いた学習済みモデルの分析を行った。SHAP はゲーム理論において個々のプレイヤーの寄与度を算出する指標であるシャープレイ値をベースにしている。

SHAP を提案した Lundberg 等は、説明の為のよりシンプルなモデルを元モデルの「解釈可能な近似モデル」として以下のように定義した。

$$g(x') = \phi_0 + \sum_{i=1}^{M} \phi_i z_i' \tag{2}$$

ここで x' は LIME[12] で提案されているのと同様に関数 h_x を用いて $x = h_x(x')$ を満たす x の 単純化ベクトルである。局所的な方法では $z' \approx x'$ の時はいつでも $g(x') \approx f(h(z'))$ が成立するように試みる。また、$z \in \{0,1\}^M$、M は単純化した入力の次元数、$\phi_i \in \mathbb{R}$ である。2 式の定義に一致する説明モデルの手法では、ϕ_i を各特徴量に対する「効果」としてみることができ、全ての変数の「効果」の和は $f(x)$ のアウトプットに近似される。LIME[12]や DeepLIFT といった他の解釈モデルも 2 式の定義に一致する。Lundberg 等によって、2 式の定義に従い、必要な性質を満たす説明モデルは 1 つであり、以下の式に従うことが示された[13]。

$$\phi_i(f, x) = \sum_{z' \subseteq x'} \frac{|z'|!\,(M - |z'| - 1)!}{M!} [f_x(z') - f_x(z' \setminus i)] \tag{3}$$

ここで $|z'|$ は $z' \in \{0,1\}^M$ の非ゼロの要素の数、f_x は特徴量 x で訓練されたモデルである。

Lundberg 等によって提案された SHAP 値は、元のモデルの条件付き期待値の関数のシャープレイ値であり、$f_x(z') = f(h_x(z'')) = E[f(z)|z_s]$ とした時の 3 式の解である。ここで、

S は z' の非ゼロ要素のインデックスの集合、z_S は集合 S に含まれていない特徴量の欠損値を持つ。SHAP 値の正確な値の計算は困難であるが、特徴量の独立性と、モデルの線形性を仮定することで近似解を求めることが可能である。これによって各特徴量が予測値に対してどのような影響を与えたのかを確認することが可能となった。

3.3 提案分析フレームワーク

3.3.1 二段階在院日数予測モデル

入院の在院日数は、現在の診療報酬の制度では各傷病、治療法によって定まる診断群分類別に DPC で基準となる在院日数が決められており、それに応じて一日あたりの診療報酬が決められている。DPC では厚生労働省が集計した診断群分類別に入院期間 I~III が定められており、定められた日数を過ぎるごとに入院一日当たりの保険点数が減少する。そのため、平均在院日数を元に設定されている入院期間 II の閾値である第Ⅱ日前後に患者を退院させるインセンティブが働く。本研究で対象とする電子カルテのデータでも在院日数が第Ⅱ日の前後に極端に集まっている。一方で、入院期間が長い患者は極端に長くなってしまうこともあるため、全ての入院に対して精度よく在院日数予測を行うことが困難である。このような問題から提案分析フレームワークにおける予測モデルの基本的な考え方に次の 3 つがある。

1. 極度に長い入院に対しては高い精度での予測は行わない
2. 細かく在院日数の予測を行う対象を絞ることにより精度の向上を目指す
3. SHAP を用いて要因分析を行いやすくする為に決定木ベースのモデルを用いる

1 の「極度に長い入院に対しては高い精度での予測は行わない」に関して、平均的な入院に比べて極度に在院日数の長いものも高精度で予測することは非常に困難である。しかし、10 日間の入院に対して 7 日間の誤差で予測を行うのと、100 日間の入院に対して 7 日間の誤差で予測を行うのでは、同じ 7 日間の誤差であってもその重大さは大きく異なる。このように異常に長い在院日数の入院に対しては高精度に在院日数の予測を行う必要性は少なく、それよりも異常に長くなってしまうかどうかの方が重要である。そのため、本研究では一段階目で入院が異常に長くなるかどうかの予想を行い、ならないと予測されたものに対してのみ細かく在院日数予測を行うモデルの提案を行う。モデルの概略図は図 1 に示した。なお「入院が異常に長くなるかどうか」の基準として、本研究では DPC で定められた各傷病の入院期間Ⅲの閾値であり、平均在院日数+2σ の値をもとに決められた日数である第Ⅲ日を使用した。

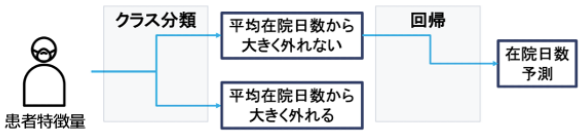

図 1 二段階予測モデルの概略図

3.3.2 分類・回帰に使用する予測モデル

本研究では、モデルの学習・予測の後に SHAP を使用して在院日数が増減する要因を調べる実験を行う。SHAP は木構造のモデルに対して適用がしやすい [14]為、第一段階、第二段階の両方で木構造をベースにしたモデルを使用して比較実験を行い適切なモデルの選択を行った。本研究では LightGBM を利用した。

3.3.3 特徴量に加えた工夫

本研究で使用する電子カルテから抜き出した特徴量は表 1 の通りである。そのデータの性質から、実験を行うにあたって様々な前処理を加える必要がある。

表 1 利用したデータの項目

データ種類	選択特徴量数	項目
DPC 情報	46	入院時年齢、性別、入院経路、転科の有無、計画入院か緊急入院か、入院目的、日常生活動作、身長、体重、BMI、入院の契機となった傷病、併存症、平均在院日数、化学療法等
バイタル情報	6	脈拍数、呼吸数、酸素飽和度、拡張期血圧、収縮期血圧、体温
検査項目	155	血圧、尿の検査結果

一つ目の前処理を行う必要のあるデータの特徴は、欠損値が多いという点である。分析対象として、入院の契機となる傷病を一定程度絞っても、各患者の状態は異なっている。その為、受けた検査は患者によって異なり各特徴量の欠損値は多くなる。その結果、比較的欠損値の数が少ないものが選ばれた表 1 内の項目に対しても欠損値の補完の処理は必要である。補完はカテゴリカルデータに対してはその項目の最頻値を、数値データに対してはその平均値を用いて行った。

二つ目の前処理を行う必要のあるデータの特徴は傷病の種類が多く、傷病に関する特徴量はスパースになりやすいという点である。例えば本研究では入院日数予測は入院の契機となった傷病に ICD10 コードの C00-D48 の新生物<腫瘍>をもつ入院を対象としている為、入院の契機となった傷病の数は限られている。しかし、その各入院において患者がもつ併存症には C00-D48 以外の傷病も含まれている為、併存症の情報をそのまま特徴量として用いる場合は非常に多くの種類のカテゴリカルデータを処理する必要が出てくる。このようなデータの持ち方ではデータがスパースになりやすいという問題点がある。そこで本研究では傷病の情報を LDA を用いてトピック化する工夫も行った。この工夫を行うことにより以下の 3 点の効果が見込まれる。

1. 入院日数予測タスクにおける学習データ以外の傷病の情報も活用して傷病の特徴量を学習出来る
2. 大量の傷病をダミー変数化して保持することによるデータのスパース化を防ぐことが出来る
3. SHAP を用いた要因分析の際にトピック単位での効果の解釈が出来る

その為に、電子カルテに含まれる各患者の診察結果のデータを用いて、患者を文章、傷病を単語と見立てて LDA を適用する。LDA を用いて各傷病 v のトピック分布 $\phi_{k,v} = p(k|v)$ を計算する。LDA において、各文章は複数のトピックから構成されており、そのトピックによって単語が生成されると仮定されている。本研究のように患者を文章、傷病を単語と見立てて適用した場合は、トピックはそのトピックに出現する傷病の原因の要素のようなものであると解釈をすることが出来る。この解釈の下で、逆に傷病がどのようなトピックに分布しているのかを表す $\phi_{k,v} = p(k|v)$ は、傷病の原因となる要素を用いて傷病をベクトル化したものであると解釈をすることが出来る。またこの時、傷病 v のそもそもの出現確率が極端に小さい場合は、各トピックにおける $p(k|v)$ の値も 0 に近くなってしまう。そのため、トピック数を N とした時、以下のように正規化を行うことで、傷病 v のトピックが極端に小さくなることを回避した。

$$\sum_{i}^{N} p(k_i|v) = 1 \quad (4)$$

このように傷病をLDAを用いて変換したのち、併存症のように複数の傷病が有る場合はトピックごとに足し合わせたものをその患者がもつ併存症の特徴量、傷病の特徴量とした。この工夫により、入院の契機となった傷病 376 種類と併存症 1,363 種類をそれぞれ 12 次元に圧縮することが可能となった。更に全ての傷病をまとめてトピックで表現するように変換した場合は、入院の契機となった傷病と併存症を合わせて 12 次元で表現することが可能である。

3.3.4 SHAP による要因分析

DPC の導入により、診療報酬は医療行為ごとに支払いが増える出来高制ではなく、在院日数に応じた 1 日あたり定額報酬を算定する方式に変更されることとなった [3]。これにより、患者の入院が長期になるほど、1 日あたりの診療報酬が低くなるようになった為、患者の健康面、国民医療費の観点からだけでなく、病院経営の観点からも長期の入院は避けたい事象となっている。そこで本研究では患者の在院日数が増減する要因を調べる為に予測結果と各特徴量の関係の分析を行う。

一般的に、機械学習モデルの解釈を行う場合、シンプルなモデルであるほど解釈可能性は高いが予測精度は低い。逆に予測精度の高いモデルは解釈が非常に困難である。このように解釈可能性と予測精度はトレードオフの関係にあるため、本研究では予測モデルと解釈で別々のモデルを使用する。そこで提案する分析フレームワークでは、二段階予測モデルの二段階目の回帰予測部分に対して SHAP を適用することで、在院日数が増減する要因の分析を行う。これにより、在院日数の高精度予測と要因分析の二つを両立することに成功した。

4. 結果

提案フレームワークの予測性能とそれを用いた分析結果を確認する為に、以下の 4 つの実験を行った。

実験1: 二段階に分けて在院日数予測を行うことの必要性の確認
実験2: 第一段階のクラス分類で使用するモデルの決定
実験3: 二段階在院日数予測モデルの精度の確認とLDAによる特徴量変換の効果確認
実験4: 在院日数の増減の要因分析

これらの実験により、提案フレームワークの有効性を確認することに成功した。ここでは実験 1 から 3 の結果と実験 4 の一部抜粋をした結果を説明する。実験では、宮崎大学医学部附属病院の DPC と DWH のデータを用いた。LDA の学習の為に 1916 年 2 月 1 日から 2017 年 7 月 10 日までの DWH 上にある 274,511 人の全 2,358,696 件を、在院日数予測の為に 2013 年 1 月 4 日から 2018 年 3 月 30 日 までに入院し、入院の契機となった傷病に ICD10 コードの C00-D48 の新生物〈腫瘍〉をもつ 10,231 人の患者の入院開始日前後 7 日以内の DPC、DWH データを使用した。在院日数予測タスクでは 2017 年 7 月 10 日までに入院した延べ 8,183 人の患者のデータを用いて学習、残りの 2,048 件を検証データとして使用した。なお、二段階予測モデルの二段階目の部分の回帰予測部分は 2017 年 7 月 10 日までに入院した延べ 8,183 人の患者のデータの内、在院日数が「平均から大きく外れていない」7,369 件のデータを用いて学習を行った。

4.1 実験 1

実験 1 では直接全ての入院に対して在院日数予測を行うことにより、二段階予測を行わない場合はどの程度の予測精度で在院日数予測を行えるのか、そして二段階に分割した場合はどの程度予測精度が向上するのかを確認した。なお評価指標として MAE(平均絶対誤差)を用いた。その為、値が低いほどモデルの精度が高いことを示す。また、各傷病の DPC で定められた平均在院日数である第 II 日を予測値として用いる平均日数予測法と決定木ベースの手法である Random Forest の二種類を比較手法として利用した。

4.1.1 実験 1-1:全て一括で予測することの困難さの確認

本実験では、二段階ではなく直接すべてのデータを用いて学習・予測を行い、その精度を確認した。そのため、在院日数の回帰予測ではあるが、学習データ 8,183 件全てを用いて学習し、検証データ 2,048 件すべてに対して予測を行った。特徴量には LDA を用いて変換したものは使用せずに学習・予測を行った。結果は表 2 に示した。LightGBM を用いた時に一番精度が高く、平均日数予測法に比べて約 12%MAE を改善できることがわかった.

表 2 実験 1.1:全てまとめて在院日数の回帰予測を行った結果

	MAE
平均日数予測法	8.10
LightGBM	7.12
Random Forest	7.69

4.1.2 実験 1-2: 二段階で予測を行う優位性の確認

本実験では外れ値(DPC データの第Ⅲ日よりも長期の在院)を除いた患者に対しての在院日数予測を行い、その結果から外れ値を除くことによりどれだけ在院日数の予測精度を上げることが出来るのかを確認した。学習データには 8,183 件全てを使用したものと、外れ値のものを抜いた 7,369 件のデータを使用したものの二種類で実験を行った。結果は表 3 に示した。この結果から、外れ値を抜いたデータに対して予測を行う際も、LightGBM を用いて予測を行うのが一番精度がよく、平均日数予測法に比べて約 19%MAE が改善されることがわかった。

表 3 実験 1.2:外れ値以外に対して在院日数の回帰予測を行った結果

	MAE 外れ値を含めて学習	MAE 外れ値を抜いて学習
平均日数予測法	6.42	6.42
LightGBM	11.81	5.23
Random Forest	11.27	5.83

4.2 実験 2

実験 2 では第一段階のクラス分類で使用するモデルを決定する為に複数のクラス分類モデルを用いて、分類精度の比較実験を行う。実験を行ったモデルは決定木ベースの手法である Random Forest と LightGBM である。この 2 種類のモデルで比較実験を行い、実験結果は表 4 に記した。この実験では評価指標として AUC(ROC 曲線下面積)を利用した。この指標はクラス分類モデルを評価する際の代表的な評価指標の一つであり、0.5 から 1 の値をとる。ランダムな識別機の場合は 0.5 の値をとり、完全な識別機の場合は 1 をとる。この結果から、LightGBM を用いた方が精度が高く予測を行うことが可能であることがわかった。

表 4　実験 2：第一段階のクラス分類のモデル間の比較結果

	LightGBM	Random Forest
AUC	0.860	0.780

4.3 実験 3

実験 3 では、二段階在院日数予測モデルの精度確認の為に、DPC で定められた在院日数Ⅱの値を用いたベースラインとして平均在院日数予測法と一段階での在院日数予測を用いて、二段階での在院日数予測の比較実験を行った。また、同時に LDA を用いて特徴量をトピック化する効果の確認の為に、他の特徴量は固定し傷病の特徴量だけ複数の持ち方を行なう比較実験も行なった。傷病の特徴量は以下の 5 種類がある。

A：入院の契機となった傷病のカテゴリカルデータ
B：併存症の個数
C：入院の契機となった契機傷病のトピックデータ
D：併存症のトピックデータ
E：傷病全てをまとめたトピックデータ

実験結果は表 4 に示した。評価指標には MAE を使用した。その為、表 5 の結果は値が低いほどモデルの予測性能が高いことを意味する。

表 5　実験３：二段階在院日数予測モデルの精度と LDA による特徴量変換の効果の確認と比較（評価指標：MAE）

	傷病の特徴量					予測方法		
	A	B	C	D	E	平均日数予測法	一段階予測	二段階予測
1	○	○				8.10	7.12	6.77
2	○	○	○	○		8.10	7.11	6.83
3		○	○	○		8.10	7.11	6.83
4		○			○	8.10	7.21	6.84
5	○	○			○	8.10	7.15	6.71

4.4 実験 4

実験 4 では在院日数が増減する要因を分析するために、フレームワークの二段階目の回帰予測モデルに対して SHAP を適用して分析を行った。まず特徴量別に SHAP 値の平均絶対値を確認して、どのような特徴量が在院日数予測に大きな影響を与えているのかを確認した。結果は図 2 に示した。その後、SHAP 値の平均絶対値の大きさが上位であった特徴量について特徴量の値と SHAP 値の関係の分析を行い、医師の方と分析を行った。ここでは一部の結果のみ抜粋して eGFR の検査値と SHAP 値の関係性についての分析を紹介する。結果は図 3 に示した。

5. 考察と結論

実験 1 では平均日数予測法と LightGBM の結果を比較すると、実験 1-1 の直接全体の在院日数を予測する方法では LightGBM は平均日数予測法に比べて約 12%改善されたのに対して、実験 1-2 の学習・検証データの両方から外れ値を抜いた場合は、LightGBM の MAE は平均日数予測法に比べて約 19%予測精度が改善されている。これらの結果から、外れ値を抜いて学習・予測を行う、つまり、二段階で予測を行うことにより、より高い精度で在院日数の予測を行えることが確認された。

実験 2 では一段階目の分類タスクにおいて、Random Forest よりも LightGBM のほうが予測精度が高いことが確認された。そのため、提案分析フレームワークの二段階予測モデルの 1 段階目でも LightGBM を用いることとした。

実験 3 における在院日数の予測実験では、A、B、E を特徴量として利用して学習・予測を行った二段階在院日数予測モデルが一番良い結果であることが確認出来た。そのことから提案フレームワークを用いることでより高い精度で予測が出来ることを示した。

実験 4 の SHAP を用いた要因分析では SHAP 値の平均絶対値が大きい上位 20 個の特徴量の確認と、eGFR についてその検査結果の値と SHAP 値の関係性の確認を行った。SHAP 値の平均絶対値が大きい上位 20 個の特徴量の中には LDA を用いてトピック化した特徴量も現れており、トピック化することにより、傷病単位で関係性の解釈を行うよりも解釈が行いやすくなる可能性を示した。詳細は省略するが、図 2 に現れた Topic5、Topic7、Topic10 について、Topic5 は目に関連する傷病が、Topic7 では循環器系の傷病が、Topic10 には婦人科系の病気が多く出現していた。また、eGFR について、eGFR は低いと腎機能に問題があることを意味する。その為この値が低いと在院日数が高くなるのは妥当であると考えられる。今回は分量の都合上省略したが、その他にも多くの特徴量で医療の知見や、現場でのオペレーションに即した特徴量とその SHAP 値の関係を得られた。予測に用いられた特徴量の SHAP 値を確認することは、モデルの予測結果をどれだけ信頼するかの参考になるだけでなく、それらの関係について、より詳しく分析を行なうことで、新しい医療の知見を獲得出来る可能性がある。また、提案フレームワークを医療現場で使用することが出来るシステムに落とし込むことで、医療資源の効率的な活用と、在院日数の長期化を防ぐだけでなく、医療の質の向上や現場の医療従事者の負担を軽減することに繋がると考えられる。

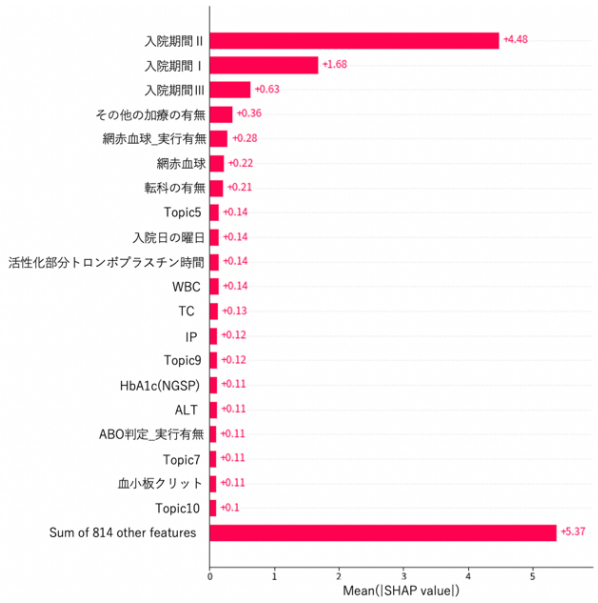

図 2　SHAP 値の平均絶対値が大きい上位 20 個の特徴量とその値

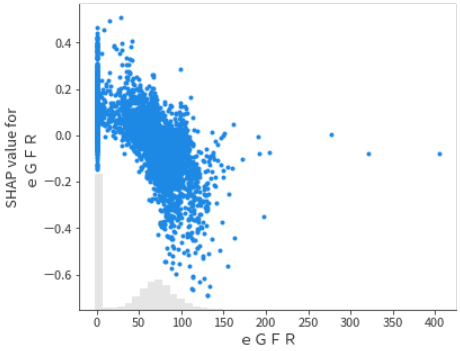

図 3　検査項目 eGFR の値と SHAP 値の関係

参考文献

1) 日本医療ネットワーク協会. 千年カルテとは｜NPO 法人　日本医療ネットワーク協会.　[https://www.gehr.jp/about/index.html (cited 2021-Jan-21)].
2) 厚生労働省. 結果の概要. [https://www.mhlw.go.jp/toukei/saikin/hw/k-iryohi/18/dl/kekka.pdf(cited 2021-Jan-22)].
3) 厚生労働省. DPC 制度の概要と基本的な考え方. [https://www.mhlw.go.jp/stf/shingi/2r9852000000uytu-att/2r9852000000uyyr.pdf(cited 2021-Jan-21)].
4) Trevor Hastie, Robert Tibshirani, and Jerome Friedman.　統計的学習の基礎: データマイニング・推論・予測.共立出版, 2014.
5) 仲濱正大, 納富一宏, 斎藤恵一, 外山比南子.DPC データを用いた在院日数予測を行う Web アプリケーションの開発.マルチメディア, 分散協調とモバイルシンポジウム 2014 論文集, 2014:703-706, 2014.
6) Tahani A Daghistani, Radwa Elshawi, Sherif Sakr, Amjad M Ahmed, Abdullah Al-Thwayee and Mouaz H Al-Mallah. Predictors of In-hospital Length of Stay Among Cardiac Patients:A Machine Learning Approach. International Journal of Cardiology, 288:140-147, 2019.
7) Guolin Ke, Qi Meng, Thomas Finley,et al. Lightgbm: A Highly Efficient Gradient Boosting Decision Tree. Advances in Neural Information Processing Systems, 30:3146-3154, 2017.
8) David M Blei, Andrew Y Ng and Michael I Jordan. Latent Dirichlet Allocation. Journal of Machine Learning Research, 3(1):993-1022, 2003.
9) 佐藤一誠, 奥村学. トピックモデルによる統計的潜在意味解析. コロナ社, 2015.
10) 高正妍, 田澤浩二, チョウイら. Ensemble LDA を用いた既存および新規顧客へのスタイリスト推薦.オペレーションズ・リサーチ, 64(2):95-101, 2019.
11) Christoph Molnar. Interpretable Machine Learning. [https://hacarus.github.io/interpretable- ml-book-ja/, 2020. (cited 2021-Jan-14)].
12) Saumitra Mishra, Bob L Sturm and Simon Dixon. Local Interpretable Model-Agnostic Explanations for Music Content Analysis. International Society for Music Information Retrieval, 537-543, 2017.
13) Scott M Lundberg and SuIn Lee. A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems 30, 4765-4774, 2017.
14) Scott M. Lundberg, Gabriel Erion, Hugh Chen, et al. From Local Explanations to Global Understanding with Explainable AI for Trees. Nature Machine Intelligence, 2(1):2522-5839, 2020.

補遺

本文中に掲載しきれなかった LDA を用いて計算された各トピックにおいて出現頻度の高い上位 20 個の傷病をワードクラウドを用いて表した図を以下の図 4 に示す。

図 4　各トピックの出現頻度の高い傷病をワードクラウドを使用して表した図.出現頻度の高い傷病ほど大きい.