

一般口演 | 医療データ解析

## 一般口演1

### 医療データ解析

2019年11月22日(金) 09:00 ~ 11:00 C会場 (国際会議場 2階国際会議室)

## [2-C-1-06] 医業未収金モデルの高度化と概念実証: スパース正則化とブートストラップ法の応用

○奥村 拓史<sup>1</sup>、高橋 知樹<sup>1</sup>、福田 健<sup>1</sup>、加塩 大吾<sup>2</sup>、辻村 浩子<sup>2</sup>、相馬 陽一<sup>2</sup> (1. 株式会社三菱総合研究所, 2. 東京慈恵会医科大学附属柏病院)

キーワード : patient of unpaid medical bills , lasso, Bootstrap Method

【1. はじめに】 医業未収金問題については、未収金発生後の回収策への対策が一般的であるが、奥村・他（2017）では、AIを用いた医業未収金予防についての新たな方策を提案した。本研究は奥村・他（2017）で提案されたモデルについて統計的な高度化を行う。次節でデータとモデルを示し、3節で推定結果とモデルの評価を行う。4節で考察を行い、最終節で今後の予定について述べる。

【2. 方法】 データは入院患者の属性データおよびDPCデータである。前回のデータに平成28年4月から平成29年3月までの1年分のデータを追加した。また、モデルの判別力を高めるため、WOE（Weight of Evidence）による変数加工を行った。

モデル選択にはlasso（least absolute shrinkage and selection operator）によるスパース推定法を採用した。モデルの精度検証は、AUC（Area Under the Curve）分布をBootstrap法により構成し誤差評価を行った。

【3. 結果】 モデル変数は29のダミー変数、10項目が選ばれた。AUCは0.8であり、高い判別力を有するモデルである。Bootstrap法のパーセンタイル法によるAUCの95%信頼区間は（0.78,0.85）となった。

【4. 考察】 本分析では未収金モデルの統計的な高度化を行った。結果、高い判別力と精度を持つモデルを構築することができた。これにより、より効率的で効果的な未収金予防対策が可能となる。

【5. 今後について】 本分析結果に基づくスコアリングモデルを構築し、未収金予防効果についてのPoCを東京慈恵会医科大学附属柏病院で行う。期間は2か月間。PoCの結果については、11月の学会の場で報告したい。

# 医業未収金モデルの高度化と概念実証 - スパース正則化とブートストラップ法の応用 -

奥村 拓史<sup>\*1</sup>、前間 孝久<sup>\*1</sup>、福田 健<sup>\*1</sup>

相馬 陽一<sup>\*2</sup>、辻村 浩子<sup>\*2</sup>、加塩 大吾<sup>\*2</sup>

<sup>\*1</sup> 株式会社三菱総合研究所、<sup>\*2</sup> 東京慈恵会医科大学附属柏病院

## Refinement and Proof of Concept of Medical Accounts Receivable Management Model: Application of Sparse Regularization and the Bootstrap Method

Hiroshi Okumura<sup>\*1</sup>, Takahisa Maema<sup>\*1</sup>, Ken Fukuda<sup>\*1</sup>

Youichi Souma<sup>\*2</sup>, Hiroko Tsujimura<sup>\*2</sup>, Daigo Kashio<sup>\*2</sup>

<sup>\*1</sup> Mitsubishi Research Institute, Inc., <sup>\*2</sup> The Jikei University Kashiwa Hospital

The strain on financial resources for the social security system with the arrival of a super aging society has caused the environment surrounding hospital management to become increasingly stricter. The issue of unpaid medical bills, as well as problems of hospital management are widely accepted as current social problems. Okumura et al. (2017) proposed a credit scoring model for estimating the probability of default using artificial intelligence (AI). In addition to revealing the characteristics of patients who are likely to default, Okumura et al. demonstrated the possibility of reducing accounts receivables by 50% by taking actions corresponding to the probability of default.

In this paper, we refined the statistical model proposed by Okumura et al. (2017). Specifically, we performed parameter estimation via sparsity regularization (lasso) and model estimation using the bootstrap method. As a result, the model's discriminatory power improved, and its robustness was also verified.

**Keywords:** patient of unpaid medical bills, lasso, Bootstrap Method

### 1. 緒論

わが国は少子高齢化の急速な進展により、世界でも類を見ない超高齢化社会を迎えている<sup>1</sup>。高齢化は医療サービスの需要を増大させる一方で、少子化は社会保障制度を支える現役世代の減少から、社会保障財源を逼迫させる要因となる。更に、近年の経済の低成長化も相俟って、わが国の社会保障システムは危機的状況にある。

政府は、診療報酬の引き下げなど、医療費削減のための政策を強力に推進している。それに伴い、病院経営を取り巻く環境は一層の厳しさを増し、一般社団法人全国公私病院連盟、一般社団法人日本病院会(2016)<sup>1)</sup>のレポートでは、調査対象の病院のうち72.9%が赤字だとしている<sup>2</sup>。医業経営が混迷するなか、未収金に再び注目が集まっている。

医業未収金については2005年、四病院団体協議会が大規模な調査を行った。そこでは、当該病院に加盟する3,270病院の累未収金額が1年間で約219億円もあり、3年間で約426億円に上がることが明らかにされた。厚生労働省は2007年、「医療機関の未収金問題に関する検討会(以下、「検討会」)を立ち上げ、未収金の分析と対応策について記した報告書をまとめた(厚生労働省(2008)<sup>2)</sup>。

これを受けて、2009年に四病院団体協議会が未収金発生防止マニュアルおよび回収マニュアルを作成し、全国約5,700か所の会員病院へ配布した。

これまで、未収金問題は個々の病院の経営上の課題として扱われ、病院が単独でその問題に取り組んできたが、ようやく、社会的問題として広く認識されるようになったといえる。

厚生労働省は2017年に「病院経営管理指標」にかかる調査の一テーマとして、10年ぶりにアンケートによる大規模な医業未収金の実体調査を行った(厚生労働省(2017)<sup>3)</sup>。ここでは、未収金問題は未解決のまま、今なお残り続け、さらにインバウンドの増加に伴い、外国人の医療費未払いも増え、新たな課題となっていることが明らかにされている。

奥村・他(2017)<sup>4)</sup>は「未然の予防策」として、AI(Artificial Intelligence, 人工知能)を使った新たな解決策の提案を行った。そこでは、AIにより患者の未収のしやすさ(確率)を求め、未収可能性の高い患者へ集中的なアクションを行うことで、約50%の未収金を減らすことが期待できるとした。

### 2. 目的

奥村・他(2017)ではさまざまなAI手法を比較検証した結果、データの特徴から最終的にはロジック・モデルを選択した。また、未収患者データのサンプル制約から、モデルの頑健性評価は行われていない。

本稿では、奥村・他(2017)で提案されたモデルの統計的な高度化を行う。併せて、モデルの頑健性の評価も行う。モデルの高度化は未収患者の判別力やモデルの精度を高め

<sup>1</sup> 一般的に65歳以上の人口の総人口に占める割合が21%を超える社会。

<sup>2</sup> 病院別にみると赤字割合はそれぞれ、自治体病院は

89.0%、その他公的病院は60.3%、私的病院は45.9%となっている。

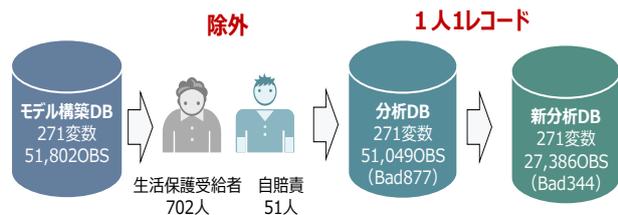


図1 分析データの概要

る。また、頑健なモデルは、安定的なパフォーマンスを期待できる。高い判別力と頑健性を持つモデルを構築できれば、より効率的で効果的な予防対策が可能となる。

分析の結果、未収患者を特徴づける新たな変数が選択され、高い判別力も持つモデルが構築された。また、モデルは頑健で安定的なパフォーマンスが期待できることがわかった。なお、本分析は奥村・他(2017)と同様、東京慈恵医科大附属柏病院と共同研究の成果である。

本稿は以下のような構成で進められる。次節では、分析に用いたデータおよび分析定義と分析方法についての説明を行う。4節では推定結果とモデルの評価を行う。続く第5節では考察を行い期待される効果を導き出す。6節では今後行う予定の概念実証の概要を示し、最後の7節で結論と今後の課題および更なる高度化の方向性について述べる。

### 3. 方法

#### 3.1 データ

データは東京慈恵医科大附属柏病院の入院患者の属性データおよびDPCデータである<sup>3</sup>。前回のデータに平成28年4月から平成29年3月までの1年分のデータを追加した。分析対象患者数は27,386人となり、その内訳は未収患者数344人、非未収患者数27,042人のデータである。分析データは行に患者を来院日単位で持ち、列に患者の属性情報と診療情報の計87変数をもつ、パネル・データである。

#### 3.2 未収患者の定義

未収患者の判定には、退院翌月の支払い状態を示す「状態理由」変数から、病院側の都合を除いた、患者事由による状態理由を未収と定義した。図2に未収患者の定義を示す。なお、被説明変数となる未収フラグは、未収=“1”、非未収=“0”のダミー変数である。

患者の中には、複数回来院している者も数多く存在する。来院時の経済環境や患者の収入状況によって、未収と非未収が混在する患者が複数人認められた。本分析においては、過去、一度でも未収の履歴がある場合は、未収患者と定義した。

#### 3.3 変数加工:WOE

モデルの判別力を高めるため、変数加工を行った。具体的にはすべての説明変数をカテゴリ化し、WOE (Weight of Evidence) 変換を行い、再カテゴリ化した上で、ダミー変数化(以下、モデル変数)した。結果、モデル構築を行う変数は696変数となった<sup>4</sup>。

<sup>3</sup> DPC (Diagnosis Procedure Combination; 診断群分類) に基づく患者の臨床情報と診療行為のデータ。

<sup>4</sup> WOEとは1950年統計学者ジョングッドによって提案さ

#### 1 月末請求翌月支払いのない人

#### 2 “状態理由”が以下に該当する人

- 「後日会計」、「死亡退院」、「所持金なし (その他)」
- 「所持金なし (高額貸付)」、「所持金なし (保険)」
- 「督促対象」、「未会計帰宅」

#### 3 複数回来院患者については過去一度でも未収があれば未収者

図2 未収患者の定義

### 3.4 モデル:lasso

WOEによるダミー変数化を行った結果、696の説明変数を持つ高次元データ・セットになった。WOE変換による再カテゴリ化は被説明変数との線形性を高め、効きが強いカテゴリ変数を抽出できる一方で、高次元で高い相関を持つデータ・セットなることから、分析には注意を要する。一般的に、このようなデータ・セットに最小二乗法による回帰分析や最尤法によるロジット・モデルを適用すると、推定量が得られない、「解の不定問題」に直面することとなる。本分析においても、ロジット・モデルではパラメータの推定ができなかった。

上記の理由により、本分析のモデル選択には疎なモデル(変数選択されたモデル)とパラメータの推定を同時に行うことができ、かつ、モデルの過適合を防止することのできるlasso (least absolute shrinkage and selection operator, スパース推定)を採用した。

lassoは次式の制約付き最小化問題を解くことで、推定値を得ることができる。

$$\min_{\beta} \frac{1}{2n} \|y - \beta\|_2^2 \text{ subject to } \|\beta\|_1 \leq s$$

lasso推定は、いくつかのパラメータの推定値を“0”に縮小する性質を持つ。lassoにより次元の圧縮とパラメータの推定が可能となる。

### 3.5 モデルの評価:Bootstrap法

モデルの頑健性評価は、Bootstrap法にて行う。具体的には、分析データに対しリサンプリングを行い、判別力指標であるAUC (Area Under the Curve)の分布を構成し、予測精度や安定性を評価する。

モデルの頑健性評価は、分析データを推定用データベースと検証用データベースに分け、予測精度を比較するクロスバリデーション法 (Cross Validation Method) が一般的である。実際的には、推定に必要なデフォルトデータが十分に確保されていることが、適用の条件となる。本分析における未収患者数は344人であり、分割した場合、推定に十分なデータが確保できない。また、クロスバリデーション法は、単に2つのデータの同質性を評価していること過ぎず、予測精度の評価として十分でない。

Bootstrap法によるAUC分布を構成し、平均値や分散を推定すれば、モデルの予測精度や安定性を精緻に評価することができる。また、信頼区間を構成すれば、AUCの下限值から、未収金回収額の最小値の計算が可能となる。

れた方法である。

表 2 lasso によるパラメータ推定結果

ID	説明変数	モデル変数	推定値	標準誤差	
1		切片	7.004 ***	0.33	
2	性別	男性	-0.436 ***	0.12	
3	年齢	0<=age <34	-0.599 ***	0.21	
4		35<=age <49	-0.308 *	0.22	
5		50<=age <69	-0.603 ***	0.17	
6		70<=age <80	0.183	0.18	
7	診療科	救急、泌尿、腫内	-1.230 ***	0.26	
8		腎内、呼内、糖内、循内、脳外	-0.897 ***	0.27	
9		外科、産婦	-0.938 ***	0.24	
10		消内、皮膚、整形、形成、小児	-0.433 **	0.25	
11	保険種別	自非、自課	-3.191 ***	0.31	
12		国保、政管、警察、地公、後期	-1.124 ***	0.18	
13		産自、労災、共済	-0.562 *	0.33	
14		545, CCU	-1.148 **	0.46	
15	病棟	07C, ICU, 01C	-0.198	0.25	
16		06C, 07A, 03C, 04A, 07B	-0.323	0.20	
17		05B, 04C, 05A	-0.207	0.19	
18		救急入院	-0.066	0.18	
19	入院形態	救急医療入院、予定外	-0.362 **	0.14	
20	入院経路	介護施設・福祉施設、その他、他の病院・診療所からの転院	-0.436 **	0.18	
判別力指標				AUC	0.710
				的中率	0.768

推定値の\*\*\*は 1%、\*\*は 5%、\*は 10%有意を表す。  
推定されたパラメータのみ記載している。

## 4. 結果

### 4.1 推定結果

表 2 に lasso によるパラメータの推定結果を示す。

未収患者の特徴量として、「性別」、「年齢」、「診療科」、「保険種別」、「病棟」、「救急入院フラグ」、「入院形態」、「入院経路」の 8 変数、19 のモデル変数のパラメータが推定され、残りのモデル変数のパラメータは”0”に縮小された。

パラメータの符号条件は全て満たしている。符号がプラスの変数は“70<=age <80”の「年齢」のモデル変数のみで、その他の変数は全てマイナスとなった。19 のモデル変数のうち 14 のモデル変数のパラメータは統計的に有意となった。

未収に最も影響を与えている変数は、「保険種別」である。“自非”、“自課”が最も悪い。次いで「診療科」の効きが強い。“救急搬送”、“泌尿器科”、“腫瘍内科”が未収リスクの高い診療科となっている。「年齢」では定年前後の中高年層(50<=年齢<69)と若年層(0<=年齢<34)のリスクが高い。

いずれも、直感的にも違和感がなく、納得性のある特徴量が選択されている。

判別力指標の AUC、的中率はともに高く、高い判別力を有するモデルといえる。

### 4.2 モデルの評価

表 3 に 1000 回の Bootstrap 標本の統計量を、図 3 にその標本から構成した AUC の経験分布とパーセントイル法による信頼区間を示す。AUC 分布の中央値は 0.71 であり、推定データの AUC とほぼ一致している。また、AUC の 95%信頼区間は[0.699,0.721]であり、信頼下限でみても 0.699 の高い判別力を有することがわかる。以上から、本モデルの予測精度は安定的であると評価でき、頑健性をもつモデルと言える。

## 5. 考察

本稿の目的は、奥村・他(2017)で提案された、未収患者を判別する AI を使ったスコアリング・モデルの統計的高度化で

表 3 Bootstrap 統計量

Bootstrap統計量				
平均値	中央値	標準誤差	95%LB	95%UB
0.710	0.710	0.0002	0.699	0.721

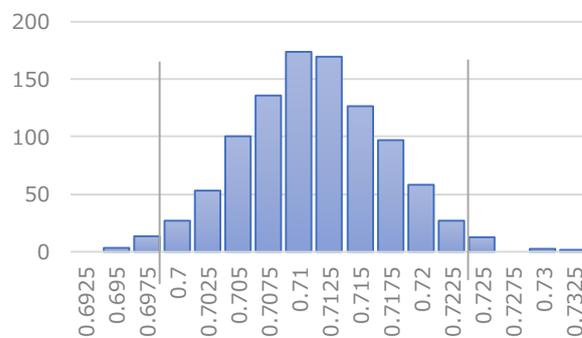


図 3 AUC 分布と信頼区間

ある。統計的な高度化は、未収患者の判別力向上や予測精度の安定性につながり、結果、高い未収金の予防につながる。

奥村・他(2017)は AI スコアリング・モデルによる未収金の予防効果を以下の式で定義し算出している。

$$\text{期待効果} = \text{未収対策患者割合} \times \text{期待予防率} (1 - \text{毀損率})$$

ここで、未収対策患者割合は、推定したパラメータを点数化し上での線形和として計算される。パラメータの推定値は期待値であり、ある一定の幅を持って上下にばらつく。上へのブレは期待以上の未収金防止につながる。一方で、下にブレた場合は、誤判別の増加により、期待した未収金削減効果を得られなくなる。

政策や実施導入の際には、将来の不確実性を考慮した費用対効果の検証が不可欠だ。不確実性はシナリオによる感度分析が一般的であるが、恣意性の排除が困難で、厳密性に欠ける。Bootstrap によるパーセントイル法を使えば、AUC 分布の信頼区間の構成が容易に求めることができ、不確実性を確率的に計量することが可能となる。

Bootstrap 法によるモデル評価の導入は、モデルの予測精度や安定性を精緻に評価するだけでなく、将来の不確実性を客観的に表現できる点で優れており、導入する意義は大きい。

## 6. 概念実証

本分析の研究成果については、共同研究先である東京慈恵会医科大学附属柏病院において、以下の日程と目的で概念実証(Proof of Concept, PoC)を行う予定である。

### ● 第 1 回 PoC

期間： 2019 年 10 月 1 日～10 月末日

目的： 精度検証、運用上の課題抽出

### ● 第 2 回 PoC

期間： 2019 年 11 月 1 日～12 月末日

目的： 未収予防効果の検証

実証実験の結果については、学会発表の場で報告したい。

## 7. 結論

本稿では奥村・他(2017)で提案された医療未収金予防のためのAIによるスコアリング・モデルについて、統計的な高度化を行った。判別力の向上やモデルの予測精度の安定性は、未収金の予防効果を高めるだけでなく、将来のリスクを精緻に計量した感度分析を可能にするなど、高度化の意義は大きい。

高度化の方法としては、WOE 変数変換による変数加工を行った。結果、高次元で変数間に高い相関をもつ分析データとなることから、疎なモデルとパラメータの推定を同時に行うことができ、かつ、モデルの過適合を防止することのできるスパース正則化(lasso)モデルを採用し、パラメータの推定を行った。結果、8変数19のモデル変数が推定され、AUCも0.7を超える高い判別力を有するモデルを構築できた。また、モデルの評価はBootstrap法によるモデル評価採用し、AUC分布の信頼区間をパーセンタイル法で構成することにより、AUCのバラツキを検証した。結果、95%信頼区間は[0.699,0.721]であり、安定した精度を持つモデルであることが分かった。

最後に課題と今後の統計的な発展可能性を述べる。高い判別力を持ち、安定した精度を持つモデルを構築できたが、現場で使うにあたり、納得性のある使いやすいモデルであるかは別問題である。実証実験を通じ、現場の声を反映したモデルにチューニングする必要がある。また、発展可能性の観点からは、より判別力を高めるために、機械学習の手法を取り入れ、精度改善を試みることも一案であろう。

## 参考文献

- 1) 社会保障改革の全体像. 厚生労働省, 2014.  
[[http://www.mhlw.go.jp/seisakunitsuite/bunya/hokabunya/shaikaihoshou/dl/260328\\_01.pdf](http://www.mhlw.go.jp/seisakunitsuite/bunya/hokabunya/shaikaihoshou/dl/260328_01.pdf)].
- 2) 平成28年 病院運営実態分析調査の概要(平成28年6月調査). 一般社団法人全国公私病院連盟, 一般社団法人日本病院会, 2017.  
[[https://www.hospital.or.jp/pdf/06\\_20170306\\_01.pdf](https://www.hospital.or.jp/pdf/06_20170306_01.pdf)].
- 3) 医療施設経営安定化推進事業 平成28年度 病院経営管理指標, 厚生労働省, 2017.  
[<https://www.mhlw.go.jp/file/06-Seisakujouhou-10800000-Iseikyoku/0000212668.pdf>].
- 4) 医療未収金についての新たな方策:AIによる未収患者の特徴療の抽出と予防効果について, 第37回日本医療情報連合大会詳細抄録, 2017.