査読論文

詐欺抵抗力判定アプリの推定性能の改善*

小久保 温†

Improved Estimation Performance of a Fraud Resistance Diagnostic Web Application

Atsushi KOKUBO

ABSTRACT

In this paper, I discuss the estimated performances of a fraud-resistance diagnostic web application, "Professor Watanabe's Fraud-Resistance Diagnostic" (hereafter referred to as "the app").

When the user answers a series of questions, the application uses logistic regression, a type of machine learning, to diagnose fraud vulnerabilities. Estimating fraud vulnerability is a type of problem known as "imbalance class" in machine learning. The app's diagnosis had poor estimation performance due to a problem in handling imbalance classes and a small number of positives in the training data. In this paper, I review previous studies and propose a method that contributes to improving estimation performance of the app.

Key Words: *machine learning, imbalance class, fraud vulnerability, web application*
キーワード: 機械学習，不均衡クラス，詐欺脆弱性，Web アプリケーション

1. アプリの開発

1.1 経緯

この論文では，詐欺抵抗力判定アプリの推定性能について論じる．まず，その開発と運用について紹介する．

科学技術振興機構(JST) 社会技術研究開発センター(RISTEX)の「安全な暮らしをつくる新しい公／私空間の構築」研究開発プロジェクトの一つに，「高齢者の詐欺被害を防ぐしなやかな地域連携モデルの研究開発」(研究代表者・渡部諭秋田県立大学教授)が 2017 年度に採択された [1)]．プロジェクトでは，2017 年 10 月～2021 年 3 月の 3 年半の期間に，特殊詐欺被害の減少を目指して詐欺に対する抵抗力を判定する Web アプリ「わたなべ教授のサギ抵抗力しんだ～ん」を開発・運用し，アプリを活用して詐欺被害防止の啓発活動に取り組んだ [2)]．アプリのタイトルは「詐欺抵抗力」となっているが，実際には逆にアプリ内では詐欺脆弱性を機械学習で推定していた．機械学習のモデル(推定手法とパラメータの導出)の開発は渡部・澁谷泰秀青森大学教授・大工泰裕京都府立医科大学特任助教(当時)が担当し，筆者はアプリのプログラムの開発と運用の責任者であった．

* 令和 4 年 10 月 28 日 受付
令和 5 年 1 月 23 日 受理

† 工学部工学科／大学院工学研究科電子電気・情報工学専攻・教授

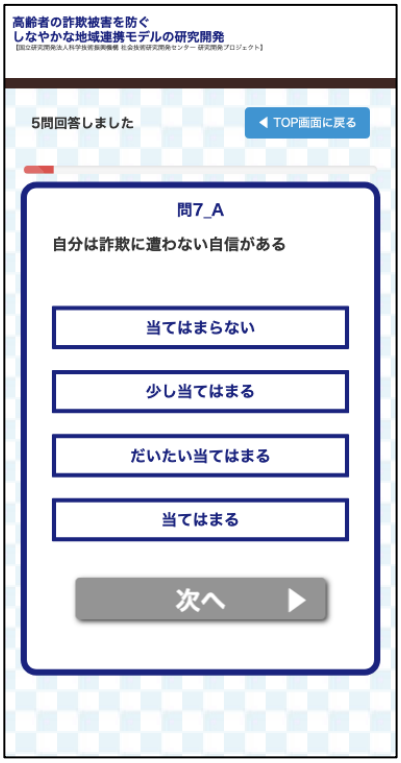

図 1 質問(説明変数)の回答画面

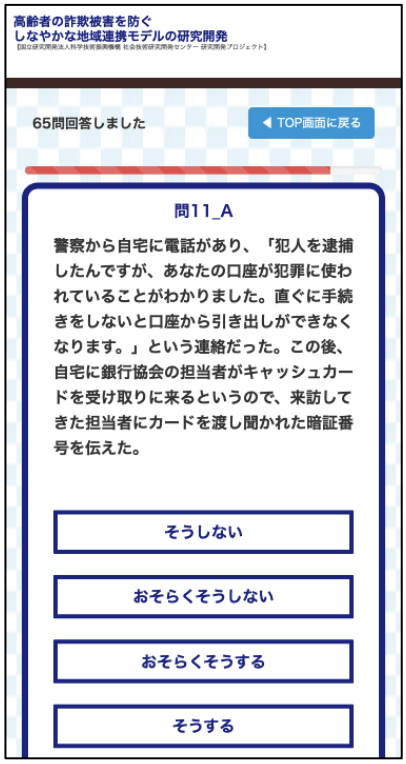

図 2 質問(目的変数)の回答画面

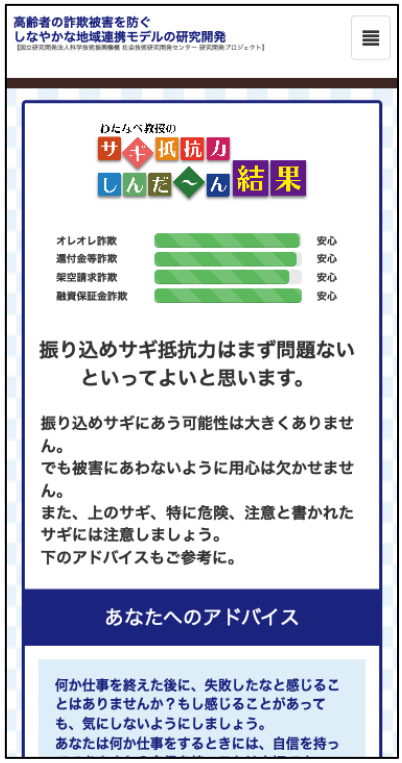

図 3 判定結果の画面

アプリは 2019 年 2 月 21 日～2021 年 3 月 31 日のおよそ 2 年間運用された．判定は数十問の質問から構成されるが，個々の質問(図 1，2．説明変数と目的変数の説明は以降の節)の回答は 860,173 件あり，一連の質問に回答すると表示される判定結果(図 3)は 11,564 回表示された．

1.2 機械学習モデルの開発

このアプリはユーザーが自分で回答する診断アプリの一種である．自答式の場合，ユーザーは質問の意図を汲んで自分の望んだ結果が出るように回答を操作しようとすることがある．これは詐欺脆弱性の推定に影響すると考えられる．この問題を回避するために，詐欺と一見関係がない心理特性などに関する質問をし，その回答から詐欺脆弱性を推定する [3]ことを最終的には目指していた．つまり，詐欺と一見関係ない心理特性などに関する質問の回答が，アプリの判定の予測に用いる説明変数である．一方，アプリの判定が予測しようとする目的変数は，詐欺に遭遇した場面のシナリオを読んでどう行動するかの回答を用いていた．目的変数の質問は意図が汲み取りやすく，ユーザーが操作しやすいので，機械学習のモデルの開発時にだけ使用し，最終的には説明変数のみを質問し目的変数にどう回答するかを予測しようとしていたのである．

また，詐欺被害は高齢者に多く，高齢者にアプリを活用して欲しい．そこで，高齢者が回答しやすいように，質問の個数は少ないことが望ましい．しかし，プロジェクト開始当初，詐欺脆弱性のシグナルとなる質問の候補がたくさんあった．そこでアプリで表示する質問をしぼりこんだ．アプリの初版の開発時にはまだアプリが存在していなかったので，質問紙で社会調査を行なって学習データを収集 [4]し，有効な回答が 690 件得られた．そして，古典的テスト理論と項目反応理論を用いてアプリで表示する質問を絞り込んだ [4]．最終的に説明変数として採用されたのは先行研究 [3]に掲載の 66 問である．これらは「あなたの性別を教えてください．」などの人口統計学的(デモグラフィック)質問 6 問，「自分は詐欺に遭わない自信がある」(図 1)など詐欺場面における行動特性 9 問，「私の人生は，これから楽しくなると思う」など未来展望 10 問，「何か仕事をするときは，自

信をもってできる」など自己効力感 16 問,「普段，自分は健康であると感じる」など生活の質 25 問である．人口統計学的質問以外は,「1 当てはまる」「2 少し当てはまる」「3 だいたい当てはまる」「4 当てはまらない」の 4 択になっている．また，これらの説明変数から予測したい目的変数は，詐欺のシナリオ問題の回答でオレオレ詐欺，架空請求詐欺，還付金等詐欺，融資保証金詐欺の詐欺 4 罪種についてそれぞれ 3 問の計 12 問であった．回答は「1 そうしない」「2 おそらくそうしない」「3 おそらくそうする」「4 そうする」の 4 択であった．詐欺のシナリオ問題は，先行研究 [3]には 3 問のみ掲載されており，12 問全体の概要が先行研究 [5]に掲載されている．質問はたとえば次のようなものであった(図 2).

> 警察から自宅に電話があり,「犯人を逮捕したんですが，あなたの口座が犯罪に使われていることがわかりました．直ぐに手続きをしないと口座から引き出しができなくなります.」という連絡だった．この後，自宅に銀行協会の担当者がキャッシュカードを受け取りに来るというので,来訪してきた担当者にカードを渡し聞かれた暗証番号を伝えた.

そして，すべて回答すると機械学習による推定をもとに判定画面(図 3)が表示される．機械学習にはロジスティック回帰を用いており，詳しくは 3 節で説明する．アプリの初版で採用された質問は説明変数 66 問，目的変数 12 問の計 78 問に絞り込まれているが，それでも回答するのに 20 分近くの時間を必要とした.

2. 本論文および先行研究での陽性・陰性の定義

本論文では，詐欺被害に遭いそうな危険な回答をする人を陽性とする．これは検査に反応する標本が陽性と定義され，本研究では質問に対する回答から詐欺被害に遭う危険性がある人を見分けようとしているからである.

なお，先行研究 [3]では，陽性・陰性の定義について異なったものが同一論文中で混在しているので参照する際には注意されたい．先行研究 [3]では，本研究同様に詐欺被害に遭いそうな危険な回答をする人を陽性とし，機械学習の決定木モデルの一種である FFTree で推定している箇所もある．また逆に，実際に詐欺被害に遭ったことがわかっている人を陽性ではなく陰性として，全回答者から陰性をどれだけ見分けられるかをロジスティック回帰で評価している箇所もある．これは同一論文中で陰性・陽性と安全・危険の組み合わせが逆のものが混在しているということである．陽性・陰性の定義を逆にすると，検査の性能の指標のうち正解率など変わらないものもあるが，実際に陽性のうち陽性と推定された割合である再現率(感度)と実際に陰性のうち陰性と推定された割合である特異度は入れ替わる．また，陽性と推定したうち実際に陽性の割合である適合率(精度,陽性的中率)と陰性と推定したうち実際に陰性の割合である陰性的中率なども入れ替わる．また，アプリの機械学習では「詐欺に遭いそうな危険な回答をする人」を見分けるモデルを構築したが,先行研究 [3]ではこれを用いて全回答者から「実際に詐欺被害に遭った人」を見分けるのに使用している点も問題である．方向性としては似ていて関連はあるが，異なる概念であり妥当な評価方法ではない．評価としては「実際に詐欺被害に遭った人」のうち,「詐欺に遭いそうな危険な回答をすると推定された人」の割合を評価するべきであった.

3. 質問紙調査で収集した目的変数のデータと陽性・陰性への分割

学習データの収集時には，機械学習で推定する値である目的変数に関する質問も含めていた．それは2節で紹介したように詐欺被害に遭いそうになる場面を読んでもらい，どのように行動するかを回答してもらう質問であった [3, 5]．これらの質問は，どのような回答をすると危険と推定されるかが予想しやすく，学習データが十分収集できたら使用しないつもりのものであった．

アプリの初版の開発に使用した質問紙のデータの目的変数の部分の回答数を表 1，割合を図 4 に示した．目的変数の値は詐欺に遭いそうなときに危険な行動を「1 そうしない」「2 おそらくそうしない」「3 おそらくそうする」「4 そうする」の 4 択である．ただし問 11_H だけは，他の質問と逆に「4 そうする」と回答した方が安全な反転項目であった．問 11_H の回答状況を見ると反転項目と気づかずに回答した人が多く，読み飛ばしなどの他の質問と異なる属性を反映していると思われた．そのため，機械学習の推定パラメータの導出に問11_Hのデータは使用されていない(データの収集は行なっている)．質問紙の有効回答 690 件のうち，回答内容 3，4 を危険な陽性だと見なすと最少で問 11_A の 3 件(0.4%)，最多で問 11_E の 71 件(10.3%)である．これは詐欺被害者の数が人口全体に比べて少ないことと対応していると思われる．全 690 件中 3 件の陽性を機械学習で識別しようとしても，たまたまその3件に含まれていた特徴を過学習してしまいかねない．もしも回答内容 2，3，4 を陽性と見なしてよければ，陽性数が増えて最少で問 11_C の 36 件(5.2%)，最多で問 11_E の 210 件(30.4%)になるので，モデルを開発した渡部・澁谷は「2 おそらくそうしない」という謙虚な回答も含んだ2，3，4を危険な陽性と見なしてパラメータを求めている．その場合，推定された結果の分類も「1 そうしない」と「2 おそらくそうしない」の間を境界としたものになる．

表 1 質問紙調査の回答の個数

回答	問 11_A	問 11_B	問 11_C	問 11_D	問 11_E	問 11_F	問 11_G	問 11_H	問 11_I	問 11_J	問 11_K	問 11_L
1 番	648	613	654	651	480	523	582	260	611	609	560	582
2 番	39	57	32	29	139	130	91	44	71	74	118	98
3 番	2	10	2	6	45	32	14	124	8	4	11	10
4 番	1	10	2	4	26	5	3	262	0	3	1	0

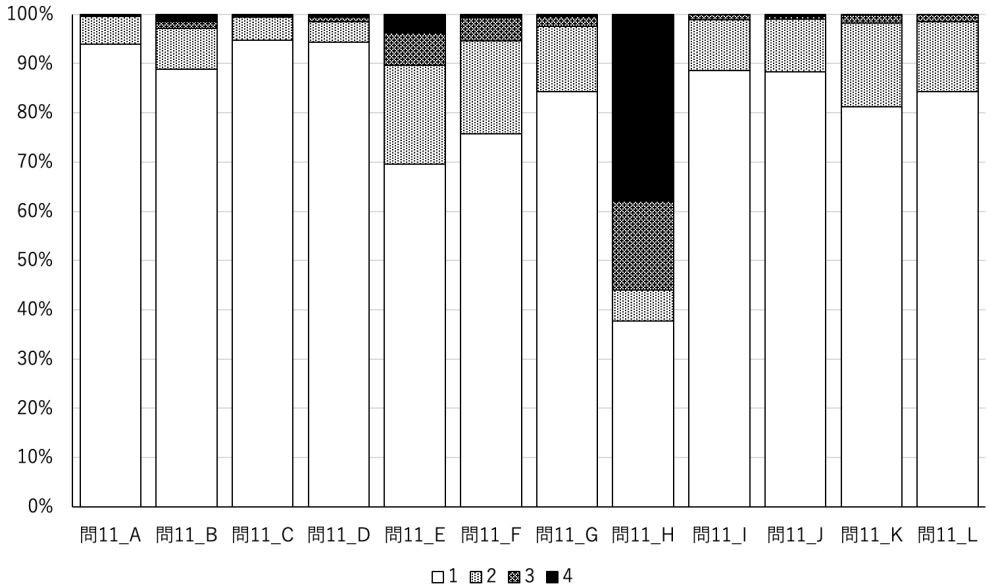

図 4 質問紙調査の回答の個数の割合

4. 質問紙調査で収集した説明変数のデータとロジスティック回帰のパラメータの推定

アプリの詐欺脆弱性の推定には機械学習の線形モデルの一種であるロジスティック回帰を用いていた．指標ごとのロジスティック回帰の推定式は先行研究[6]にすべて掲載されている．たとえばオレオレ詐欺に関する問 11_A の回答を目的変数とし，自己効力感に関する問 9_A～P の回答を説明変数としたときの詐欺脆弱性の確率 P は，以下の式のようになる．この推定式で想定している説明変数の値域は 1～4 である．

$$P_{問11\_A を問9\ で推定} = \frac{1}{1+e^{-\lambda_{問11\_A を問9\ で推定}}}$$

$$\lambda_{問11\_A を問9\ で推定} = -1.763 + 2.191 \times 問9\_D - 1.534 \times 問9\_E - 1.660 \times 問9\_F - 0.990 \times 問9\_J$$

確率 P は変数 λ のロジスティック関数で，これを λ について解くと P のロジット関数 log(P) - log(1 - P)になる．ロジスティック関数は λ の単調増加関数で，S 字カーブを描くのでシグモイド関数とも呼ばれる．λ が 0 のとき確率 P は 1/2 になり，一般にこれを閾値として推定結果を陽性・陰性に分類する．問題に求められる性質に応じて異なる閾値を使うこともできる．λ は説明変数の一次式なので，ロジスティック回帰の分類は，結果として質問ごとにポイントを割り振って，質問の回答とポイントの線形和が 0 以上か否かで判定しているのと同等である．そのとき用いるポイントは機械学習，すなわち人間が経験や直観から決めるのではなく，データを予測し説明するようにフィッティングして決める．ロジスティック回帰ではさらに，分類の決定境界，つまり説明変数の高次元空間での分類の境界の平面，そこからの距離 λ に応じてロジスティック関数の値を分類の予測確率とみなす．アプリでは 1 から予測確率を引いたものを抵抗力として，詐欺の種類ごとに棒グラフで表示(図 3)し，利用者に注意を促すのに使用していた．

アプリの初版のロジスティック回帰の推定パラメータの導出は，先行研究[3]に記載されている．具体的には説明変数を未来展望・自己効力感・生活の質の 3 グループに分け，3 グループ×目的変数 11 個=33 種類の指標ごとに求めている．目的変数は反転項目で異なる種類の特徴を拾っていると思われる問 11_H を除いたので 11 個である．そして，得られた推定式をオレオレ詐欺・架空請求詐欺・還付金等詐欺・融資保証金詐欺の詐欺 4 罪種について，それぞれ推定性能が上位 2 つの計 8 個を採用している．

なお，アプリの初版の開発に用いた質問紙のデータは前節のとおり，詐欺に遭いそうな危険な回答をした人が少なかった．そのため，すべてのデータを用いてそのデータを説明するパラメータを求めている．そして，アプリでは質問数を減らすために古典的テスト理論や項目反応理論を用いて絞り込んでいたが，更に推定パラメータを導出するときにも赤池情報量規準を指標に stepwise 法を用いて，説明変数を絞り込んでいる[3]．機械学習では学習の結果得られたモデルが学習に用いたデータを説明するだけでは不十分で，モデルにとって未知の学習に用いなかったデータも予測できる汎化性能が求められる．というのは，仮に誤差を含んだデータがあり，プロットすると細かく振動するような場合でも，推定式の次数を上げパラメータの個数を増やした複雑な関数形を使用すれば誤差も含めてフィットできる可能性がある．しかし，そのような誤差も含めてデータにフィットさせすぎた激しく振動する複雑な形状の関数は，新たに収集したデータをうまく予測できない．本当に求めたいものは，未知のデータも含めて予測できる汎化能力が高いモデルである．そこで推定に用いる説明変数を絞り込んで，関数の複雑さを抑えようとしていたのである．

5. アプリ初版で収集したデータと先行研究に記載された推定性能

アプリの初版(先行研究 [6]にパラメータが記載)の汎化性能を調べるには，アプリにとって未知の開発後に実環境で収集したデータを使う必要がある．開発したアプリではおよそ2年間で11,564件のデータが収集された．そのうち，機械学習の性能評価やその改善に使えるデータは，最初のおよそ 1 年間にアプリの初版で収集された有効な回答 8,778 件(無効なものも含めると 9,237 件)である．2 年目にリリースされた第 2 版以降では，一般向けに公開しているバージョンのアプリでは質問の個数を減らすために目的変数に相当する質問を削減しているため汎化性能の評価に用いることができない．また，2 年目から新型コロナ感染症の流行により，プロジェクトのイベントなどの活動は困難になり研究用のデータの収集も難しくなった．

アプリの初版の性能はその一部が先行研究 [3]の表 1 に掲載されている．この先行研究では，論文の執筆時点のアプリの全回答 8,690 件のうち，A 警察の協力により実際に詐欺被害に遭ったことがはっきりしている人の回答 25 件を識別できたか否かが混同行列として掲載されている．なおこの評価では，安全な場合が陽性，危険な場合が陰性で本研究とは逆である．そして，実際に詐欺被害者である陰性 25 人のうちアプリが陰性と推定したのはほぼ 0 人で全く推定できなかった．推定指標は全部で 8 種類あるが，6 種類が 0 人で 2 種類が 1 人ずつ推定されたと記載されている．ただし，これはアプリの評価としては適切でない．なぜなら，アプリが推定するのは回答者のうち詐欺に遭いそうな危険な回答をする人であり，全回答者のうち A 警察の協力で回答した実際の詐欺被害者を推定するものではないからである．実際の詐欺被害者のうち，何人が危険と推定されたかを評価するべきであった．

6. アプリ初版の推定性能

前節で，先行研究 [3] でアプリの初版は全回答者から詐欺被害者 25 人を 8 種類の指標で推定しようとしたが，6 種類では 0 人，2 種類では 1 人ずつしか推定できなかったこと，そしてこれはアプリが学習したものと似ているが異なるものを推定していることを紹介した．アプリの初版の推定性能はこれまで適切に検証されたことがない．そこで，本論文ではじめて検証することにした．

アプリは，詐欺に遭いそうな危険な回答をする人を推定するものである．アプリ初版の汎化性能を調べるために，アプリの初版の学習には用いていない，アプリの初版で収集されたデータ 8,778 件を用いた．そして，収集したデータにアプリの初版のパラメータを用いて推定した目的変数の値と，収集したデータに含まれていた目的変数の実際の値を比較することにした．比較に当たって，質問紙データの目的変数の表 1，図 3 に相当するものを，アプリのデータで表 2，図 4 として作成した．質問紙とアプリでは，概ね同じ傾向が見られることがわかった．

次にアプリの初版で用いられた先行研究 [6]のパラメータを用いて，アプリで収集したデータが推定できるかを計算し表3に示した．アプリの初版のパラメータは質問紙のデータから算出されたもので，アプリで収集したデータを算出には利用していないので，これは汎化性能を評価したことになる．表3に記載の通り，詐欺は4種類ありそれぞれ指標が2つずつある．指標にはいずれかの目的変数が対応している．TP，FN，FP，TN は，真陽性数，偽陰性数，偽陽性数，真陰性数である．表 3 の ROC は受信者操作特性(ROC)の下の面積，表の PR は横軸 Precision(適合率)，縦軸 Recall(再現率)をプロットした線の下の面積である．8 つの指標の平均を見ると，実際の陽性は TP+FN で 1,126 件あり，そのうち陽性と推定されたのは 9 件で大変少ないことがわかる．このデータでは陰性が多数で陽性が少数のため，正解率を見ると 86.9%あるが，実際に陽性のものを陽性と

表 2 アプリの初版で収集された回答の個数

回答	問 11_A	問 11_B	問 11_C	問 11_D	問 11_E	問 11_F	問 11_G	問 11_H	問 11_I	問 11_J	問 11_K	問 11_L
1 番	7586	7766	8026	7851	7336	7263	7472	3559	7872	7674	7744	7604
2 番	992	761	585	704	942	1130	961	729	716	857	826	925
3 番	144	176	110	159	359	315	261	1617	124	178	150	191
4 番	56	75	57	64	141	70	84	2873	66	69	58	58

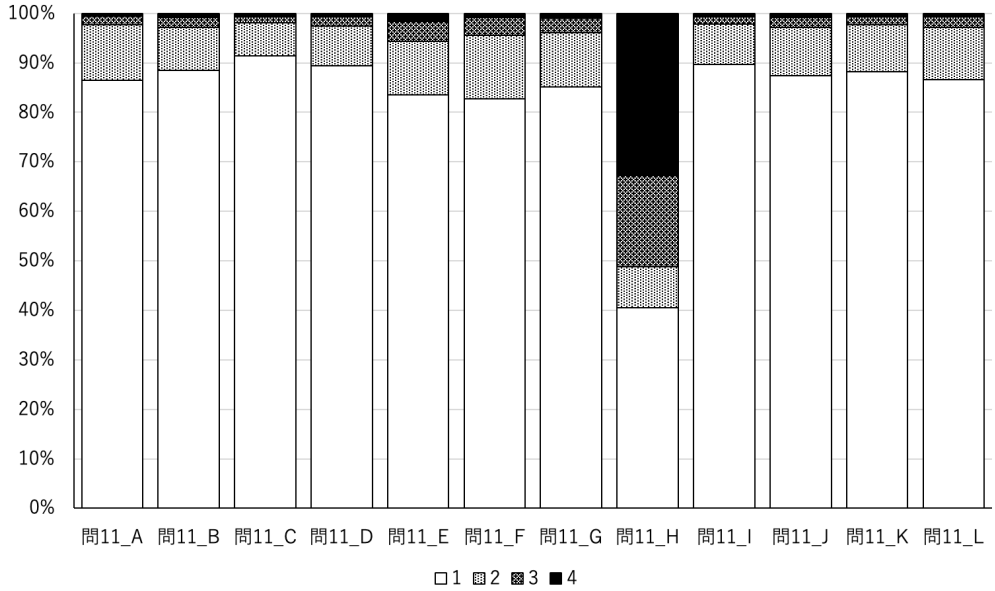

図 4 アプリの初版で収集された回答の個数の割合

推定できた割合である再現率は 0.8%しかなく，99.2%の陽性を陰性と誤分類している．陽性・陰性の分類の閾値を確率 P=1/2 から変更すると適合率と再現率は変動する．そこで閾値を変更しても変動しない ROC 下面積で見ると 53.3%で，これはランダムに推定する場合の 50%に近い(100%正解する場合は 100%になる)．また，不均衡クラスの場合の性能の傾向を示す PR 曲線下の面積は 14.4%であった．なお，PR 曲線下の面積は ROC 下面積のようにランダムが 50%といったわかりやすい解釈はできない．図 5 に ROC 下面積が最大の指標 p05 の ROC を示した．横軸に偽陽性率 fpr，縦軸に真陽性率 tpr を取ってプロットしたのが ROC である．この結果ではほぼ直線で ROC 下面積が 50%に近いことがわかる．図の赤バツは偽陽性率と真陽性率の平均の位置である．平均の赤バツが左下にあり，ほとんど陰性と推定していたことを示している．

先行研究 [3]では全回答者から詐欺被害者 25 人を推定しようとして 0 か 1 人しか推定できなかった．これはアプリの初版の学習したものとは異なるものを推定していたが，本来学習したものと全く関連がないわけでもない．アプリの初版では，再現率が低く実際には陽性の 99.2%を陰性と誤認識していたことが，詐欺被害者の推定うまくいかなかったことと関連していると思われる．

表 3 質問紙データで学習させたアプリ初版のロジスティック回帰の推定性能

詐欺	指標	目的変数	TP	FN	FP	TN	正解率	適合率	再現率	f1 値	ROC	PR
オレオレ詐欺	p02	問 11_A	3	1189	9	7577	86.4%	25.0%	0.3%	0.5%	56.1%	16.0%
	p29	問 11_K	12	1022	44	7700	87.9%	21.4%	1.2%	2.2%	55.8%	14.0%
架空請求詐欺	p05	問 11_B	1	1011	7	7759	88.4%	12.5%	0.1%	0.2%	58.2%	14.6%
	p06	問 11_B	6	1006	26	7740	88.2%	18.8%	0.6%	1.1%	54.5%	13.5%
還付金等詐欺	p18	問 11_G	10	1296	19	7453	85.0%	34.5%	0.8%	1.5%	50.8%	15.6%
	p26	問 11_J	12	1092	48	7626	87.0%	20.0%	1.1%	2.1%	50.3%	13.6%
融資保証金詐欺	p31	問 11_L	22	1152	43	7561	86.4%	33.8%	1.9%	3.6%	53.1%	15.1%
	p32	問 11_L	6	1168	48	7556	86.1%	11.1%	0.5%	1.0%	47.8%	12.8%
平均			9	1117	30.5	7621.5	86.9%	22.1%	0.8%	1.5%	53.3%	14.4%

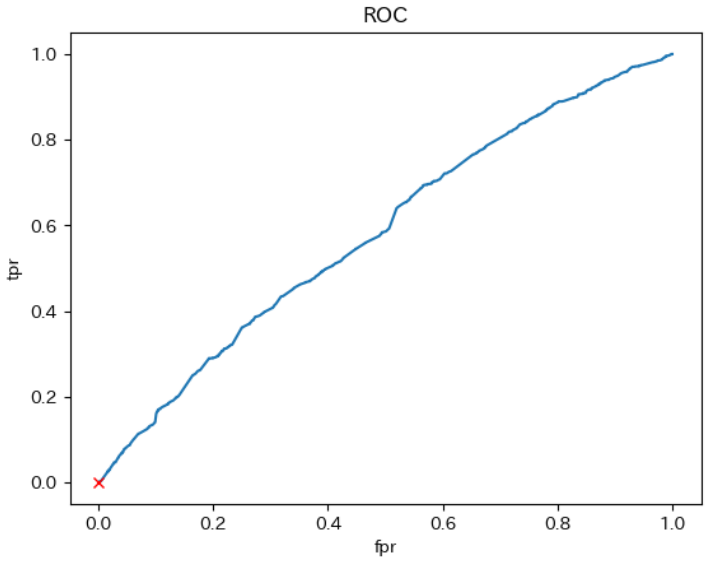

図 5 アプリ初版の ROC 下面積が最大であった指標 p05 の ROC

7. 不均衡クラスの取り扱いの改善

先行研究 [3, 6]のアプリ初版のロジスティック回帰のパラメータの導出時には，データが不均衡クラスであることは考慮されていない．不均衡クラスであることを考慮しないと，陽性・陰性のうち数の多い方であると常に推定するだけで正解率が向上してしまう．この問題に対応するために，パラメータの算出の際に陽性・陰性の個数の逆数の重みをかけて，多寡の逆数に応じた重みで扱われるようにすることを提案する．先行研究 [6]のプログラムは公開されていないため，推定に用いる説明変数の選択は先行研究 [6]で成功していると仮定し，不均衡クラスの重みを考慮し，アプリの初版と同様に質問紙で得られた回答をすべて使用し，それを説明するように学習させた．そして学習には利用していないアプリで収集したデータを用いて汎化性能を計算した結果を表 4 に示す．計算には scikit-learn 1.1.2 の LogisticRegression を用い，不均衡クラスを考慮するために class_weight オプションに balanced を指定した．そして，不均衡クラスの場合の性能の傾向である PR 曲線下の面積を最大化するパラメータを求めた．すると，再現率は 0.8%から 68.7%まで上昇し，真陽性数の方が偽陰性数よりも多くなり，陽性を取り逃すことが大幅に減った．これは危険性のあるものを判定する際には望ましい方向の変化であった．ただし，正解率は 86.9%からおよそ半分の 44.3%

に低下し，ROC 下の面積は 53.3%から 57.0%に，PR 曲線下の面積は 14.4%から 15.7%にやや上昇しただけである．

図 6 に指標 p05 の ROC を示した．これは図 5 と同じくほぼ直線であった．しかし，陽性と推定した数が増加し，偽陽性率と真陽性率の平均を示す赤バツは ROC 上で右上に移動した．これにより再現率が高くなり，危険なものを予測する際に望ましい変化が見られた．詐欺脆弱性のように危険なものを拾うことが重要な問題では，不均衡クラスを考慮することは重要で，それにより危険なものを拾うようになったということである．

表 4 不均衡クラスを考慮したロジスティック回帰の推定性能

詐欺	指標	目的変数	TP	FN	FP	TN	正解率	適合率	再現率	f1 値	ROC	PR
オレオレ詐欺	p02	問 11_A	857	335	4604	2982	43.7%	15.7%	71.9%	25.8%	58.5%	17.5%
	p29	問 11_K	770	264	4865	2879	41.6%	13.7%	74.5%	23.1%	58.7%	15.4%
架空請求詐欺	p05	問 11_B	772	240	4991	2775	40.4%	13.4%	76.3%	22.8%	57.5%	14.3%
	p06	問 11_B	675	337	4699	3067	42.6%	12.6%	66.7%	21.1%	56.1%	14.2%
還付金等詐欺	p18	問 11_G	912	394	4241	3231	47.2%	17.7%	69.8%	28.2%	57.9%	17.9%
	p26	問 11_J	687	417	4422	3252	44.9%	13.4%	62.2%	22.1%	55.0%	15.4%
融資保証金詐欺	p31	問 11_L	727	447	4270	3334	46.3%	14.5%	61.9%	23.6%	55.9%	15.6%
	p32	問 11_L	774	400	4210	3394	47.5%	15.5%	65.9%	25.1%	56.4%	15.3%
平均			772	354.3	4538	3114.25	44.3%	14.6%	68.7%	24.0%	57.0%	15.7%

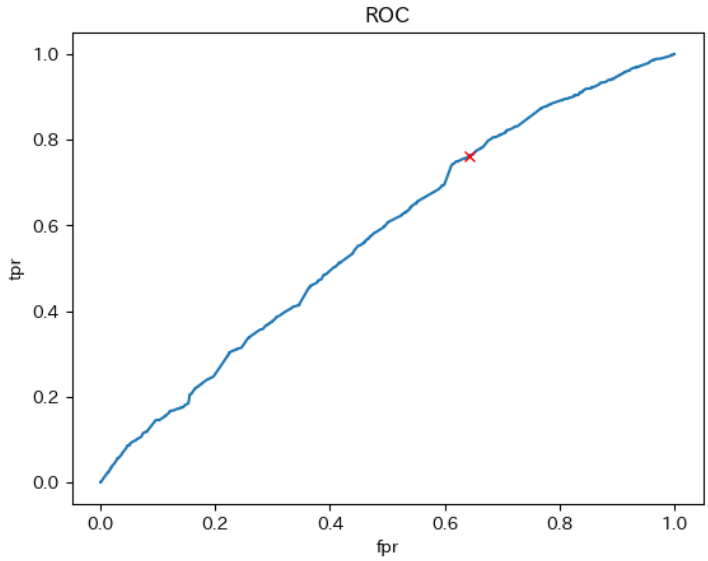

図 6 不均衡クラスを考慮したロジスティック回帰の指標 p05 の ROC

8. 学習データの増加，正則化による変数選択，自然な陽性・陰性

本論文の 6 節で，アプリの初版では再現率が低く 0.8%で，実際には陽性の 99.2%を陰性と誤認識し，ROC 下面積は 53.3%でランダムの 50%に近かったことを示した．7 節で，不均衡クラスを考慮すると，ROC 下の面積は 53.3%から 57.0%にやや上昇した程度であったが，陽性を取りこぼすことが減り，偽陽性率と真陽性率の平均の位置が ROC 上で右上に移動し，再現率は 0.8%から 68.7%まで上昇したことを示した．3 節に示したように，アプリの初版で学習に用いた質問紙 690 件のデ

ータは，含まれる陽性の数が少なかった．そのため先行研究[3]では目的変数を陽性・陰性に分ける際に謙虚な回答も危険と見なして陽性を増やしていたが，それでも学習データが足りなかった可能性がある．そこで，アプリで収集したデータ 8,778 件を学習に用いて性能が向上するかを調べることにした．アプリの初版の開発では質問紙調査で得られた 690 件の少ないデータしか利用できなかったので，すべてのデータですべてのデータを説明するパラメータを算出した．汎化性能は，アプリの初版のモデルの学習には利用していない，アプリで収集したデータを用いて評価した．今回は 8,778 件のデータを学習にも評価にも利用する．汎化性能は学習時に使っていないデータを用いて評価する必要があるので全データを学習 80%と最終的な評価に用いるテスト 20%に分割した．さらに学習には，ハイパーパラメータとパラメータの 2 種類に関するものがあり，それぞれ異なるデータを用いる必要がある．ハイパーパラメータとは推定に用いる関数の形など，パラメータの学習前に設定する必要があるものである．データをなるべく有効に活用するため，学習データを分割し交差検証を行った．具体的には，学習データを 5 等分した．4 つをパラメータの学習に使う訓練データ，1 つをハイパーパラメータの学習に用いる検証データにする．あるハイパーパラメータを設定して，訓練データで最適なパラメータを求める．そして検証データでハイパーパラメータの汎化性能を求める．検証データを交換して 5 回同じことを繰り返し，あるハイパーパラメータの設定のときの平均の汎化性能を求める．次に別のハイパーパラメータを設定し，同様のことを繰り返し，最適なハイパーパラメータの設定を求める．最後にあらかじめ取り分けておき，ハイパーパラメータとパラメータの学習には使わなかったテストデータで最終的な汎化性能を求める．

なお，アプリの初版の開発の際には，まだアプリが存在していなかったので，質問紙調査のデータを用いるしかなく，陽性の個数は最少の場合 3 個と少なかった．これを 3 分割以上すると陽性は平均 1 個以下しか含まれない．0 個の標本を分類することはできないので，3 分割交差検証すら行えなかった．そのため，アプリの初版ではデータが大量に得られなかった時代によく利用された方法，すなわち回答データを全部利用して，それを説明するモデルのパラメータを求める方法を採用していた．そのため，モデルの開発の段階では汎化性能が評価できなかった．そこで本研究では，モデルの開発後に実環境で収集したデータで評価した．

また，アプリの初版では過学習をおさえるために，推定に用いる説明変数を種類ごと分け赤池情報量基準で絞り込んで[3]モデルの複雑さをおさえていた．このとき少ないデータしか利用できなかったが，データが増えると変数選択も変わってくる可能性がある．また，説明変数を人間が経験から種類ごとに分けたがそれが最適という保証はない．そこで全説明変数を用いて正則化を行った．正則化はモデルのパラメータの値の絶対値が大きくならないように抑えるものである．モデルのパラメータの絶対値が大きすぎるということは，モデルの関数が複雑になっていたり，たまたま学習データに含まれていた特異なデータの影響を大きく受けていた可能性があるということである．これを避けようとするのが正則化である．今回の正則化は，パラメータの絶対値の和を用いる L1，2 乗の和を用いる L2 の 2 種類のいずれかで，強さ C は 1 から 10,000 までを対数スケールで 10 等分した値で，合計 2×10=20 種類を検証した．そして PR 曲線下の面積を最大化するハイパーパラメータとパラメータの組み合わせを見つけている．

また，アプリの初版の学習に用いた質問紙のデータは 690 件で，目的変数の値は「1 そうしない」「2 おそらくそうしない」「3 おそらくそうする」「4 そうする」の 4 択であった．アプリの初版では 2，3，4 番を危険な回答とみなしていたが，「2 おそらくそうしない」は危険というよりは謙虚な回答である．また，その場合，得られたアプリの推定も境界が「1 そうしない」「2 おそらくそうしない」の間になってしまう．一方，3，4 番を危険な回答とみなした場合，問 11_A は 3 人だけ

になり，機械学習の推定に用いるとたまたまこの 3 人に共通する特徴を学習してしまう過学習の危険性があった．しかし，アプリで収集した 8,778 件のデータを学習に用いるなら，1，2 番を安全，3，4 番を危険な回答とみなしても，最少の問 11_C でも陽性が 167 件になる．

そこで，アプリ初版と同様に目的変数の陽性・陰性を不自然な分類にした場合と，自然にした場合の 2 種類を評価した．まず，表 5 にアプリの初版と同様に安全な回答を 1，危険な回答を 2，3，4 として計算し，テストデータで汎化性能を評価したもの，表 6 に安全な回答を 1，２，危険な回答を 3，4 と自然な分類で評価したものを示した．また，指標 p05 に対応する問 11_B の ROC を図 7(安全 1，危険 2，3，４)，図 8(安全 1，2，危険 3，４)に示した．

アプリの初版と同様に安全な回答を 1，危険な回答を 2，3，4 とした場合(表 5)，再現率は 68.7%から 64.2%に若干減少したが，正解率は 44.3%から 70.9%に，ROC 下の面積(図 7)は 57.0%から 75.1%に大きく上昇した．また，PR 曲線下の面積も 15.7%から 34.5%に上昇した．

表 5 アプリで収集したデータで 5 分割交差検証 安全な回答 1，危険な回答 2，3，4

詐欺	目的変数	TP	FN	FP	TN	正解率	適合率	再現率	f1 値	ROC	PR
オレオレ詐欺	問 11_A	161	78	453	1065	69.8%	26.2%	67.4%	37.7%	76.0%	34.9%
	問 11_F	186	117	442	1011	68.2%	29.6%	61.4%	40.0%	71.6%	35.0%
	問 11_K	129	78	445	1104	70.2%	22.5%	62.3%	33.0%	73.3%	29.1%
架空請求詐欺	問 11_B	126	77	430	1124	71.1%	22.7%	62.1%	33.2%	75.0%	31.9%
	問 11_E	168	121	457	1011	67.1%	26.9%	58.1%	36.8%	71.6%	34.8%
還付金等詐欺	問 11_D	131	55	399	1172	74.2%	24.7%	70.4%	36.6%	81.1%	39.6%
	問 11_G	168	94	403	1092	71.7%	29.4%	64.1%	40.3%	75.8%	37.0%
	問 11_J	154	67	399	1136	73.5%	27.8%	69.7%	39.8%	79.9%	46.0%
融資保証金詐欺	問 11_C	105	46	439	1167	72.4%	19.3%	69.5%	30.2%	76.8%	32.4%
	問 11_I	109	73	443	1132	70.6%	19.7%	59.9%	29.7%	72.2%	24.7%
	問 11_L	144	91	420	1101	70.9%	25.5%	61.3%	36.0%	72.9%	34.3%
平均		143.7	81.5	430.0	1101.4	70.9%	24.9%	64.2%	35.8%	75.1%	34.5%

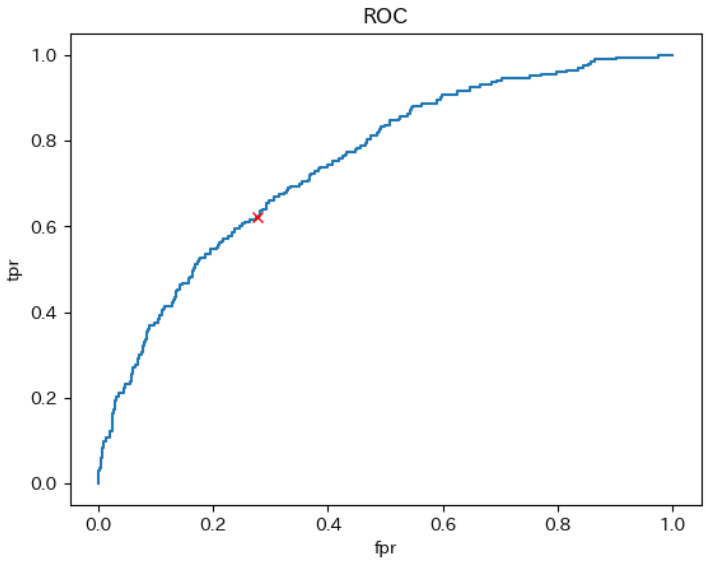

図 7 目的変数 問 11_B の ROC．アプリで収集したデータで 5 分割交差検証 安全な回答 1，危険な回答 2，3，4

安全な回答を 1，2，危険な回答を 3，4 と自然な分類にすると(表 6)，正解率は 80.0%，再現率は 68.9%，ROC 下の面積(図 8)は 82.2%と表 5 からさらに上昇した．PR 曲線下の面積は 26.6%と表 5 の 34.5%からは減少している．ただし，安全 1，危険 2，3，4 では真陽性が 16.7%だが，安全 1，2，危険 3，4 では真陽性 7.1%と少なく，問題設定が若干異なっているために PR 曲線下面積は単純には比較できない．それでも，学習データの量が少なかったアプリ初版(表 3)の 14.4%や，それに不均衡クラスを考慮した表 4 の 15.7%からは上昇している．

なお，性能向上に一番大きい寄与があったのはデータ数が 10 倍以上に増えたことである．データ数が 690 件の状態で正則化を用いても用いなくても影響はほぼなく，今回データ数が 10 倍以上に増えて違いが見られた．また，説明変数の値域はデータでは 1〜4 だが，これを 0〜1 にスケールさせる正規化や，平均 0 で標準偏差 1 にスケールさせる標準化は，データ数が少ない場合も多い場合も結果はほぼ変わらなかった．

表 6 アプリで収集したデータで 5 分割交差検証 安全な回答 1，2，危険な回答 3，4

詐欺	目的変数	TP	FN	FP	TN	正解率	適合率	再現率	f1 値	ROC	PR
オレオレ詐欺	問 11_A	32	8	358	1358	79.2%	8.2%	80.0%	14.9%	85.4%	34.5%
	問 11_F	59	18	381	1298	77.3%	13.4%	76.6%	22.8%	82.8%	25.6%
	問 11_K	26	16	329	1385	80.4%	7.3%	61.9%	13.1%	80.7%	23.9%
架空請求詐欺	問 11_B	29	22	313	1393	80.9%	8.5%	56.9%	14.8%	76.7%	25.9%
	問 11_E	66	34	388	1268	76.0%	14.5%	66.0%	23.8%	75.6%	29.3%
還付金等詐欺	問 11_D	32	13	286	1425	83.0%	10.1%	71.1%	17.6%	84.0%	17.1%
	問 11_G	53	16	374	1313	77.8%	12.4%	76.8%	21.4%	85.7%	29.0%
	問 11_J	37	13	321	1386	81.0%	10.3%	74.0%	18.1%	85.7%	26.7%
融資保証金詐欺	問 11_C	25	9	272	1451	84.0%	8.4%	73.5%	15.1%	90.5%	38.1%
	問 11_I	23	15	324	1394	80.7%	6.6%	60.5%	11.9%	77.8%	27.8%
	問 11_L	30	20	333	1373	79.9%	8.3%	60.0%	14.5%	79.6%	14.9%
平均		37.5	16.7	334.5	1367.6	80.0%	9.8%	68.9%	17.1%	82.2%	26.6%

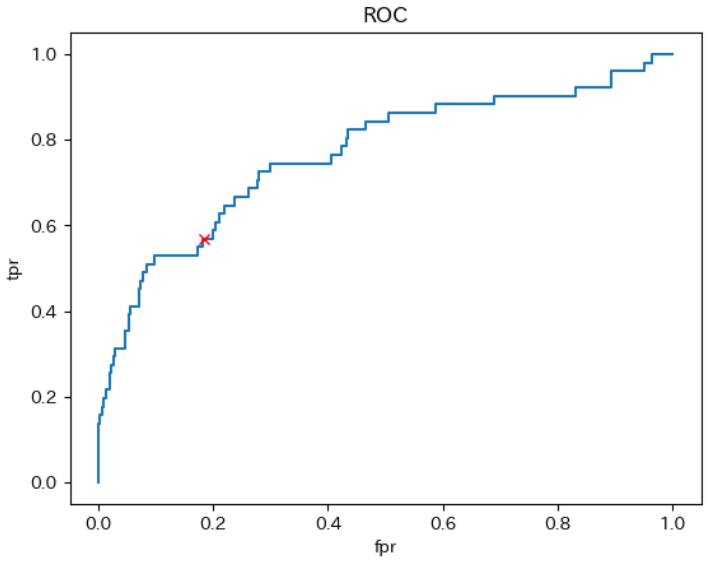

図 8 目的変数 問 11_B の ROC．アプリで収集したデータで 5 分割交差検証 安全な回答 1，2，危険な回答 3，4

9. まとめと考察

詐欺抵抗力判定アプリは機械学習の一種のロジスティック回帰を用いて詐欺抵抗力を判定していた．本論文では，その推定の汎化性能を評価し，性能を改善する方法を論じた．

アプリの初版では，不均衡クラスの問題に対応して陽性を増やすために不自然な陽性と陰性の分類を採用していた．これは詐欺に遭遇した場面ではっきり安全な行動をとると回答する人だけを安全な陰性とし，おそらく安全な行動をとるという謙虚な回答した人は危険な陽性と分類するものであった．それでも再現率は0.2%と低く，99.2%の陽性をうまく検出できないことが本研究により明らかになった．また ROC 下面積も 53.5%で，ランダムな推定の場合の 50%に近かった．ただし，正解率は 86.9%と高く，これは陽性が少なく，多い陰性が推定に強く効き，データの多くを陰性と推定したからである．

再現率が低かったのは，陽性が少ない不均衡クラスを考慮せず多数の陰性にひっぱられていたからである．不均衡クラスを考慮して，学習時に陽性・陰性の多寡の逆数で重みづけすると，陽性を拾うことができるようになり，再現率が 68.7%に向上した．一方，ROC 下面積は 57.0%であまり変化はなく，推定が陽性を拾う方向に変化したため，正解率は 44.3%に低下した．

学習データの量を増やすことで，陽性の数が増え，自然な陽性・陰性の分類を使うことができるようになった．再現率は 68.9%で少ないデータで不均衡クラスを考慮した場合とあまり変わらなかった．しかし，ROC 下面積は 82.2%に向上し全体的な性能が向上し，正解率も 80.0%になった．

よって，不均衡データの問題ではそれを考慮し，学習時に陽性・陰性の多寡の逆数の重みを付けることが重要である．これにより個数の少ないクラスもきちんと拾う方向に機械学習のモデルが変化する．また，機械学習の指標はいくつかあるが，トレードオフの関係になっているものがあり，学習データの量が少ないとさまざまな指標を同時に向上させることは難しいが，量を増やすと全体的に向上させることができる．学習データの量を増やすには，学習データを継続的に収集し続けることが重要である．そして，収集したデータで推定に用いるモデルも更新し続けることも重要である．

謝辞

本研究の元になった研究開発は，国立研究開発法人科学技術振興機構(JST)の社会技術研究開発センター(RISTEX)の戦略的創造研究推進事業(社会技術研究開発)「安全な暮らしをつくる新しい公／私空間の構築」研究開発領域に平成 29 年度に採択された「高齢者の詐欺被害を防ぐしなやかな地域連携モデルの研究開発」(研究代表者・渡部諭秋田県立大学教授)のものである．詐欺抵抗力判定アプリ「わたなべ教授のサギ抵抗力しんだ〜ん」に回答くださったみなさん，JST，研究開発領域の領域総括およびアドバイザーのみなさん，プロジェクトのメンバーみなさんに感謝いたします．また，八戸工業大学特定研究の令和 2，3，4 年度の助成を受けて実施しました．ありがとうございます．

参考文献

1) 科学技術振興機構 社会技術研究開発センター. 高齢者の詐欺被害を防ぐしなやかな地域連携モデルの研究開発. https://www.jst.go.jp/ristex/pp/project/h29_5.html <2022 年 10 月 28 日アクセス>

2) 渡部 諭, 岩田 美奈子, 上野 大介, 江口 洋子, 小久保 温, 澁谷 泰秀, 大工 泰裕, & 藤田 卓仙. (2018). 高齢者の詐欺被害を防

ぐしなやかな地域連携モデルの研究開発. 秋田県立大学ウェブジャーナル A (地域貢献部門), 5, 64-72. http://id.nii.ac.jp/1180/00000765/ <2022 年 10 月 28 日アクセス>

3) 渡部 諭. (2020). 高齢者の特殊詐欺抵抗力判定ルールの修正の試み. 国民生活研究, 60(1), 5-28.

4) 澁谷 泰秀, 吉野 諒三, 渡部 諭, 角谷 快彦, 藤田 卓仙, 小出 哲彰, 田中 康裕, & 大工 泰裕. (2019). 社会調査データに基づく特殊詐欺脆弱性判定の試み. 日本世論調査協会報 「よろん」, 123, 40-49. https://doi.org/10.18969/yoron.123.0_40

5) 渡部 諭, & 澁谷 泰秀. (2021). 高速倹約決定木による特殊詐欺抵抗力の判定. データ分析の理論と応用, 10(1), 29-44. https://doi.org/10.32146/bdajcs.10.28

6) 渡部 諭, & 澁谷 泰秀. (2021). 特殊詐欺抵抗力判定式改良の試み. 秋田県立大学総合科学研究彙報, (22), 1-6. http://id.nii.ac.jp/1180/00001192/ <2022 年 10 月 28 日アクセス>

要　旨

この論文では，詐欺抵抗力判定 Web アプリ「わたなべ教授のサギ抵抗力しんだ〜ん」の推定性能を論じる．

ユーザーが一連の質問に回答すると，このアプリは機械学習の一種であるロジスティック回帰などを用いて，詐欺脆弱性を判定する．詐欺脆弱性を推定する問題は機械学習で「不均衡クラス」と言われる種類の問題である．このアプリの判定には不均衡クラスの取り扱いに問題があり，学習データの陽性の数が少なかったため，推定性能が低かった．この論文では先行研究の検証を行い，推定性能の向上に寄与する手法を提案する．

キーワード: 機械学習，不均衡クラス，詐欺脆弱性，Web アプリケーション