

適応型テストを受検した場合の不適切解答者の検知

○歌川 真一郎*, 登藤 直弥**, 甲斐 江里***, 尾崎 幸謙****

筑波大学大学院ビジネス科学研究科博士後期課程*, 筑波大学人間系**, 株式会社リクルートマネジメントソリューションズ***, 筑波大学ビジネスサイエンス系****

1 研究の背景と目的

コンピュータを利用して行うテスト (Computer-based testing: CBT) が普及したこともあり, Web 上でのテストなど実施形態も多様化しているが, テストにおいて公正性が担保されることは非常に重要である。そこで, 本研究では Web 上でのテストにも用いられる適応型テスト (Computerized adaptive testing: CAT) において, 不適切解答をした者を検出することを目的とする。なお, 本研究における不適切解答とは, 参考書を見て, 友人に聞いて, あるいはネット検索を用いての解答など, Web で実施するがゆえに可能となる不正行為を指すこととする。不適切解答者の検出は, ①テスト問題を刷新するなどの対策の必要性の指標となること, ②受検者の解答結果から項目パラメタを推定したり等化を行う場合に, データから除外すべき受検者を同定できる, という意味で有益と考える。

機械学習を用いた不正行為の検出に関しては Yamana, Ohzeki and Aurélien(2015) でも行われているが, テスト会場において席が近い受検者間で行われるような不正行為を, 受検者すべてのペアごとの, 座席の配置や提示した項目間の関連性を考慮した不正行為に関する相関を用いて検出することを目的としたものであり, 本研究とは検出すべき不適切解答の対象が異なる。

2 研究方法

同一能力を測定するテストを複数回受検した者の中から不適切解答者を検知する方法について, 機械学習を用いたシミュレーションにより考える。

最初に CAT を用いたシミュレーションを行う。まず, 初期項目を提示し, それに対する反応から能力値 θ の初期推定値 $\hat{\theta}$ を求める。そして, その時点での $\hat{\theta}$ から最適な項目を提示し, 次の推定値 $\hat{\theta}$ を求める。項目提示の終了条件を満たすまでこれを繰り返し, テスト終了後に最終的な $\hat{\theta}$ を求める。この過

程を同一能力を測定する testA と testB のテスト 2 回分行う。その際, 適切に解答する者, testA のみ不適切解答する者, testB のみ不適切解答する者, testA, B とも不適切解答する者に分けておく。

この CAT の結果から, 適切に解答した者と不適切解答者では, 測定した 2 回の能力の推定値及び各項目の解答状況に異なる特徴が見られることを踏まえ, 有効な説明変数を作成し, 機械学習を行う。その後, 不適切解答と判別した者を除いて項目パラメタの推定を行い, 精度の改善について考える。

3 シミュレーションの設定

3.1 CAT

項目 各項目には 2 パラメタロジスティックモデルを当てはめる。困難度パラメタを $N(0, 1)$, 識別力パラメタを $U(0.5, 1.5)$ から発生させ, 各テスト 500 項目ずつの項目プールを作成する。

真値 受検者数 10,000 人として各テスト実施時の能力の真値について, 相関 $r = 0.8$ の多変量標準正規分布から発生させる。

解答 項目パラメタ及び受検者の能力の真値に基づいて正答確率を算出する。その確率に基づき全受検者の全項目分の解答を 2 値で作成する。

不適切解答の解答 受検者 10,000 人のうち 1,000 人を testA のみ不適切解答者, 1,000 人を testB のみ不適切解答者, 1,000 人を testA, B 両方の不適切解答者とする。作成した解答のうち不適切解答者の誤答の項目を正答に変更する割合 (以下, 誤答変更割合とする) を 8 割とし, 誤答項目からこの割合に応じて正答に変更する作業により不適切解答をしたものとする。

初期項目 各テストの最初の 3 項目は, $\theta = -1, 0, 1$ にもっとも困難度パラメタに近い 10 項目からランダムに出題する。初期項目以降は CAT によって項目提示を行う。なお, 項目選択基準にはフィッシャー情報量基準 (MFI) を用いる。

推定法 EAP 推定を用いる。事前分布は $N(0, 1)$ とする。

提示終了条件 θ の事後分布の標準偏差が 0.35 を下回った時点で提示終了とする。

3.2 機械学習

目的変数 少なくとも 1 つのテストで不適切解答をしている者を T, 適切に解答している者を F とする 2 値変数とする。

説明変数 次に示す 10 変数を説明変数とする。 $\hat{\theta}_A - \hat{\theta}_B$ の値 (変数 1), $|\hat{\theta}_A - \hat{\theta}_B|$ の値 (変数 2), $\hat{\theta}_A, \hat{\theta}_B$ 及び 2 回の推定値のマハラノビス距離 (変数 3-5), $(\hat{\theta}_A - \hat{\theta}_{regA})^2 + (\hat{\theta}_B - \hat{\theta}_{regB})^2$ (変数 6), 各テストごとの各受検者の解答した項目の困難度と解答の正誤 2 値データの点双列相関係数 (変数 7,8), 変数 7- 変数 8 の値 (変数 9), 変数 9 の値の絶対値 (変数 10)。ただし, $\hat{\theta}_A, \hat{\theta}_B$ は CAT による各テストの能力推定値, $\hat{\theta}_{regA}, \hat{\theta}_{regB}$ は $\hat{\theta}_{regA} = 0.0 + r\hat{\theta}_B$ のようにして得られた回帰推定値を表す。

予測モデル アンダーサンプリングにより TF 各々 1,000 人分ずつ抽出して訓練データとし, 残り 8,000 人分をテストデータとする。機械学習の方法としては, ロジスティック回帰モデル, K 最近傍法, 決定木, ランダムフォレスト, ブースティング, リッジ回帰, lasso を用いて比較する。ランダムフォレストでは訓練データからのブートストラップ法を, それ以外では訓練データに対する 10 分割交差検証を行い, パラメタのチューニングをしつつモデルを構築する。

4 シミュレーション結果と考察

機械学習による各方法についてその精度 (Accuracy: 正しく判断した割合), 適合率 (Precision: 予測が T のうち実際に T の割合), 再現率 (Recall: 実際に T のうち予測が T の割合) の値を表 1 に示す。表 1 より, どの方法も高い精度で不適切解答者を検出できたといえる。なお, 誤答変更割合を 5 割まで下げて同様のシミュレーションを行うと精度は機械学習の方法にもよるが 70% 程度まで低下した。

次に, 機械学習により検出した不適切解答者を除いて各テストごとに項目パラメタの推定を行ったとき, その推定精度が不適切解答者を除かずに推定し

表 1 機械学習ごとの精度, 適合率, 再現率

	Accuracy	Precision	Recall
Logistic	97.1%	94.8%	93.4%
KNN	93.7%	82.1%	95.6%
Tree	93.5%	83.5%	92.5%
RF	95.0%	86.1%	95.6%
Boosting	95.4%	87.1%	95.7%
Ridge	95.7%	90.2%	93.1%
Lasso	95.0%	86.8%	94.3%

たときと比べ改善が見られるかを調べた。なお, 項目パラメタの推定は全ての方法で 400 人以上の解答があった項目 (testA で 83 項目, testB で 80 項目) を対象に行い, 項目パラメタの bias や RMSE, 真値との誤差が 0.5 以上ある項目数について確認した。識別力パラメタについては改善は見られなかったが, 困難度パラメタについては bias, RMSE とも改善し, 10 程度あった誤差が 0.5 以上の項目は, 機械学習により検出した不適切解答者を除くことにより, 全項目について誤差が 0.5 以内に収まった。

さらに, テストが testA の 1 つだけ実施された場合の不適切解答者の検出を試みた。テストが 1 つだけのため, 説明変数には変数 3 及び変数 7 の 2 変数のみ用いたが, 誤答変更割合が 8 割のときの精度ではテストが 2 つの場合と同程度の結果が得られた。

5 今後の課題

本研究はシミュレーションのため, 実用化に向けては実際の CAT のデータに即して諸々の設定を調整していく必要がある。また, 実際には不適切解答をする受検者はあまり多くない。加えて, 当該データが確実に不適切解答によるものと特定することは難しい。実際のテストに用いる際には訓練データをどのように収集するかということが課題である。

文献

Yamanaka Shogo, Ohzeki Masayuki and Aurélien Decelle(2015) Detection of Cheating by Decimation Algorithm. *Journal of the Physical Society of Japan*, 84(2), 024801.