

Web-testingにおけるMLPとSMOTEを用いた筆記認証モデル

林 大介[†] 川又 泰介[†] 赤倉 貴子^{††}

[†] 東京理科大学大学院工学研究科 〒125-8585 東京都葛飾区新宿 6-3-1

^{††} 東京理科大学工学部 〒125-8585 東京都葛飾区新宿 6-3-1

E-mail: [†]{4418518,4417701}@ed.tus.ac.jp, ^{††}takakura@rs.tus.ac.jp

あらまし 現在の Web-testing における受験者認証法は、試験開始時の ID とパスワードのみであるため、試験中のなりすましが容易に行える。そこで、著者らは、受験者がタブレット PC に解答文字を記入した際の、筆圧やペン傾斜などのオンライン特徴量を基に受験者認証を行った。しかし、先行研究の認証精度は不十分であり、実用的に Web-testing へ適用するのは困難であった。その理由として、複数の特徴量を結合させるにあたり、線形回帰を用いていたことが挙げられる。そこで本研究では、距離結合に MLP を用いた筆記認証モデルを構築した。正例と負例の不均衡データの調整には SMOTE を適用した。結果として、先行研究より認証精度が向上し、MLP と SMOTE を用いた提案手法の有効性を確認できた。

キーワード Web-testing, MLP, SMOTE, 筆記認証, 機械学習

Writing authentication model using MLP and SMOTE in Web-testing

Daisuke HAYASHI[†], Taisuke KAWAMATA[†], and Takako AKAKURA^{††}

[†] Graduate School of Engineering, Tokyo University of Science 6-3-1 Nijuku, Katsushika-ku, Tokyo, 125-8585 Japan

^{††} Faculty of Engineering, Tokyo University of Science 6-3-1 Nijuku, Katsushika-ku, Tokyo, 125-8585 Japan

E-mail: [†]{4418518,4417701}@ed.tus.ac.jp, ^{††}takakura@rs.tus.ac.jp

Abstract Since the common examinee authentication method in Web-testing is based only on the ID and password at the beginning of the test, it is easy to spoof during the test. To solve this problem, we proposed an examinee authentication method based on online feature values such as pen pressure and pen tilt when the examinee wrote the answer characters on a tablet PC. However, the accuracy of our previous research was insufficient. Hence, it was practically difficult to apply the examinee authentication method of previous research to Web-testing. One reason for this is because linear regression was used to combine multiple features. In this research, we constructed a handwriting authentication model using MLP for distance coupling. In addition, we applied SMOTE to adjust the imbalanced data with positive and negative cases. As a result, the accuracy of the authentication has improved compared to our previous research. Therefore, the effectiveness of the proposed method using MLP and SMOTE was confirmed.

Key words Web-testing, MLP, SMOTE, Writing authentication, Machine learning

1. はじめに

情報化の進展を受けて、高等教育機関で e-Learning の導入が進んでおり、講義における時間的・空間的制約は緩和されつつある。しかし、試験における制約は依然として存在する。そのため、大学への通学が困難な社会人学生などは、単位認定試験などで大学に行く必要がある。

試験の制約を緩和するために、Web 上で試験を受験できる

Web-testing が有効であると考えられる。しかし、高等教育機関でその導入は進んでいない。原因として、現状の認証方式は、試験開始時の ID とパスワードのみであるため、試験中に「なりすまし」や「カンニング」といった不正行為が容易であることがあげられる。ゆえに、試験時間全体を通じた継続的認証手法を確立する必要があるといえる。

継続的認証手法を確立するために、バイオメトリクスに着目する [2]。バイオメトリクスは、指紋や顔などの身体的特徴を用

いたものと、音声や署名などの行動的特徴を用いたものに分けられる。近年では、指紋認証でログオンする PC や、顔認証を利用した入館ゲートなど、様々な場面でバイオメトリクスが用いられている。

ここで、Web-testing で受験者認証するためには、試験で自然に行われる行為であり、受験者の負担にならないことが必要である。さらに、継続的に受験者認証用のデータを取得できなければならない。指紋認証に関しては、受験者が問題を解くごとに指紋取得の動作を行わせる必要がある。ゆえに、試験の妨げとなり、身体的負担を伴ってしまう。一方、顔認証に関しては、顔を録画されながらテストを行うのは、受験者に精神的不快感を与えてしまう恐れがある。したがって、プライバシーの問題が存在するといえる。

そこで、図 1 で示した、タブレット PC を用いて、試験中の解答文字より受験者認証を行った筆記認証が提案されている [1]。受験者が解答を記入するごとに、解答文字の筆記データより受験者認証を行った。そのため、受験者の負担にならずに、継続的認証が可能である。しかし、実試験を想定した条件での認証精度は十分ではない。その理由として、複数の特徴量を結合させるにあたり、式 (1) の線形回帰を用いていたためと考えられる (X : 複合距離ベクトル, w : 重みベクトル, x : 距離ベクトル, e : 誤差ベクトル, i : 各特徴量の添字)。

$$X = \sum_{i=1}^6 w_i x_i + e \quad (1)$$

図 2, 図 3 と図 4 は、ある受験者が 3 と筆記した際の登録時 1, 登録時 2 と試験時の筆圧 (P), x 傾斜 (T_x), y 傾斜 (T_y) を表している。横軸の j, k, l は、筆記時間の要素数を意味する。 y 傾斜に着目すると、登録時に比べて試験時は大きく低くなっていることがわかる。 y 傾斜とは、筆記面の垂直方向を 0 度として、受験者側を正の角度、反対側を負の角度と設定したものである。調査した結果、登録時と試験時の姿勢の違いが原因であると分かった。登録時は筆記数字が指定されていたため、画面の指示文を注視する必要はなかった。しかし、試験時は画面に表示される設問と選択肢を注意深く読む必要があるため、姿勢が前かがみになっていた。ゆえに、ペンが受験者の反対側に傾きやすくなるため、 y 傾斜が低い値となっていた。したがって、先行研究 [1] で用いていた線形回帰では、正規の受験者となりすましを判別するのが困難であった。

以上の背景より、本研究では Web-testing における「なりすまし」「カンニング」といった不正行為のうち、なりすましを防止することを目的とする。先行研究 [1] より高精度の受験者認証を実現するために、特徴量の距離結合に MLP (Multi-Layer Perceptron) [3] を用いた筆記認証モデルを構築する。正例と負例の不均衡データの調整には SMOTE (Synthetic Minority Over-sampling Technique) [4] を適用する。

2 章でオンライン特徴量による受験者認証法と SMOTE, 3 章で MLP を用いた筆記認証モデルについて述べる。4 章では、実際の Web-testing 環境下で受験者認証を適用した際の認証精度を評価する。

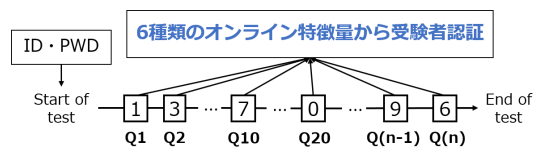


図 1 Web-testing の解答文字を利用した認証方式

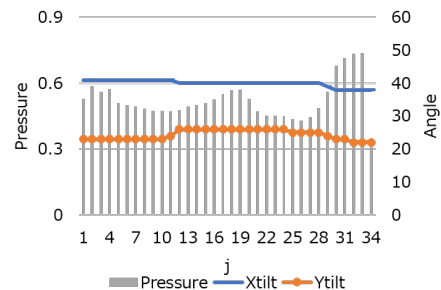


図 2 登録時 1 のペン傾斜と筆圧

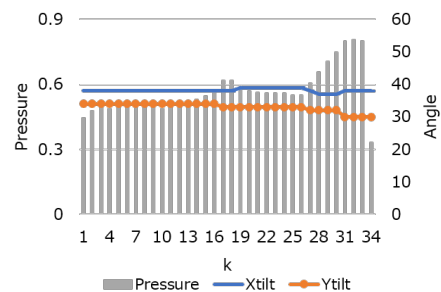


図 3 登録時 2 のペン傾斜と筆圧

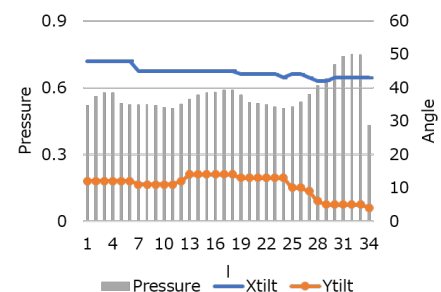


図 4 試験時のペン傾斜と筆圧

2. 受験者認証法

提案するアルゴリズムを図 5 に示す。提案アルゴリズムは、学習フェーズと試験フェーズからなる。

学習フェーズ: 各受験者における「0~9」の数字を登録する。1 回目に取得したものを学習データ 1 として扱う。学習データ 1 の取得から 1 ヶ月後に取得したものを学習データ 2 とする。学習データ 1 と学習データ 2 を DTW によって類似度計算を行い、距離ベクトルを算出する。距離ベクトルに MLP を適用することで、認証モデルを構築する。正例と負例の不均衡データの調整には SMOTE を適用する。

試験フェーズ: Web-testing 中に得られた解答文字を試験データとして扱う。試験データを学習フェーズで構築した認証モデルに入力し、推定確率を算出する。その値を閾値処理すること

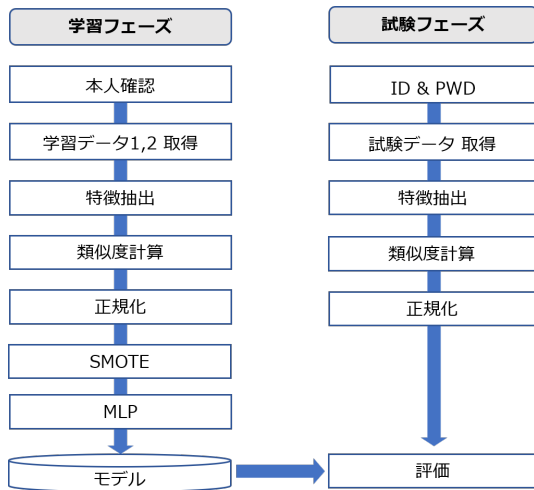


図5 アルゴリズム

で、正規の受験者となりすましの判定を行う。

2.1 特徴抽出

本研究ではオンライン特徴量を抽出する。オンライン特徴量を用いた理由は、静止画像よりも他人が真似ることが困難なため、個人認証に適合しているためである [5]。受験者が解答文字を筆記するたびに、筆記時間、 x 座標、 y 座標、筆圧、 x 傾斜、 y 傾斜の特徴量を抽出する。これら 6 種類の特徴量は、時系列データとして保存される。

ある受験者が 3 と筆記した際の、時系列データの例を図 6 に示す。表の行が時系列データの数、列が各特徴量を表している。本研究では、ペンダウン（筆圧 > 0 ）の時系列データを利用する。ペンダウンの行数は 92 行から 172 行までの計 81 行あり、特徴量は 6 種類ある。したがって、この例では、 $81 \times 6 = 486$ 個の時系列データを取得できる。

また、抽出した時系列データの前処理は、以下の 3 段階により行う。

- (1) ペンアップ時間（筆圧 = 0）の時系列データを除外する。
- (2) 左利きの受験者には、 x 傾斜の正負を反転することで、右利きの受験者同様に対応する。
- (3) 筆記時間の初期値を 0 に統一する。

2.2 類似度計算

データを分析するにあたり、筆記データの類似度（距離）を数値化する必要がある。しかし、筆記データは、字画の長さや筆記速度の違いにより、同じ文字を記入しても長さ（要素数）が異なる場合がある。そのため、単純な差分では距離を求めることはできない。そこで、本研究では、異なる長さの時系列データ間の距離計算に、動的時間伸縮法である DTW [2] を用いる。DTW は、2 つの時系列データが類似しているほど距離が小さくなる。ゆえに、この距離を基準に正規の受験者かなりすましかを判定する。DTW による距離計算の手順を以下に述べる。

比較する 2 つの時系列データである R 、 Q を式 (2) と式 (3) で表す。 I と J は時系列データ R 、 Q の要素数である。 r_i と q_j は、時系列データの i 番目と j 番目のデータである。

$$R = r_1, r_2, \dots, r_i, \dots, r_I \quad (2)$$

$$Q = q_1, q_2, \dots, q_j, \dots, q_J \quad (3)$$

初めに、距離の初期値を式 (4) とし、式 (5) で $g(i, j)$ の値を $g(I, J)$ まで変化させながら、最短距離を逐次計算していく。

$$g(i, j) = d(0, 0) \quad (4)$$

$$g(i, j) = d(i, j) + \min \left\{ \begin{array}{l} g(i, j-1) \\ g(i-1, j-1) \\ g(i-1, j) \end{array} \right\} \quad (5)$$

$d(i, j)$ は r_i と q_j について計算される距離である。筆記時間 (T)、 x 座標 (C_x)、 y 座標 (C_y)、筆圧 (P)、 x 傾斜 (T_x)、 y 傾斜 (T_y) における局所距離計算には、式 (6) のように差分の絶対値を用いた。

$$d(i, j) = |T_i - T_j| \quad (6)$$

逐次的に求めた距離 $g(I, J)$ を 2 つの時系列データの要素数の合計値を用いて式 (7) で正規化する。正規化された距離 x を 2 つの時系列データ R と Q の距離ベクトルとする。

$$x = \frac{g(I, J)}{I + J} \quad (7)$$

2.3 正規化

DTW によって算出した各特徴量における距離ベクトルの列を、式 (8) によって最小値 0、最大値 1 に正規化する [6]。 $x^{(i)}$ は特定のサンプルであり、 x_{min} は距離ベクトルの列における最小値、 x_{max} は最大値を表す。 $x_{norm}^{(i)}$ は、 $x^{(i)}$ の正規化後のサンプルである。

正規化は、個人内距離と個人間距離を対象にして行う。個人内距離は、受験者本人を対象にして距離計算したものである。一方、個人間距離は、受験者本人をほかの受験者全員と比較して距離計算したものである。MLP を適用するにあたり、個人内距離を正例、個人間距離を負例にラベル付けした。

正例ラベルの数 34,000 に対して、負例ラベル数は 306,000 となっている。ゆえに、パラメータを最適化するためには、不均

行数	時間	x 座標	y 座標	筆圧	x 傾斜	y 傾斜
1	2.603	85	28	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
92	18.885	52	59	0.3340	46	32
93	18.913	52	59	0.5078	46	32
94	18.929	52	59	0.5830	46	32
⋮	⋮	⋮	⋮	⋮	⋮	⋮
170	20.198	54	105	0.7783	39	17
171	20.215	54	105	0.7715	39	17
172	20.231	54	104	0.5410	39	17
⋮	⋮	⋮	⋮	⋮	⋮	⋮
179	20.395	60	148	0	0	0

図6 時系列データの例

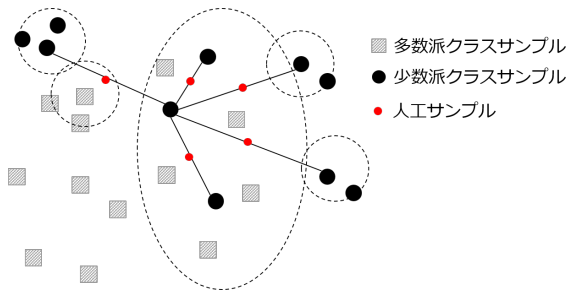


図7 SMOTE

衡データを同数に調整する必要がある。

$$\mathbf{x}_{norm}^{(i)} = \frac{\mathbf{x}^{(i)} - \mathbf{x}_{min}}{\mathbf{x}_{max} - \mathbf{x}_{min}} \quad (8)$$

2.4 SMOTE

本研究では、不均衡データへの調整に SMOTE(Synthetic Minority Over-sampling Technique) [4] を適用する。SMOTE の概要を図7に示す。SMOTE は、Oversampling の代表的な手法として用いられている。少数派クラスサンプルの最近傍点を線でつなぎ、その線分上の任意の点をランダムに人工データとして生成する。少数派クラスのデータを単に複製する方法 (Oversampling) では過学習してしまう。ゆえに、SMOTE によって人工データを生成することで、重大な過学習を伴わずにデータのバランスを調整する。

2.5 評価方法

2.5.1 EER

照合性能を表す指標として、本人拒否率 (False Rejection Rate; FRR), 他人受入率 (False Acceptance Rate; FAR), 等価エラー率 (Equal Error Rate; EER) [7], ROC 曲線下面積 (Area Under the ROC Curve; AUC) [8] を求める。AUC に関しては、表1より、高精度を意味する 0.9 以上となることを目指す。

FRR と FAR は以下の式で算出される。

$$FRR(\theta) = 1 - \int_0^\theta P(s|E=1)ds \quad (9)$$

$$FAR(\theta) = \int_0^\theta P(s|E=0)ds \quad (10)$$

$P(s|E)$ は、受験者 E が本人 ($E=1$) もしくは他人 ($E=0$) であったときに、類似度が s であった割合である。FRR(θ) は、類似度の閾値を θ と定めたときに、正規の受験者をなりすましとする誤りである。一方、FAR(θ) は、なりすましを正規の受験者とする誤りである。FRR と FAR は、閾値 θ の変化に対してトレードオフの関係にある。ゆえに、FRR と FAR が等しく

表1 AUC の推定指標 [8]

AUC	Evaluation
0.9 - 1.0	High accuracy
0.7 - 0.9	Moderate accuracy
0.5 - 0.7	Low accuracy

なる最適閾値 θ の誤り率 EER が認証精度の評価に用いられる。

$$\theta = \arg \min_{\theta} \frac{FRR(\theta) + FAR(\theta)}{2} \quad (11)$$

$$EER = \min_{\theta} \frac{FRR(\theta) + FAR(\theta)}{2} \quad (12)$$

3. 筆記認証モデル

3.1 MLP

本研究では、距離結合に MLP (Multi-Layer Perceptron) [3] を利用する。構築した MLP の構造を図8に示す。ネットワーク構造は、入力層、隠れ層、出力層の3層からなる多層パーセプトロンである。隠れ層のユニットは入力層と完全に結合しており、出力層は隠れ層と完全に結合している。

入力層には、6次元の距離ベクトルを入力する。隠れ層の直後に、活性化関数 Tanh と Dropout [9] を用いる。Tanh は、出力を $[-1, 1]$ の範囲に調整する。Dropout とは、ネットワークを学習する際に、ある更新で層の中のいくつかのノードを無効にして学習を行い、次の更新では別のノードを無効にして学習を繰り返すことを意味する。これにより、学習時にネットワークの自由度を強制的に小さくして汎化性能を上げ、過学習を抑制する。本研究では、Dropout の確率を $p=0.5$ に設定した。

出力層は式 (13) の Sigmoid 関数を用いる。 $\phi(z)$ は活性化関数、 x は距離ベクトル、 w は重みベクトルを表している。 z は式 (14) で表される総入力である。 $\phi(z)$ は z が大きいほど 1 に近づき、小さいほど 0 に近づく関数である。 $\phi(z)$ が 0.5 以上で正例、0.5 未満で負例と推定する。

$$\phi(z) = \frac{1}{1 + e^{-z}} \quad (13)$$

$$z = \mathbf{w}^T \mathbf{x} = w_1x_1 + w_2x_2 + \dots + w_6x_6 \quad (14)$$

MLP の出力を計算する順伝搬法は、以下の3つのプロセスから成り立つ [6]。複数の手順でこれらの手順を繰り返し、MLP の重みを学習した後、順伝搬法を使ってネットワークの出力を計算する。

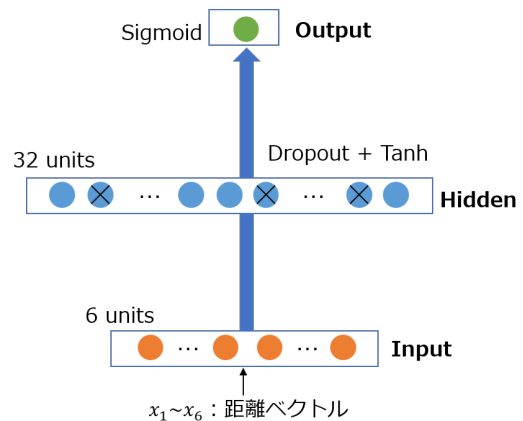


図8 MLP の構造

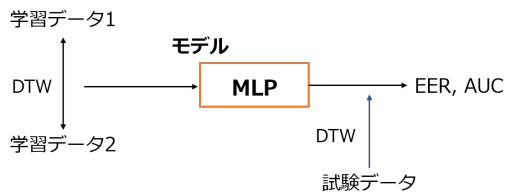


図9 筆記認証モデル

表2 実験データ

データ	人数	取得時期	記述内容	データ数
学習データ1	10	試験の1ヶ月前	「0~9」を17セット	1700文字
学習データ2	10	試験の2週間前	「0~9」を20セット	2000文字
試験データ	10		英語テスト40問 (四肢択一式)	400文字

(1) 入力層を出発点として、学習データのパターンをネットワーク経由で順方向に伝搬させ、出力を生成する。

(2) ネットワークの出力に基づき、活性化関数を使って誤差を計算する。この誤差を最小化することが目的となる。

(3) 誤差を逆方向に伝搬させることで、ネットワーク内の各重みに対する偏導関数を求め、モデルを更新する。

学習の最適化手法には、Adam [10] を $\alpha=0.001$, $\beta_1=0.9$, $\beta_2=0.999$ に設定して用いる。ミニバッチサイズを128に設定し、50エポックを上限として誤差が減少しなくなった時点で学習を終了することにする。MLPの実装は、PythonのライブラリであるKeras [11] を用いる

3.2 筆記認証モデル

試験前に、距離結合にMLPを用いた筆記認証モデルを構築する。筆記認証モデルの概要を図9に示す。利用するデータを表2に表す。受験者10名から1ヶ月の間隔をあけて、学習データ1、学習データ2を取得する。データ取得に用いたデバイスは、タブレットPC (Surface Pro) である。学習データ1と学習データ2をDTWによって比較した距離ベクトルをMLPに適用して、認証モデルを作成する。

学習結果を図10に示す。エポックを増やすごとに誤差が収束していることが分かる。本研究では、誤差が収束した7エポックまで学習を継続した。

4. 実試験環境実験

4.1 実験概要

本実験では、Web-testing 中になりすましを防止して、先行

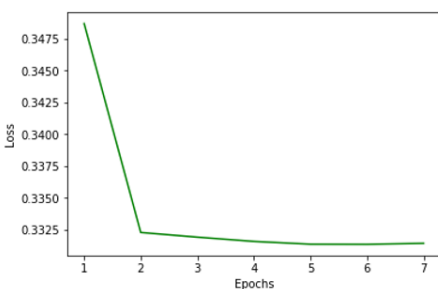


図10 学習結果

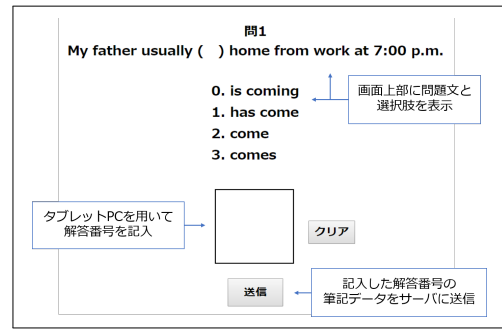


図11 Web-testing システム

研究 [1] よりも高精度な受験者認証を実現できるかを検証する。実際の Web-testing の環境下で受験者 10 名から、タブレット PC により四肢択一式の英語テスト 40 問を解答してもらい。試験の実施は、学習データ 2 の取得から 2 週間後に行った。試験より取得したデータを照合データとして扱う。受験者 10 名は認証モデルを構築したときと同じ被験者である。照合データを認証モデルに入力することで、認証精度を出力する。

4.2 出題画面の例

図 11 に本実験で利用した Web-testing システムを示す。受験者は、画面中央の枠内にタブレット PC を用いて解答番号を記入する。記入後に送信ボタンを押すと、座標位置や筆圧、ペン傾斜などの時系列データがサーバに送信される。

出題形式は、「0~9」の数字を均等に出题するために四肢問題の解答選択肢を、問 1~8: 「0, 1, 2, 3」, 問 9~16: 「4, 5, 6, 7」, 問 17~24: 「8, 9, 0, 1」, 問 25~32: 「2, 3, 4, 5」, 問 33~40: 「6, 7, 8, 9」と設定した。このように問題設定することで「0~9」の数字をすべて選択肢に含めた場合でも、条件を満たす認証精度を算出できることを意図とした。

4.3 結果

実験結果を表 3 に、ROC 曲線を図 12 に示す。試験データにおいて、正例ラベル数 : 708 に対し、負例ラベル数 : 14,618 となっている。MLP の有効性を確認するために、先行研究 [1] の線形回帰に加えて、NB (Naive bayes) [12], SVM (Support vector machine) [13], RDF (Random decision forest) [14], KNN (K-nearest neighbor) [15] とも比較した。結果より、MLP は EER, AUC ともに分類手法の中で最も良い精度を示していることがわかる。MLP の EER は 13.55% となっており、線形回帰 [1] の EER : 19.50% よりも 5.95% 優れた精度を示している。また、MLP の AUC は 0.9384 であり、高精度を意味する 0.9 以上の値を得ることができた。したがって、距離結合には MLP が有効であることを確認した。

一方、ROC 曲線を見ると、線形回帰 [1] の線は折れ線になっていることがわかる。ゆえに、閾値を変化させていく際に、正規の受験者かなりすましかを判定するのは困難であるといえる。

4.4 考察

不均衡データへの対処法に用いた SMOTE の有効性を確認するために、異なる Over-sampling 法と比較した結果を表 4 に示す。SVM-SMOTE [16] は、過学習を抑制するために SVM を用いて人工サンプルを生成したものである。ADASYN [17] は、

表 3 実験結果

Model	EER[%]	AUC
MLP	13.55	0.9384
NB	15.56	0.9242
SVM	15.83	0.8416
RDF	17.71	0.8962
KNN	18.02	0.8586
線形回帰 [1]	19.50	0.8049

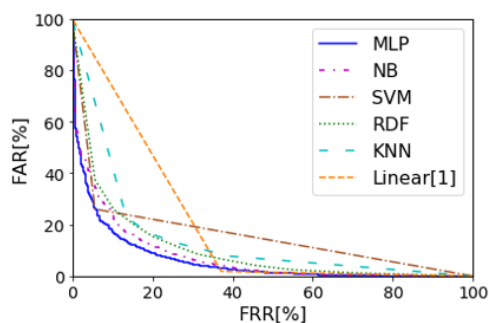


図 12 ROC 曲線

表 4 異なる Over-sampling 法との比較

MLP	SMOTE	SVM-SMOTE	ADASYN	B-SMOTE
EER[%]	13.55	13.95	14.40	15.06
AUC	0.9384	0.9349	0.9320	0.9276

KNN を利用して誤分類した元のサンプルの隣に新しいサンプルを生成する手法である。B-SMOTE [16] は、サンプルの近傍とクラスの関係より、ノイズ・危険・安全に分類する。危険と分類されたサンプルに対して新しいサンプルを作成するものである。

結果より、SMOTE は Over-sampling 法の中で、EER と AUC とともに最も良い精度を示していることがわかる。ゆえに、MLP と SMOTE を用いた提案手法の有効性を確認できた。

5. むすび

本研究では Web-testing において、受験者認証を先行研究 [1] よりも高精度に行うことで、なりすましを防止することを目的とした。受験者がタブレット PC に解答文字を記入した際の、筆圧やペン傾斜など 6 種類のオンライン特徴量を基に受験者認証を行った。先行研究 [1] は、特徴量の距離結合に線形回帰を用いており、認証精度に課題が残るものであった。そこで本研究では、距離結合に MLP を用いた筆記認証モデルを構築することを提案した。また、正例と負例の不均衡データの調整には SMOTE を適用した。

実験結果より、MLP の EER は 13.55% となっており、線形回帰 [1] の EER : 19.50% よりも 5.95% 優れた精度を示した。また、MLP の AUC は 0.9384 であり、高精度を意味する 0.9 以上の値を得ることができた。したがって、距離結合には MLP が有効であることを確認した。

SMOTE をほかの Over-sampling 法と比較した結果、EER と AUC とともに SMOTE が最も良い精度を示した。ゆえに、

MLP と SMOTE を用いた提案手法の有効性を確認できた。

今後の課題は、認証精度の更なる向上を目標に、オンライン特徴量の認証に加えて、解答文字の静止画像によるオフライン認証を追加することなどが挙げられる。

文 献

- [1] 林 大介, 赤倉貴子, “e-Testing におけるタブレット PC とオンライン筆記情報を用いた筆記認証法の提案,” 日本教育工学会論文誌, vol.42, no.Suppl, pp.101-104, 2018.
- [2] 半谷精一郎, “バイオメトリクス教科書～原理からプログラミングまで～,” (社)映像メディア学会, (編)コロナ社, 東京, 2012.
- [3] S. Maheshwary, S. Ganguly, V. Pudi, “Deep Secure: A Fast and Simple Neural Network based approach for User Authentication and Identification via Keystroke Dynamics,” International Joint Conference on Artificial Intelligence, Aug. 2017.
- [4] N. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” Journal of Artificial Intelligence Research, vol.16, no.1, pp.321-357, 2002.
- [5] 中村善一, 木戸出正継, “筆跡鑑定の知見に基づく特性値を用いたオンライン筆者照合,” システム制御情報学会論文誌, vol.22, no.1, pp.37-47, 2009.
- [6] S. Raschka, V. Mirjalili, “Python 機械学習プログラミング達人データサイエンティストによる理論と実践,” 高橋隆志 (編), インプレス, 東京, 2018.
- [7] 川又泰介, 赤倉貴子, “e-Testing における Web カメラとペンタブレットを用いた逐次受験者認証システムの開発,” 電子情報通信学会論文誌 D, vol.J102-D, no.3, pp.163-172, 2019.
- [8] A. Akobeng, “Understanding diagnostic tests 3: Receiver operating characteristic curves,” Acta Paediatr, vol.96, no.5, pp.644-647, 2007.
- [9] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” Journal of Machine Learning Research, vol.15, no.1, pp.1929-1958, 2014.
- [10] D. P. Kingma, J. L. Ba, “Adam: A Method for Stochastic Optimization,” arXiv:1412.6980 [cs. LG], 2014.
- [11] F. Chollet, “Keras Documentary,” <https://keras.io>, Nov 2019.
- [12] H. Zhang, “The Optimality of Naive Bayes,” in Proceedings of the Seventeenth International Artificial Intelligence Research Society Conference, (V. Barr and Z. Markov, eds.), Miami Beach, FL: AAAI Press, 2004.
- [13] T. Harris, “Credit scoring using the clustered support vector machine,” Expert Systems with Applications, vol.42, no.2, pp.741-750, 2015.
- [14] 浜田祐一, 栗山繁, 向井智彦, “ランダムフォレスト法を用いた動作検出,” 電子情報通信学会 D, vol.J93-D, no.11, pp.2516-2524, 2010.
- [15] D. A. Adeniyi, Z. Wei, Y. Yongquan, “Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method,” Applied Computing and Informatics, vol.12, no.1, pp.90-108, 2016.
- [16] H. M. Nguyen, E. W. Cooper, K. Kamei, “Borderline Over-sampling for Imbalanced Data Classification,” IEEE SMC Hiroshima Chapter, vol.2009, no.1, pp.24-29, 2009.
- [17] H. He, Y. Bai, E. A. Gracia and S. Li, “ADASYN: Adaptive synthetic sampling approach for imbalanced learning,” 2008 International Joint Conference on Neural Networks, pp.1322-1328, 2008.