クラウド型 CAPTCHA サービスにおける機械学習を用いたボットの検知

荒井　毅[1,*]　岡部 寿男[1]　松本　悦宜[2]　川村　剛司[2]

概要：近年，パスワードリスト攻撃を始めとしたボットを用いた不正アクセスによる被害が増加している．ボットによる自動アクセスを防ぐ手段として CAPTCHA が利用されているが，ボットの高度化に対応して CAPTCHA の難度が高くなり，ユーザの利便性の低下が問題視されている．この問題の解決策の 1 つとして，ボット検知技術を CAPTCHA に応用し，ボット利用のリスクが高いアクセスにのみ難度の高い CAPTCHA を出すことが検討されている．本研究では，商用サービスとして広く利用されているクラウド型 CAPTCHA サービスである Capy パズル CAPTCHA において，CAPTCHA 難度を変更するためのアクセスごとのボット利用リスクの判定について検討した。実際のサービスのアクセスログに対して過去に攻撃を検知した実績からボットであるかの判別フラグを付け，教師あり機械学習での判別を試みた．判別のモデルには XGBoost を用い，特徴量として DNS 逆引き情報や地理情報，UserAgent および CAPTCHA 回答時間などを用いた．また，結果における False Positive データについて解析し，ボット判別フラグの付与されていないデータにおけるボット利用アクセスの検知を行った．

キーワード：CAPTCHA, ボット検知，機械学習，クラウド

Detection of Bots in CAPTCHA as a Cloud Service Utilizing Machine Learning

Tsuyoshi Arai[1,*]　Yasuo Okabe[1] Yoshinori Matsumoto[2]　Koji Kawamura[2]

Abstract: In recent years, the damage caused by unauthorized access using bots has increased. Compared with attacks on conventional login screens, the success rate is high and detection is difficult. It is considered that the user's convenience declines because the difficulty of the CAPTCHA becomes high corresponding to the advancement of the bot. As a solution, applying bot detection technologies to CAPTCHA is considered. In this research, we focus on capy puzzle CAPTCHA which is widely used in commercial service. We examined to estimate the risk of each access to change the difficulty of CAPTCHA. Based on the attacks detected in the past, we added a flag to determine bot. We tried to discriminate the flags using supervised learning. We used XGBoost as a model. We used reverse DNS response, Http-User-Agent and response time of CAPTCHA as features. Moreover, we analyzed the False Positive data and detected the bot from the data which has no bot discrimination flag.

Keywords: CAPTCHA, bot detection, machine learning, cloud

1. はじめに

近年，不正アクセスによる被害が急激に増加している．不正アクセス被害の代表的なものとして，認証の突破による個人情報の漏洩や，不正なポイント変換による金銭被害などが挙げられる．

不正アクセスの手口としてボットを用いたパスワードリスト攻撃が挙げられる．パスワードリスト攻撃とは，不正に入手した ID とパスワードのリストを用いて，複数の web サイトに正規のルートから不正アクセスを行う攻撃である．このような攻撃は，従来の総当たりの攻撃と比較して ID あたりのログイン試行回数が少なく，検知が難しいという特徴がある [3]．ボットとは，人間を装い Web サイトの操作などを行う自動プログラムの総称である．従来型のユーザ ID とパスワードによる認証はボットによる短時間の自動アクセスによって破られやすいという弱点がある．

ボットによる自動アクセスを防ぐ手段として CAPTCHA が存在する [15]．CAPTCHA とは，機械と人間の判別を自動で行うチューリングテストである．代表的な CAPTCHA は Google 社の運用する reCAPTCHA などが挙げられ，ボットのアクセスを防ぐ手段として広く利用されている [14]．その一方で，高度な光学文字認識技術を使用したボットを用いて自動的に CAPTCHA を破る手法が生み出され，認証精度の低下が問題となっている [5],[6],[11]．ボットの光学文字認識技術の高度化による CAPTCHA の突破の対抗手段として CAPTCHA の複雑化が進み，ユーザの利便性の低下も問題化している [4],[16].

CAPTCHA のユーザビリティと強度の問題に対する解決策の 1 つとして，ボット検知技術を CAPTCHA に応用する

1 京都大学
Kyoto University
2 Capy 株式会社
Capy Japan Inc.

* arai@net.ist.i.kyoto-u.ac.jp

といった手法が考えられる．具体的には，ボット利用のリスクが高いアクセスについてのみ難度の高いCAPTCHAを出すことで，人間の利便性を確保するとともに，ボットによる不正アクセスの成功確率を下げることが可能である．CAPTCHA の難易度をリスクベース認証のアプローチを用いて動的に変更する手法は Google 社の開発した Invisible-reCAPTCHA によって実用化されている [9].

本研究では，商用サービスとして広く利用されているクラウド型CAPTCHAサービスに対してボット検知技術を付加し，ボットを検知した際に CAPTCHA の難度を変更し，ボット利用を制限できるようなシステムについて検討する．具体的には，実際のクラウド型CAPTCHAサービスのアクセスログに過去の検知の実績からボット判別フラグを付与し，データセットの特徴を教師あり機械学習モデルに学習させることでボット検知を行う．なお，本論文におけるボットという単語は，パスワードリスト攻撃で利用されるボットを指す．

本研究の主要な貢献は，手動でブラックリストを作成し，機械的に付与を行なったボット判別フラグを学習させたモデルにおいて高い精度のボット判別フラグの予測ができたこと，およびボット判別フラグの付与されていないデータのうち，モデルがボット利用であると判定したもの(FalsePositive データ)から複数のボット判別フラグが付与されたものと同様のボット利用アクセスを検出できたことである．教師あり機械学習モデルに XGBoost, LightGBM を用いて上記のシステムを実装することで，ROC-AUC 0.99 以上の高い精度が得られ，別期間のデータと照らし合わせることで 2 件の ISP のボット判別フラグの付与されていないデータから，同様のボットを検知することができた．

本論文では，まず 2 章で前提となるクラウド型 CAPTCHA サービスについての説明する．3 章では目標とするボット検知システムと検知に用いる情報について説明する．4 章ではボット検知システムの構築のためのアクセスデータの調査とボット利用データの特徴の調査を行い，5 章においては予測に用いる教師あり機械学習モデルの構築について示す．6 章では予測のテスト結果を示す．7 章ではモデルにおける False Positive(誤検知)に該当するデータについて調査を行い，それらのデータの中にボット利用データが含まれていないかの考察を行う．8 章にて本論文をまとめる．

2. クラウド型 CAPTCHA サービス

本節では，本研究の前提となるクラウドを利用して提供される CAPTCHA サービスである Capy パズル CAPTCHA に関して説明する．

パズル CAPTCHA は，Capy 株式会社の提供するクラウド型 CAPTCHA サービスである(a)．パズル CAPTCHA では一般的なクラウド型 CAPTCHA と同様に，導入する Web サイトが現在使用している認証プラットフォームにカプセル化した CAPTCHA に関するスクリプトを追加することで実装する．

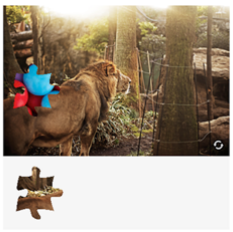

図 1 パズル CAPTCHA の認証画面(b)

3. ボット検知システム

本節では，本研究で提案するクラウド型 CAPTCHA サービスにボット検知技術を付加したクラウド型 CAPTCHA システムついて説明する．本システムのゴールは，クラウド型 CAPTCHA サービスにボット検知技術を取り入れ，ボット利用を検知した際に CAPTCHA の難度を高くすることで，ボットによるアクセスを防ぐことである．ボット検知の手法の 1 つとして，アクセスログから抽出した静的なルールをベースにボットをブロックする手法が考えられる [1],[12]．また，CAPTCHA を解く際の挙動を用いてボット検知を行う手法も提案されている [13]．しかし，日々新しく更新される新種のボットを検知することが難しいという欠点がある．そこで，教師あり機械学習をボット検知に用いることで，膨大なログの中からボットの特徴の効率的な抽出を行う手法が提案されている [2].

3.1 ボット利用アクセスの定義

本研究ではボット利用アクセスの定義として，ボット判別フラグを用いる．なお，本論文におけるボットは，パスワードリスト攻撃に用いられるボットを指すこととする．

本システムでは，ボット判別フラグが付与されているアクセスをボット利用の危険があるアクセスとして扱う．本システムは Web サイトアクセス時点での情報からボット判別フラグの予測を行い，失敗すると予測されたアクセスについて，ボット利用の危険があるアクセスであるとの判定を行う．

本研究では 2 節で挙げたパズル CAPTCHA サービスの実際のユーザのアクセスログの一部である 2,618,899 件にボット判別フラグを付与したものをボット検知のためのデータセットとして用いる．本研究に用いるアクセスログは複数の環境のデータを混合したものである．

(a) https://www.capy.me/jp/products/puzzle_captcha/

(b) https://www.capy.me/products/

ボット判別フラグの付与手順については以下の通りである．CAPTCHA を設置しているサイトのうち，ログ期間内に特定の環境においてパスワードリスト攻撃の疑いが強いアクセスを確認した．本研究でこれらの事例で使用されたアクセスをボットと判定し，ボット判別フラグの付与を行った．具体的には以下の事例である

・事例 A

該当する環境の通常の CAPTCHA では取得できないログと照らし合わせ，疑いのあるアクセスの中で，これらのアクセスのうち通常運用では考えられない短期間での大量アクセスがあった IP アドレスをボットとして判定した．

・事例 B

事例 A とは別の環境において，特定の IP レンジから同様のボットによる攻撃を確認した．これらのアクセスは IP アドレスが頻繁にかわるものの，特定の ISP レンジからアクセスされていた．該当する ISP の IP アドレスレンジをボットとして判定した．

最後に，本研究に用いるデータセットに対して上記事例に該当する IP アドレスからのアクセスに機械的にフィルタリングを行うことでボット判別フラグの付与を行った．ブラックリスト上の IP アドレスのレンジは最大で CIDR 表記/16 単位であった．

3.2 ボット検知に利用できる情報

本研究では 3.1 節の手法でボット判別フラグを付与したパズル CAPTCHA サービスの実際のユーザの試行ログの一部である 2,618,899 回分をボット検知のためのデータセットとして用いる．

本研究で利用したデータにおいてボット判別フラグの予測に用いることができる情報は，アクセス時刻，IP アドレス，HTTP User Agent, HTTP Accept Language，パズル ID，セッション ID，クッキーID, CAPTCHA の軌跡，CAPTCHA 回答時間の 9 つである．また，IP アドレスの逆引きで得られる FQDN に関する情報，GeoIP2 を用いて得られる地理情報や ISP に関する情報もボット検知に利用することが可能である [8]．さらには，HTTP User Agent を元に回答に利用した端末や OS の情報を取得し，利用することが可能である(c).

4. データセットとボット特徴

本節では，機械学習アルゴリズムの実装の前段階として，本研究に使用したデータセットの詳細とボット判別フラグのつけられたデータの持つ特徴について調査を行った結果について述べる．

4.1 データセットの詳細

本研究に用いるデータセットは日本時間 2019 年 3 月 1 日 8 時 59 分 59 秒から 2019 年 7 月 1 日 8 時 59 分 59 秒までの 4 ヶ月分のパズル CAPTCHA の実際のアクセスログである．データセットには，Web サイトへのアクセス，CAPTCHA 画像取得，CAPTCHA 回答の 3 手順のアクセスログの一部にあたる 2,618,899 件が含まれており，それぞれの手順で HTTP User Agent, IP アドレスをはじめとした共通の情報とパズル ID，CAPTCHA 回答時間などの固有の情報が記録される．データセットに含まれるログの情報は下記のように記録される．

表 1 データセットのアクセスログの例

datetime	captcha_key_num	name	result
2019-06-16 09:27:18	491	verify	incorrect-answer
challenge_key	challenge_info	cookie_id	answer
Vez1pAsWk…	NaN	NaN	14,F,0x...
remote_address	enduser_ip_address	elapsed_time	method
A.B.143.215	C.D.68.215	4583	POST
http_user_agent	http_accept_language	duration	is_bot
HTTPClient/1.0…	NaN	0.025	False

4.2 データセットの解析

本節では，利用するデータセットに含まれる情報と，IP アドレス，HTTP User-Agent から取得可能な地理情報や OS などの情報の分布について解析を行った結果を示す．

4.2.1 アクセス元の国情報

アクセス元の国情報について，GeoIP2 を用いて IP アドレスから取得し，解析を行った．データセット内に登場した国の数は 225 件であり，国情報が不明なものが 1%程度存在した．そのうちの上位 7 件を表 2 に示す．国情報が不明なものは，GeoIP2 に登録されていない IP アドレスからのアクセスであり，全体の傾向としてはアメリカからのアクセスが多数である．

表 2 アクセス元の国情報の内訳

国名	件数	比率
日本	2,157,142	82.4%
アメリカ	141,472	5.4%
国情報不明	29,440	1.1%
フランス	24,766	0.9%
ブラジル	22,856	0.9%
イギリス	19,434	0.7%
中国	16,377	0.6%

(c) https://github.com/selwin/python-user-agents

4.2.2 OS 情報

アクセス元の端末の OS について，Python の user_agents モジュールを利用して解析を行った(c). 計 15 種類の OS と OS の不明な User-Agent が存在した．OS が不明なものは Other と記載されている. そのうち上位 6 件を表 3 に示す. クラウド型 CAPTCHA の特性上，CAPTCHA の回答を送信するアクセスの User-Agent はアクセス先 web サイトのサーバのものが記録されるため，エンドユーザの利用 OS を判別することができない．そのようなアクセスの OS 情報が Other と記載されていると考えられる.

表 3 利用 OS 情報の内訳

OS	件数	比率
Windows	1,025,338	39.2%
iOS	539,933	20.7%
Android	476,536	18.2%
Other	303,144	11.6%
Linux	85,443	3.3%
Mac OS X	84,878	3.2%

4.2.3 アクセス時刻

データセットにおけるアクセス時刻の分布を図 2 に示す. アクセス時刻は分単位を切り捨てて日本標準時を基準としている．4.2.1 節から分かるように，日本からのアクセスが多数となっているため，日本における深夜から早朝の時間はアクセス数が顕著に少なくなっている.

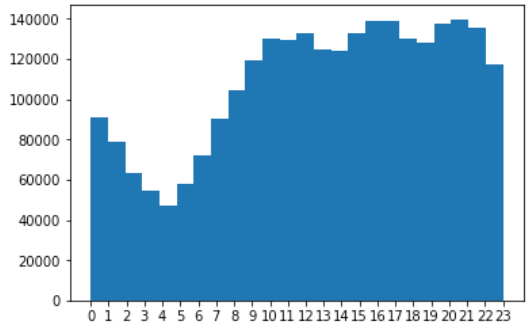

図 2 アクセス時刻の分布(日本標準時)

4.2.4 その他

本データセットにおいて，DNS の逆引き情報が存在する IP アドレスからのアクセスは全体の 86%であり，高い割合を占めている．また，利用ブラウザは Firefox, Mobile Safari など，モバイル端末のものと PC 端末のものがほぼ均等に広く分布しており．全体のモバイル端末の割合は 37%であった．また，Cookie がセットされたアクセスの割合は 25%であった.

4.3 ボット利用アクセスの特徴

本節では, 3.1 節の手法を用いて過去の検知実績より付与したボット判別フラグにより，ボット利用のアクセスである可能性が高いと判断されたアクセスデータの持つ特徴について述べる. これらのアクセスの傾向を分析することで，データ全体の傾向との違いを把握し，ボット判別フラグの予測に有用なボット特徴の抽出を行う．ボット判別フラグが付与されたデータは 6,762 件でありデータセット全体の 0.26%程度である．また，ボット判別フラグは人間の手で作成した IP アドレスのブラックリストから機械的に付与したものであり，ボット判別フラグのついたアクセス以外にもボット利用のアクセスがデータセット内に存在することが考えられる.

4.3.1 ボット利用アクセスの国情報

ボット利用アクセスにおけるアクセス元の国情報を表 4 に示す．国情報が不明となっているものは調査の結果アメリカからのアクセスであると判明している．今までの検知の実績において，発見できているボット利用のアクセスはアメリカの特定 ISP のものが大半である．日本からのボット利用のアクセスはクラウドサービスを経由して行われているものであることが判明している.

表 4 ボット利用アクセスの国情報の内訳

国名	件数
アメリカ	6,746
日本	9
国情報不明	7

4.3.2 ボット利用アクセスの OS 情報

ボット利用アクセスの OS 情報を表 5 に示す．クラウド型 CAPTCHA サービスの特性上エンドユーザの User-Agent が取得できないものについては Other と記載している．ボット利用のアクセスについて，現在検知できているものは Windows を利用したものである.

表 5 ボット利用アクセスの OS 情報の内訳

OS	件数	比率
Windows	5,731	84.8%
Other	1,031	15.2%

4.3.3 ボット利用アクセスの時刻

ボット利用のアクセス時刻の分布を図 3 に示す. 4.3.1 節よりボット利用アクセスのアクセス元はアメリカからが大半であると判明しているため，日本時刻から 13 時間戻した時刻で集計を行った．図 3 の結果から，ボット利用のアクセスは通常のアクセスが集中する時間だけでなく，深夜帯アクセスが少なくなる時間にも一定間隔でアクセスしているものが多く存在すると考えられる.

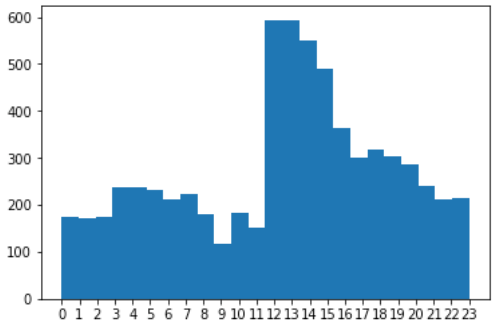

図 3 ボット利用アクセス時刻の分布(ワシントン標準時)

4.3.4 その他のボット特徴

ボット判別フラグの付与されたアクセスにおいて，DNS 逆引き情報が存在する IP アドレスからのアクセスは 1.5%未満であった．また，利用ブラウザのほとんどが Firefox であり，User-Agent 上はモバイル端末からのアクセスは存在しなかった. Cookie がセットされたアクセスは全体の 0.1%未満であった．

5. アルゴリズムの検討と実装

本節では，3 節で示したボット検知システムにおけるボット判別フラグの予測を行うアルゴリズムの実装とその詳細について述べる．

本研究では，ボット判別フラグの予測に教師あり機械学習のアルゴリズムを用いた [7]．教師あり機械学習を選択した理由は，現状のルールベースのボット判別には限界があり，日々変化するボットを見逃してしまう危険があること，また，ルール更新による対処が大きな人的負荷となってしまうことが現状問題となっているためである．

本研究では，ボット判別フラグの付与に人間の手で作成した IP アドレスのブラックリストを用いている. このようなルールベースに近い手法で作成されたフラグの予測に対しては，決定木系のアルゴリズムの適用が妥当であると考えられる．そこで，本研究では XGBoost と LightGBM の 2 つのアルゴリズムを用いた．XGBoost と LightGBM はどちらもメジャーな決定木の勾配ブースティングアルゴリズムである．

5.1 特徴量の選択

本研究におけるボット検知に用いる特徴量の選択は，4.3 節におけるボット利用データの特徴を元に行った．今回用いた特徴量はアクセス元の国情報，OS，曜日，時刻[1h 単位]，PC の User-Agent か，Mobile の User-Agent か，Cookie の有無，DNS 逆引き情報の有無 の 8 つである．

5.2 ハイパーパラメータの設定

本研究に用いたデータセットにおけるボット判別フラグには大きな偏りが存在する．ボット判別フラグの付与されたデータ（以下陽性データ）は全体の 400 分の 1 程度であり，ハイパーパラメータの調整を行わずに予測を行った場合に陽性データの検知を十分に行うことができない．そこで, XGBoost における陽性データに 400 倍の重みを与える．LightGBM においては学習データにおける陽性データの比率を逆数として重みに与えている．また，XGBoost においては学習における評価指標に AUC スコアを指定している．

5.3 データセットの前処理

本研究における予測のターゲットとなるボット判別フラグは最大で/16 単位で指定された IP アドレスのブラックリストにより付与されている. 同一の IP アドレスルールによりボット判別フラグが付与されたアクセスが学習データとテストデータに存在し，そのような判別フラグを検知することはルールベースによる検知と同等である．そのような事態を避けるため，学習データは IPv4 アドレスの 16bit 目が奇数のもの，テストデータは 16bit 目が偶数のものとなるようにデータセットの分割を行った．

5.1 節における特徴量のうち国情報，OS，曜日，時刻はカテゴリ型変数として取り扱い，前処理としてダミー変数処理を施した．また，bool 型の特徴量の欠損値は False として補完を行い, カテゴリ型のものについては Unknown として補完を行った．

6. 実験結果

本節では，5 節に示したアルゴルズムによってボット判別フラグを予測した結果を示し，その評価を行う．5.3 節の手法によるデータ分割の結果, 学習データは 1,450,388 件，テストデータは 1,168,511 件となり. ボット判別フラグの付与されたデータはそれぞれ 3,374 件，3,388 件であった．

6.1 誤分類数の評価

XGBoost, LightGBM の各モデルにおける予測値とその正答数は表 6,7 の通りである．それぞれのモデルにおける正答率は 97%を超える高いスコアとなった．誤分類数から考えられる評価として，XGBoost は LightGBM に比べボット利用フラグのついたデータの誤判別(False Negative)が少なく，ボット特徴を強く学習していることが見て取れる．

表 6 XGBoost の予測値

		正解	
		ボット利用	ボットでない
予測	ボット利用	3,380	29,718
	ボットでない	8	1,135,405

表 7 LightGBM の予測値

		正解	
		ボット利用	ボットでない
予測	ボット利用	3,331	26,251
	ボットでない	57	1,138,872

6.2 ROC - AUC による評価

XGBoost, LightGBM の各モデルの ROC 曲線と AUC スコアを図 4, 5 に示す．各モデルにおける AUC スコアは 0.99 を超える高いスコアとなった．

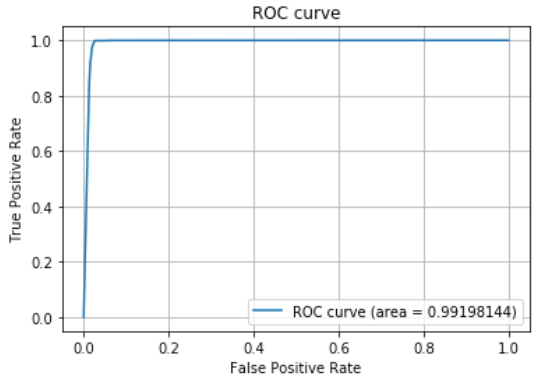

図 4　XGBoost の ROC 曲線と AUC スコア

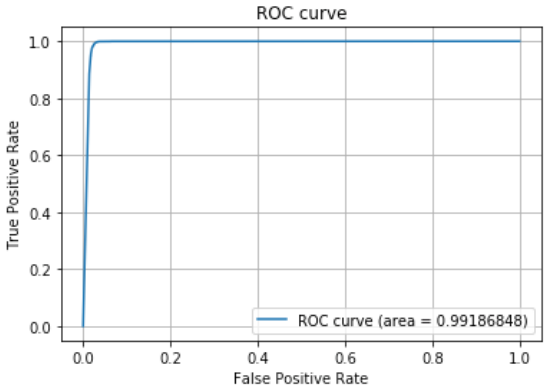

図 5　LightGBM の ROC 曲線と AUC スコア

6.3 Precision - Recall による評価

XGBoost, LightGBM の各モデルの Precision スコアと Recall スコアの関係を図 6, 7 に示す．Precision-Recall の評価指標は ROC に比べて多数派を占める陰性データの誤分類(False Positive)に敏感であると言える [10]. 陰性データと陽性データの比率が大きく偏った今回のデータセットにおいては，Precision-Recall における AUC スコアは ROC における AUC スコアと比較して顕著に低くなっている．

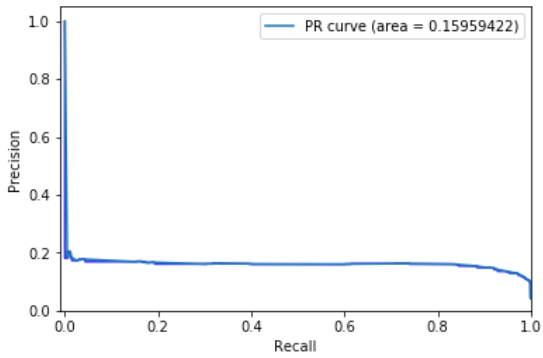

図 6 XGBoost の PR 曲線と AUC スコア

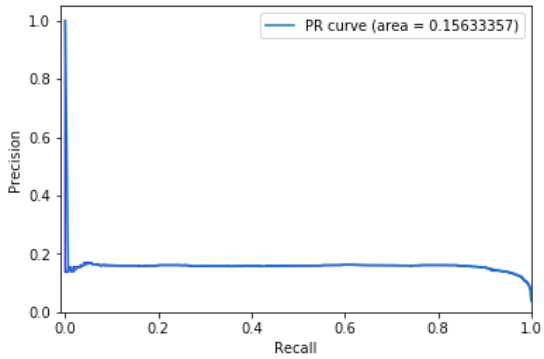

図 7 LightGBM の PR 曲線と AUC スコア

7. ボット予測データの解析

本節では，6.1 節において誤分類の 1 つとして取り扱った，機械学習アルゴリズムがボット利用アクセスであると判定したものの中で，実際にはボット判別フラグのついていないデータ(False Positive データ)について調査を行った．

本研究におけるボット判別フラグが IP アドレスのブラックリストを元にルールベースで付与されている特性上，本来はボット利用アクセスであるがボット判別フラグが付与されていないデータが存在すると考えられる．本研究では，そのようなデータの存在について教師あり機械学習がボット利用アクセスであると予測したアクセスに着目して解析をすることで評価を行った．

7.1 False Positive データの特徴

本節では，各モデルにおける False Positive に該当するアクセスがどのような特徴をもつものであったかの調査の結果を示す．XGBoost における False positive データ数は 29,718 件，LightGBM では 26,251 件であった．そのうち，2 モデルが共通で誤分類したものが 25,301 件存在したため，False Positive データ数の総計は 30,668 件であった．

7.1.1 False Positive データの国情報

False Positive に該当するアクセスデータの国情報を表 8 に示す. 4.2.1 節におけるボット判別フラグの大半がアメリカからのアクセスであったことから，アメリカからのアクセスに対して検知が多く行われていることが見て取れる.

表 8　False Positive データの国情報

国名	件数
アメリカ	30,057
国情報不明	338
日本	273

7.1.2 False Positive データの OS 情報

False Positive に該当するアクセスデータの OS 情報を表 9 に示す. 7.1.1 節と同様, ボット判別フラグの特徴を捉えた結果になっていると考えられる.

表 9　False Positive データの OS 情報

OS	件数
Windows	19,280
Other	11,372
Linux	14
Ubuntu	1
Chrome OS	1

7.1.3 False Positive データのブラウザ

False Positive に該当するアクセスデータのブラウザのうち, 代表的なものを表 10 に示す. 全体のデータセットにおける調査においては Mobile Safari のようなモバイル端末のブラウザが 37%程度存在したが，False Positive データにおいてはほぼ全てが PC 端末からと考えられる.

表 10　False Positive データのブラウザ情報

ブラウザ	件数
Firefox	17,237
Chrome	683
StatusCakebot	429
bingbot	10
AdsBot-Google	4

7.2 考察

7.1 節の調査から, 本研究において作成したモデルによる False Positive データは, 4.3 節で考察したボット利用データの特徴と類似しているものが非常に多いことがわかる．これは，教師あり機械学習モデルがボット判別フラグの特徴をよく学習しているためと考えられる.

これらのデータの ISP 情報ごとに，CAPTCHA のログを解析した．その結果，2,383 件のうち 2 件の ISP において，ログ集計期間外のログから同様のボットによる攻撃の疑いが強いアクセスが発見された．しかしながら，アクセス数など他要素との相関関係については，本論文の計算結果だけでは断定はできないことや，ISP 単位の評価としては正当なアクセスも存在するため，汎用的な評価には至っておらず，他のデータやログなどを用いて検討を続ける必要があることがわかった.

8. まとめ

本研究では，商用サービスとして広く利用されているクラウド型 CAPTCHA サービスである Capy パズル CAPTCHA の実際のアクセスログを用い，教師あり機械学習を用いてボット判別フラグの予測を行うことで，ボット利用アクセスの検知をするとともに，その False Positive データからの未知のボット利用アクセスの検知を試みた.

ボット判別フラグは過去のボット検知の事例から得た IP アドレスのブラックリストを用いて機械的にフィルタリングを行うことで付与した.

また，データセットのアクセスの傾向とボット利用アクセスのアクセスの傾向に関して調査を行った．さらには本データセットにおけるボット利用アクセスの特徴に関して考察を行い，それらの結果を元に機械学習の特徴選択とデータの前処理を行った．ボット判別フラグの予測は教師あり機械学習モデルである XGBoost と LightGBM を用い，それぞれの予測結果について複数の評価指標で評価を行った.

さらには，予測結果においてボット利用アクセスであると判別されたボット判別フラグ無しデータ(False Positive データ)に関して調査を行い，これまでの検知とは別の期間のログデータにおける特定 ISP からのボット利用のアクセスを新たに発見した.

今後の課題として，得られた False Positive データに関して，さらに期間を広げてのアクセス分析や，他の事例から得られる新規のボット判別フラグ付きデータを元にアルゴリズムの再構築を行い，比較することでアルゴリズムの検討を引き続き行う.

また，精度向上のために，データセットに対して教師なしの機械学習アルゴリズムを適用することで，ボット利用データと人間の利用データの分割を行い，未知のボットの検知を試みる手法についても検討を行っている.

参考文献

[1] A. Ramachandran, N. Feamster, and D. Dagon, “Revealing botnet membership using DNSBL counter-intelligence,” in Proceedings of the 2nd Conference on Steps to Reducing Unwanted Traffic on the Internet Workshop (SRUTI ’06), vol. 2, p. 8 (2006).

[2] E. Biglar Beigi, H. Hadian Jazi, N. Stakhanova and A. A. Ghorbani, "Towards effective feature selection in machine learning-based botnet detection approaches," *2014 IEEE Conference on Communications and Network Security*, pp. 247-255(2014).

[3] ebookjapan, “不正ログインに関する最終報告”, https://www.ebookjapan.jp/ebj/information/20130405_access.asp

[4] Fidas, C. A., Voyiatzis, A. G. and Avouris, N. M.: On the Necessity of User-friendly CAPTCHA, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’11, ACM, pp. 2623–2626 (2011).

[5] Hernåndez-Castro, C. J., R-Moreno, M. D. and Bar- rero, D. F.: Side-Channel Attack against the Capy HIP, 2014 Fifth International Conference on Emerging Security Technologies, pp. 99–104 (2014).

[6] Hernåndez-Castro, C. J., R-Moreno, M. D. and Barrero, D. F.: Using JPEG to Measure Image Continuity and Break Capy and Other Puzzle CAPTCHAs, IEEE Internet Computing, Vol. 19, No. 6, pp. 46–53 (2015).

[7] Kotsiantis, Sotiris B., I. Zaharakis, and P. Pintelas. "Supervised machine learning: A review of classification techniques." *Emerging artificial intelligence applications in computer engineering* 160 pp.3-24(2007).

[8] MaxMind,Inc. “GeoIP® Databases & Services: Industry Leading IP Intelligence”, https://www.maxmind.com/en/geoip2-services-and-databases

[9] Powell, B. M., Singh, R., Vatsa, M. and Noore, A.: Poster: Adaptcha: An Adaptive CAPTCHA for Improved User Experience, system, Vol. 4, p. 6.

[10] Saito T, Rehmsmeier M, “The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets”, PLOS ONE 10(3): e0118432(2015).

[11] Sivakorn, S., Polakis, J. and Keromytis, A. D.: I’m not a human: Breaking the Google reCAPTCHA， Black Hat (2016).

[12] T.-F. Yen et al., “Beehive: Large-scale log analysis for detecting suspicious activity in enterprise networks,” in Proc. ACM Annu. Comput. Security Appl. Conf., pp. 199–208(2013).

[13] Traore, I., Woungang, I., Obaidat, M. S., Nakkabi, Y. and Lai, I.: Combining Mouse and Keystroke Dynamics Biometrics for Risk-Based Authentication in Web Environments, 2012 Fourth International Conference on Digital Home, pp. 138–145 (2012).

[14] von Ahn, L., Maurer, B., McMillen, C., Abraham, D. and Blum, M.: reCAPTCHA: Human-Based Character Recognition via Web Security Measures, Science, Vol. 321, No. 5895, pp. 1465–1468 (2008).

[15] von Ahn, L., Blum, M. and Langford, J.: Telling Humans and Computers Apart Automatically, Commun. ACM, Vol. 47, No. 2, pp. 56–60 (2004).

[16] Yan, J. and El Ahmad, A. S.: Usability of CAPTCHAs or Usability Issues in CAPTCHA Design, Proceedings of the 4th Symposium on Usable Privacy and Security, SOUPS ’08, New York, NY, USA, ACM, pp. 44–52 (2008).