5ZA-08

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出

荒井　毅 [1,†1]　岡部　寿男 [1,†2]　岡田　満雄 [2,†3]　渡辺　孝信 [2,†4]　松本　悦宜 [2,†5]
京都大学 [1]　Capy 株式会社 [2]

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，ネットバンキングや仮想通貨口座からの不正送金事案が挙げられる.

不正アクセスの手口としてボットを用いたパスワードの推測や総当たり法を用いた攻撃が挙げられる. 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大量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．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 [6]．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 [2],[3],[5]．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 [1].

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 [4].

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.

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 Inc.の提供するクラウド型 CAPTCHA サービスである*1．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.

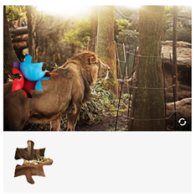

図 1 パズル CAPTCHA の認証画面*2

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 の難度を高くし，ボットによるアクセスを防ぐことである.

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，CAPTCHA の成否を用いる. 本システムでは, CAPTCHA に失敗し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 CAPTCHA 出題時点での情報から CAPTCHA の成否の予測を行い，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

Detection of Bots in CAPTCHA as Cloud Service Utilizing Machine Learning
†1 TSUYOSHI ARAI
†2 YASUO OKABE
†3 MITSUO OKADA
†4 TAKANOBU WATANABE
†5 YOSHINORI MATSUMOTO
1 Kyoto University
2 Capy Japan Inc.
*1 https://www.capy.me/jp/products/puzzle¥_captcha/
*2 https://www.capy.me/products/

るとの判定を行う．実際に CAPTCHA に成功するようなボットは存在すると考えられるが, 本研究では, CAPTCHA に失敗するようなボットを確実に検知することを目標とし，このような分類を行う．

データセットに含まれる具体的なボット利用の危険があるアクセスを表 1 に示す．表 1 にある 2 つのアクセスログは同一ユーザと考えられるが，CAPTCHA 回答の時間，アクセス間隔が非常に短いといった特徴がある．しかし，CAPTCHA が失敗しているアクセスの中には自明な特徴を持たないものも多く存在する．

表 1: ボット利用の疑いのあるログの例

パズル回答時刻	画像取得時刻	IP アドレス
1502390076 1502390087	1502390075 1502390087	A.B.54.239 A.B.54.239

User Agent	Accept Language	パズル識別子
…5.0(Windows… …5.0(Windows…	ja, en-US;... ja, en-US;…	/qXhHb9O… /qXhHb9O…

セッション識別子	回答の軌跡	結果
3c0IIFk4z… WNn2+m…	27,T,0xax84x0xax… 55,F,0xax8ex0xkx…	incorrect-answer incorrect-answer

3.2 ボット検知に利用できる情報

本研究では 2.1 節で挙げたパズル CAPTCHA サービスの実際のユーザの試行ログ 13,808,070 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 CAPTCHA 出題時点の成否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，User Agent, Accept Language，パズル ID，セッション ID の 6 つである．また，これらの情報からアクセス元の国名, 利用 ISP の情報などを得ることができる. また，ブラウザのフィンガープリントのような情報もボット検知に応用することが可能である．

4. ボット検知に用いるアルゴリズム

提案するシステムでは，CAPTCHA に失敗するようなアクセスを事前に予測することを目的とする．システムへの入力は 3.2 節に示した CAPTCHA 出題前に得られる情報であり，出力は CAPTCHA の成否である．

実際のシステムにおいては，IP アドレスごとの試行回数や，User Agent と ISP の関係性など，統計データや複数情報の組み合わせを入力として扱うことを検討する．

CAPTCHA の性質から，CAPTCHA に失敗するようなアクセスを誤って成功と判断することには，実用上大きなリスクがある．その一方で，CAPTCHA に成功するアクセスを誤って失敗と判断することのリスクは比較的大きくない．そのため，システムの要件として多少の誤検知は許容できるが，誤通過の可能性は限りなく低くする必要がある．

CAPTCHA に失敗するアクセスの判定について，自明な特徴をもつアクセスが少ないことから，ルールベースでの判定は難しい. また, データセットの量が十分大きいため，ボット検知システムの実装には，教師あり機械学習を用いる．CAPTCHA の成否を教師ラベルとして用い，時系列に沿って教師データとテストデータを分割する．利用するアルゴリズムについてはロジスティック回帰，ランダムフォレスト, ニューラルネットなどの複数の手法で試行を行い，より良い精度で識別が行える手法を検討する．

5. まとめ

本研究では，クラウド型 CAPTCHA サービスのユーザの利便性の向上および認証精度の向上を目的として，ボット検知システムを用いた CAPTCHA の難度変更を提案した．具体的には, IP アドレスから得られる回線や国の情報, User Agent, アクセス時刻などを用いて CAPTCHA の成否を出題前に予測することを試みる．ボット検知システムの実装には複数の機械学習アルゴリズムを用いて比較検討を行う．本研究の今後の課題として，まずは CAPTCHA の成否識別精度の向上が挙げられる．さらには CAPTCHA を解く挙動の人間らしさなどに着目し，ボット利用のラベル付けを行うことを検討している．このようなラベルを作成することで，CAPTCHA の成否で分類するよりさらに高度なボット利用の分類が可能である．そのように作成したラベルでより高度なボット利用の検知を行うことを本研究の最終的な目標としている．

参考文献

[1] 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).
[2] Hern ́andez-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).
[3] Hern ́andez-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).
[4] Powell, B. M., Singh, R., Vatsa, M. and Noore, A.: Poster: Adaptcha: An Adaptive CAPTCHA for Improved User Experience, system, Vol. 4, p. 6.
[5] Sivakorn, S., Polakis, J. and Keromytis, A. D.: I'm not a human: Breaking the Google reCAPTCHA, Black Hat (2016).
[6] von Ahn, L., Maurer, B., McMillen, C., Abraham, D. and Blum, M.: reCAPTCHA: Human-Based Charac- ter Recognition via Web Security Measures, Science, Vol. 321, No. 5895, pp. 1465–1468 (2008).