4W-02

Deep Learningの学習曲線を用いたCAPTCHA

坂井　麻守†　　森田　光†
†神奈川大学大学院工学研究科

1　はじめに

ネットからBotのように自動的なコンピュータプログラムによる攻撃が可能となり、CAPTCHA画像によるコンピュータの機械的なアクセスを排除する手段も脅威にさらされている．

将来の、コンピュータのDeep Learningなどによる学習能力の大幅な向上を見据え、CAPTCHAで使われる画像が攻撃者の学習材料となる脅威に対して、著者らは、Deep Learningの学習に伴う攻撃能力の向上として検討した．Deep Learningに大量の画像を与えれば、ほとんど人間並の識別能力になるので、攻撃者に対する画像の与え方が重要になることが明らかになった．

そこで著者らは、Deep Learningの汎化作用を表す学習曲線を用いれば、攻撃者の学習の進行を予想できるので、CAPTCHA画像を入れ替えるタイミングを設定できる方法を提案した．

2　先行研究

人間の高度な認知能力を用いるCAPTCHAが多数登場してきた．Googleが開発したreCAPTCHA[1]は、9つの画像の中から例と同じ種類の画像を選択する画像選択型CAPTCHAである．他の画像選択型には、佐野らによる3次元のCAPTCHA[2]もある．これは3次元オブジェクトの向きをユーザに選択させるのである．画像とは異なるアプローチとして可児らによる4コマ漫画の並び順を解答させるCAPTCHA[3]がある．

新しいCAPTCHAが登場する度に、それを突破する新たな攻撃手法が表れ、バリエーションが増加している．その中でも体系的に全体像を、確率的アプローチで考察しようとする試みがあった[4]．また、Starkらは、Active Deep Learningという独自のアルゴリズムを考案し、約80%以上の確率で文字判別型CAPTCHAを認識できることを示している[5]．

3　提案法

Deep Learningの性能を調べるために、著者らはまず画像に対する予備実験を行い、それを学習曲線の形にまとめることを試みる．

実験結果は各カテゴリーに同数の学習画像を与え、最小自乗法による、シグモイド関数による回帰曲線を次の近似式で与える：

$$p_{DL}(x) = \frac{1}{1 + e^{-bx+d}} \tag{1}$$

ここで、$p_{DL}(x)$は、学習用の既知画像数xに対して画像認識の成功確率の学習曲線を示す．なお、論文[6]を単純化して、本式を導いた．

カテゴリ数はc、総学習画像数はx以下（カテゴリ毎ではx/c）とした．

次に、画像認識を複数組み合わせるCAPTCHAの適応対象を次のようにした（図1）．このアクセス制御方法は比較的実用的なCAPTCHA例であり、アクセス者は、右上に例示された画像を見て、9つの個々の画像から同種類の画像を選択する．

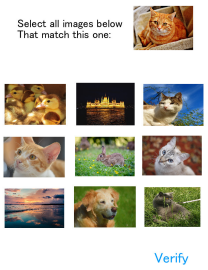

図1: 考察対象とするCAPTCHA認証画面

ここで、図1におけるDeep Learningへの攻撃の成功確率を考え、CAPTCHAを構成する画像の与え方を提案する．

例示画像と同じカテゴリの画像がi枚の半数程度あるとし、成功したときその半数が同一カテゴリとしての既知の学習画像として、攻撃者に使われることになる．ただし、例示画像と同じカテゴリでない残りの画像は、未知とする．

以上の前提により、攻撃者による個々の画像の推定の成功確率$p_{DL}(x)$から、$i = 9$として、期待値E_{DL}が導かれる．

$$E_{DL} = \sum_{t=1}^{T} \Delta E_{DL}(t)$$

ここでは、前述の無脳アプローチと同様に、トライ回

CAPTCHA by using a Sigmoid Curve of Deep Learning
†Mashu SAKAI　†Hikaru MORITA
†Graduate School of Engineering, Kanagawa University

Copyright ©2017 Information Processing Society of Japan.
All Rights Reserved.

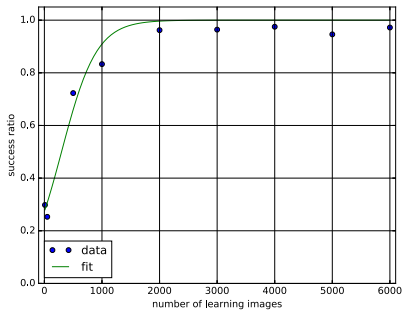

図 2: Deep Learning の学習曲線

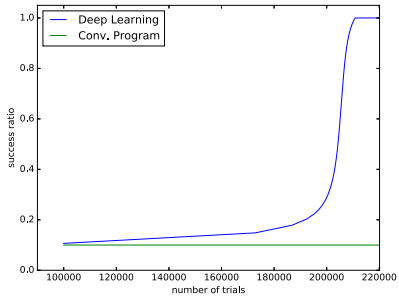

図 3: CAPTCHA に対する攻撃の成功確率 $p_{DL}(x)$ $(i = 5)$

数 T の間隔を ΔT とし、既知の画像の数 $x(t)$ を更新（増加）するとし、

$$\Delta E_{DL} = \Delta T\ p_{DL}(x)^i \geqq 1 \text{ になったとき}$$

$$x \leftarrow x + \frac{i}{2}\left(1 - \frac{x}{n}\right)$$

で更新するとともに $\Delta T = 0$ にリセットする．

更新するトライ回数の間隔は

$$\Delta T = p_{DL}(x)^{-i}$$

なので、ここでは $p_{DL}(x)$ が小さいので ΔT は大きくなる．

4 評価と考察

実験結果をうけ、図 2 に深層学習の学習曲線を示す．学習の画像枚数 x による識別成功率を実験し成功率を測定した．なお、得られた近似式 (1) においては、$b = 3.28 \times 10^{-3}, d = 0.979$ であり、$c = 10, x = 10000$ 以下（カテゴリ毎では 1000）とした．

また、$i = 5$ について、トライ回数に対して個々の画像に対する推定の成功確率を図 3 に示す．

仮に画像 10000 枚であっても、評価実験 1 回あたりの Deep Learning で学習に要する時間は表 1 の環境で 5 分程度なので、1 つの学習用の画像を得るのに、最初は疎であっても 10^{-i} の確率となり、つまり 10^i トライ回数から 1 回分の既知画像を獲得し、以下既知画像を増大させていく．

表 1: Deep Learning の学習環境

CPU	Core i7 6950X Extrem Edition 3.0GHz
GPU	GeForce GTX 1080 (2WAY-SLI)
メモリ	64GB

Deep Learning の場合、1 カテゴリあたり 1000 枚、全体で 10000 枚の学習画像があればほとんど 100% となるので、Deep Learning させる前の効率の悪い学習にランダムなトライアルだけが行われるとすると、10^{i+4} トライが必要となる．

したがって、自分のもつ CAPTCHA が外部からどのくらいアクセスされているかを目安に、全く別種の学習対象とするカテゴリに変更すれば、安全性が確保されることになる．

5 まとめ

Bot 攻撃が起きた場合、外部に提示した画像が解析されることを想定し、Deep Learning の攻撃として、主に画像を扱う CAPTCHA に対して、その汎化の脅威とその対策について考察した．CAPTCHA で守る側としては、画像の種類、対象とするカテゴリの総入れ替えのタイミングなどに関するライフタイムを、外部からの攻撃のトライアル回数から推計する方法も提案した．

参考文献

[1] Google Inc, “reCAPTCHA: Easy on Humans, Hard on Bots,” Google Inc,https://www.google.com/recaptcha/intro/index.html, 参照 Aug.7,2016.

[2] 佐野絢音, 藤田真浩, 西垣正勝, “Spatiometric 型メンタルローテーション CAPTCHA の提案,” 暗号と情報セキュリティシンポジウム 2016（SCIS2016）予稿集, 3C2-1, 2016.

[3] 可児潤也, 鈴木徳一郎, 上原章敬, 山本匠, 西垣正勝, “4コマ漫画 CAPTCHA,” 情処学論, vol.54, no.9, pp.2232-2243, 2013.

[4] M. Chew, J. D. Tygar, “Image Recognition CHAPTCHAs,” *Information Security, 7th International Conference, ISC 2004*, pp.268-279, 2004.

[5] F. Stark, C. Hazırbaş, R. Triebel, and D. Cremers, “CAPTCHA recognition with active deep learning,” *GCPR Workshop on New Challenges in Neural Computation*, Aachen, Germany 2015.

[6] 坂井麻守, 森田光, “Deep Learning の学習曲線を用いた CAPTCHA または画像パスワード認証,” 暗号と情報セキュリティシンポジウム 2017（SCIS2017）予稿集, 3B4-3, 2017.

Copyright ©2017 Information Processing Society of Japan.
All Rights Reserved.