Using Machine Learning to Break Visual Human Interaction Proofs (HIPs)

Kumar Chellapilla
kumarc@microsoft.com
Microsoft Research
One Microsoft Way
Redmond, WA 98052

Patrice Y. Simard
patrice@microsoft.com
Microsoft Research
One Microsoft Way
Redmond, WA 98052

1 Introduction

Machine learning is often used to automatically solve human tasks. In this paper, the author uses it to differentiate between a human and a computer. Human Interactive Proofs (HIPs) or Completely Automated Public Turing Tests to Tell Computers and Humans Apart (CAPTCHAs) are automated tests that humans can pass, but computer programs can't pass. These tests enable the construction of automatic filters that can be used to prevent automated scripts from utilizing services intended for humans. These tests are crucial these days; in the era of computer security, as if these are not there then now and then our email inbox will be spammed by some random bot.

2 Human Interactive Proofs (HIPs)

Over the last five years, efforts to defend Web services against abuse by programs ('bots') have led to this new family of security protocols - HIPs that can distinguish between humans and machine users. Most HIPs are pure recognition tasks that can be easily broken using machine learning. Harder HIPs are built using the combination of recognition and segmentation tasks. Hence segmentation is the most effective way to confuse the machine learning algorithms. Yet some of these HIPs are harder than others and could be made even harder by identifying the recognition and segmentation and emphasizing the latter.

Although HIPs are quite effective to tell computers and humans apart, they may fail too. HIPs are text-based systems and are interesting and useful for several reasons:

- Text is a new medium
- Blind people cannot pass a visual CAPTCHA test
- Text-based interface would be more convenient in certain situations, such as console on a Unix system.

Specifically seven different HIPs namely Mailblocks, MSN, Ticketmaster, Yahoo, Yahoo v2, Register, and Google are discussed.The methods for solving six of them are discussed.

3 Pros and Cons of HIPs

3.1 Pros

- Blocks automated increased usage of services that are intended for humans.
- Makes online activity safer.
- Differenciates between a humans and computers.
- Easy to implement.
- It is a cryptographic protocol whose underlying hardness assumption is based on an AI problem.
- Prevent Dictionary Attacks.

3.2 Cons

- Poor User experience.
- It can only limit spam but unable to prevent spam completely.
- They are considered an annoyance by most people.
- Sometimes they are illegible, as not all HIPs are readable to humans.

4 Using machine learning to break HIPs

The author says that a general method to solve all the existing HIPs can't be specified. He explained, how to approach the solution of some common HIPs like in Mailblocks, Register, Yahoo / EZ-Gimpy, Ticketmaster, Yahoo version 2, Google/GMail. One of them is given as example below:

4.1 Mailblocks

To solve the HIP,

- Select the red channel, binarize and erode it.
- Extract the largest connected components (CCs) and breakup CCs that are too large into two or three adjacent CCs.
- Vertically overlapping half character size CCs are merged. The resulting rough segmentation works most of the time. Here is an example:

The end-to-end success rate is 88.8% for segmentation, 95.9% for recognition (given correct

segmentation) and 66.2 % in total.
Similarly the methods to solve the other five HIPs are discussed in the paper by the author.

5 Technical Disabilities

Many technical issues that have been systematically studied by the document image analysis (DIA) community is relevant to the HIP research.

- All commercial uses of HIPs known to us exploit the gap between computer ability and human ability. But the problem is asymmetric: we want to exploit things which humans can do, but which computers cannot do. Is the complementary problem, IPC (Interactive Proof of being a Computer), of any interest?

- Computer security people have dealt with 'Identification and Authentication' issues for a long time. The general wisdom in that field is that authentication is based on one or more of three factors:

  - Something you know (eg. a password)
  - Something you have (eg. a badge)
  - Something you are (eg. a fingerprint)

  Are we proposing a new authentication factor

  - Something you can do (eg recognize beauty)?

Even if we consider fingerprint as the authentication factor, there are some tools that can generate random fingerprints; these tools exist to test some machine learning algorithms.

6 Technical Suggestions

6.1 Replacing HIPs with another system

Apart from HIPs there is another preliminary system that can be developed that exploits human's ability to analyze the meaning in sentences.The computer administering the test selects one word from the source text. It replaces the chosen word with another "bogus" word selected randomly - either from a set of words of the same part of speech as the chosen word, or, more generally, according to some probability distribution. To pass the test, the user must identify the bogus word, this task would be really hard for computer.

Problems with this system are:

- It is possible for a computer to pass the test when the bogus word is chosen based on part of speech.

- We must find a secure way to obtain a coherent sentence in which to put the bogus word.

6.2 Building better/harder HIPs

Harder HIPs can be built by making the segmentation difficult. The idea is that the additional arcs are themselves good candidates for false characters.

The previous segmentation attacks would fail on this HIP. Despite the apparent difficulty of these HIPs, humans are far better than computers at segmentation.

6.3 Using BaffleText

- BaffleText is a CAPTCHA which uses non-English pronounceable character strings to defend against dictionary attacks, and Gestalt-motivated image-masking degradations to defend against image restoration attacks.

7 Future Research

The security level of HIPs can be enhanced in future by turning it more intellectual and based on artificial intelligence problem which require mental ability to solve the HIPs. HIPs should be easy, effective, logical and non-breakable that can be solved by human but almost impossible for robots.