Software-Agenten im Internet

Florian Pircher
Technologische Fachoberschule
Oberschulzentrum Fallmerayer
Brixen, Italien

19. März 2016

Abstract

Menschen sind schon lange nicht mehr die einzigen Nutzer des Internets. Beinahe 50 % aller Webseiten-Aufrufe werden von autonomer Software getätigt. Diese auch als Bots bezeichnete Softwares agieren in den Schatten des Netzes. Unbemerkt indexieren sie Webseiten, verbreiten Spam, legen gefälscht Profile an oder versuchen in Datenbanken einzubrechen.
Hindernisse wie CAPTCHAs oder Honeypots galten bislang als effektive Gegenmittel, allerdings verhilft der rapide Fortschritt im Feld der Künstlichen Intelligenz modernen Bots auch derartige Barrieren zu durchbrechen. Im Bereich E-Mail-Spam findet ein unablässiger Kampf zwischen Spam-Bots und Klassifizierungsalgorithmen statt.
Im ersten Teil dieser Arbeit wird das Verhalten und Vermögen von Software-Agenten untersucht. Schwerpunkte bilden dabei die Themen Spam und Sicherheit. Der zweite Teil beschreibt die Anwendung des gewonnenen Wissens in Form der Entwicklung eines eigenen Bots der autonom durch das World Wide Web navigiert und anhand von maschinellem Lernen versucht CAPTCHAs zu knacken.

Inhaltsverzeichnis

1 Einleitung

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

2 Überblick

2.1 Software-Agenten

2.2 Definition

Ein Software-Agent ist eine Computer-Applikation, welche autonom vordefinierte Ziele verfolgt. Alternative Namen umfassen: *Bot*, *autonome Software*, oder im Bereich der Künstlichen Intelligenz auch *KI* oder *A.I.*

Es gibt verschiedene Arten von Software-Agenten die verschiedenste Aufgaben übernehmen. Diese Arbeit befasst sich mit Bots, die das Internet als *Lebensraum* nutzen. Dazu zählen unter anderen:

Webcrawler/Spider Bots, welche Webseiten im *World Wide Web* indexieren und/oder analysieren. Die bekanntesten Vertreter sind *Web Spiders*[1]: Agenten, welche von Suchmaschinen eingesetzt werden um neue Webseiten zu finden und den bestehenden Suchindex zu aktualisieren. Webcrawler können auch genutzt werden um E-Mail-Adressen aufzuspüren, beispielsweise für den Versand von Spam-Nachrichten.

Spambot Diese Software-Agenten versenden Spam-Nachrichten. Da ihre Spam-Nachrichten jedoch meist auf sie zurückverfolgt werden können (zum Beispiel mittels Absender-Adresse oder Benutzername) sehen sich Spambots genötigt massenhaft Accounts anzulegen um nicht blockiert zu werden. Im Kapitel 3.3.1 wird eine Methode vorgestellt, um dies zu vermeiden und in 4.1 wird ein Versuch beschrieben diese Methode zu überlisten.

Brute-Force-Bots Brute-Force (englisch für *rohe Gewalt*) beschreibt den Akt, bei welchem eine Software sämtliche möglichen Optionen erproben um in fremdes System einzudringen. In der Regel besteht dies darin Accounts zu entwenden, indem eine Kombination von Benutzernamen und Passwörtern ausprobiert werden. Der Benutzername kann beispielsweise in Form einer E-Mail-Adresse bezogen worden sein, welche ein Webcrawler auftreiben hat. Die Passwörter, die erprobt werden, entstammen einer Tabelle von geläufigen Passwörtern.

[1] Spetka, *The TkWWW Robot: Beyond Browsing*.

3 Beobachten und Verstehen

3.1 Spam

3.1.1 Definition

Unerwünschte Nachrichten im Internet, größtenteils Werbeangebote, Phishing-Attacken oder Übermittler von Schadsoftware, werden als Spam bezeichnet. Spam-Nachrichten stellen per definitionem eine leidige Kommunikationsform dar. Die alltäglichste Art, namhaft durch ihre Omnipräsenz im Leben mit dem Internet, stellt der E-Mail-Spam dar.

Der Begriff fand noch vor dem Erscheinen des *World Wide Web* anklang, beschrieb damals jedoch eine Flut an unerwünschten Nachrichten.[2] Erste Aufkommen des Phänomens wurden im ARPANET[3] und im USENET beobachtet.

Der Name *Spam* entspringt dem gleichnamigen Dosenfleisch, welches im Zweiten Weltkrieg in großen Mengen an Soldaten verteilt wurde.[4] Der Markenname Spam® (kurz für *Spiced Ham*) setzte sich auf diese Weise rasch im Vereinigten Königreich als Deonym für Frühstücksfleisch durch. Die britische Comedy-Gruppe Monty Python griff im Sketch *Spam*[5] der BBC Serie *Monty Python's Flying Circus* die Allgegenwärtigkeit des Fleischprodukts auf, weshalb dieses bis heute als Symbol für einen unerwünschten Überschuss fungiert.

Mit dem Begriff *Spam* werden heutzutage diverse Arten von Nachrichten bezeichnet. Werbenachrichten sind das klassische Beispiel einer Spam-Nachricht. Sie bieten zumeist dubiose Produkte oder Dienstleistungen an.

[2] Templeton, *Origin of the term "spam" to mean net abuse*.
[3] Postel, *On the Junk Mail Problem*.
[4] Longmate, ≫How We Lived Then: A History of Everyday Life During the Second World War≪.
[5] Monty Python, *Spam - Monty Python's The Flying Circus*.

3.2 Phishing

3.2.1 Definition

Das Bestreben persönliche Daten im Internet durch ein gezielt verfälschtes Auftreten abzugreifen wird als Phishing (aus dem Englischen *fishing*, für *fischen* oder *angeln*) bezeichnet. In der Regel handelt es sich dabei um E-Mail-Nachrichten oder Webseiten, welche Informationen wie Name, Adresse, Benutzername/Passwort oder Kreditkarten-Daten anfordern. Dem Empfänger wird dargelegt, sein Account auf einer Webseite würde in Gefahr schweben, sollten die geforderten Informationen nicht rasch übergeben werden. Dieser künstliche Drang soll den Adressaten der Phishing-Nachricht davon abhalten seinen Verstand zu benutzen und stattdessen eine panische, reflexartige Handlung auslösen. Diese Handlung wird vom Betrüger geschickt dirigiert: dem Opfer wird ein großer Button oder ein einfach erreichbarer Link präsentiert, der auf die Webseite des Betrügers führt. Diese Webseite imitiert das Aussehen des Originals, leitet jedoch die vom Benutzer eingetragenen Daten an den Sender der Phishing-Nachricht weiter.

Phishing ist eine besonders elegante Art des Spams, da die Opfer häufig nicht mitbekommen, dass ihre Daten von Dritten abgegriffen werden. Nachdem man die geforderten Daten eingegeben und abgesendet hat, leitet die Webseite des Betrügers das Opfer auf die tatsächliche Webseite weiter.

3.3 Künstliche Intelligenz

3.3.1 CAPTCHA

3.3.2 Honeypot

Als Honeypot (zu Deutsch *Honigtopf*, sinngemäß *Fettnäpfchen*) wird in der Informatik eine Falle bezeichnet, die sich als begehrenswertes Objekt tarnt.

Ein exemplarischer Einsatz ist das von Ordnungshütern durchgeführte Veröffentlichen von scheinbar illegalen Dateien im Internet. Nutzer, welche auf diese Dateien zugreifen, werden vom Honeypot registriert und können daraufhin strafrechtlich verfolgt werden.

Dieselbe Methodik lässt sich bei der Unterscheidung zwischen Bots und Menschen zum Einsatz bringen. Eine geläufige Praxis ist es, online Formulare (zum Beispiel ein Kommentar-Formular) mit einem zusätzlichen Textfeld zu bestücken und dieses visuell zu verbergen. Ein menschlicher Benutzer kann dieses Feld weder erreichen noch wahrnehmen; konträr stoßen Software-Agenten auf das Textfeld – sie nehmen schließlich nur den Programmcode des Formulars und nicht die visuelle Aufmachung wahr – und füllen es aus. Beim Absenden des Formulars kann die Formular-Software überprüfen, ob das Honeypot-Feld ausgefüllt wurde oder nicht; im ersten Falle würde diese Software das Absenden unterbinden.

4 Eigener Software-Agent

4.1 Maschinelles Lernen: CAPTCHA

Literatur

Longmate, Norman. »How We Lived Then: A History of Everyday Life During the Second World War«. In: Random House UK, 1971, S. 142, 159.

Monty Python. *Spam - Monty Python's The Flying Circus*. 2009. URL: https://www.youtube.com/watch?v=M_eYSuPKP3Y (besucht am 27. 01. 2016).

Postel, Jon. *On the Junk Mail Problem*. 1975. URL: http://tools.ietf.org/html/rfc706 (besucht am 05. 02. 2016).

Spetka, Scott. *The TkWWW Robot: Beyond Browsing*. 2001. URL: https://web.archive.org/web/20040903174942/archive.ncsa.uiuc.edu/SDG/IT94/Proceedings/Agents/spetka/spetka.html (besucht am 19. 03. 2016).

Templeton, Brad. *Origin of the term "spam" to mean net abuse*. URL: http://www.templetons.com/brad/spamterm.html (besucht am 05. 02. 2016).