Monika Ermert, Christian Grothoff

Data Mining für den Drohnenkrieg

Lexikon des NSA-Skandals: Skynet

Die Überwachungsprogramme der NSA sammeln unzählige Daten. Diese werden automatisiert ausgewertet, offenbar auch zur Suche nach Zielen für den Drohnenkrieg.

Als der ehemalige NSA-Chef Michael Hayden erklärte, dass die USA „auf der Basis von Metadaten" töten, spielte er vermutlich auf das Programm Skynet an. Eine 20 Folien umfassende interne NSA-Präsentation dazu hatte das Online-Magazin The Intercept erstmals im Mai 2015 öffentlich gemacht. Skynet ähnelt Programmen zum Data Mining im Börsenhandel oder Kunden-Profiling, dient aber der Suche nach Zielen für die US-Drohnen im „Krieg gegen den Terror".

Die Folien zu Skynet basieren auf Daten aus Pakistan, wo der US-Geheimdienst CIA seit 2004 Hunderte Ziele mit Drohnen angegriffen hat. Die pakistanische Regierung hat die US-Regierung wiederholt aufgefordert, diese Angriffe einzustellen. Immerhin werden dabei Menschen ohne Gerichtsverfahren getötet. Die USA jedoch klassifizieren alle Opfer systematisch als „im Kampf getötete Feinde", obwohl es sich bei der Mehrheit nicht einmal um Zielpersonen der USA handelt. Nach welchen Kriterien die bestimmt werden, ist allenfalls bruchstückhaft bekannt. Eines dieser Puzzlestücke ist Skynet.

Datenfutter für Skynet

Das Programm Skynet nutzt Methoden, die im Umgang mit Big Data inzwischen alltäglich sind. Grundlage sind die zahlreichen Metadaten über Individuen, die mit Überwachungsprogrammen wie XKeyscore gesammelt werden. Im Fall von Pakistan, das direkt im Visier des US-Anti-Terrorkampfs steht, betrifft das weite Teile der Bevölkerung.

Insbesondere aus Mobilfunkdaten gewonnene Bewegungsprofile interessieren die NSA-Analysten. Dafür werden die Wege von SIM-Karten verfolgt, dank deren IMSI-Nummern sogar dann, wenn sie in ein anderes Mobiltelefon gesteckt werden. Auch die Hardware-Nummer (ESN/MEID) hilft bei der Verfolgung von Geräten und damit letztlich von Individuen. Zu den gesammelten Rohdaten gehört auch der sogenannte Social Graph, also das Netzwerk an Personen, mit denen der Mobilfunkteilnehmer in Kontakt steht. Darüber lässt sich etwa feststellen, wenn jemand häufig die Geräte wechselt.

Aus den Bewegungsprofilen und den Listen der geführten Gespräche von Millionen Menschen errechnet Skynet deren typische Tagesroutinen: Wer reist mit wem, wann und wohin; wer steht mit wem in Verbindung? Wer übernachtet bei Freunden, wer reist ins Ausland oder zieht dauerhaft an einen anderen Ort? Insgesamt erhebt die NSA laut den Skynet-Folien Daten zu 80 verschiedenen Merkmalen. Diese Werte bilden die Basis für die maschinelle Klassifikation. Sie charakterisieren das Verhalten individueller Bürger und liefern zumindest vermeintliche Erklärungen. Diesem Vorgehen liegt die Annahme zugrunde, dass sich eine potenzielle Zielperson anders verhält als ein normaler Bürger.

Lexikon des NSA-Skandals

XKeyscore	c't 17/15, Seite 134
Tempora	c't 18/15, Seite 72
Fashioncleft	c't 19/15, Seite 66
Prism	c't 22/15, Seite 84

Das andere Schlüsselelement, mit dem Skynet gefüttert wird, sind Daten aus einem Referenzdatensatz. Diese beim maschinellen Lernen als „Ground Truth" (Grundwahrheit) bezeichneten Datensätze geben dem lernenden Algorithmus „richtige" Ergebnisse beispielhaft vor. Im Fall von Skynet klassifizieren die Referenzdaten einzelne Individuen („Selektoren") als „Terroristen" und „Unverdächtige". Solch ein Referenzdatenset zu erzeugen ist schwierig. Denn wer würde im NSA-Fragebogen freiwillig das Kästchen hinter „Planen Sie einen terroristischen Anschlag?" ankreuzen?

Die Dokumente legen nahe, dass die NSA für ihre Modelldaten auf Informationen über bereits als Terroristen eingestufte „Kuriere" terroristischer Gruppen zurückgreift. Diese müssen nicht direkt an Anschlägen oder deren Vorbereitung beteiligt sein, sondern übermitteln beispielsweise Botschaften zwischen Terroristen. Zur Profilbildung im Referenzdatensatz werden sie als die zu identifizierenden Ziele klassifiziert. Der Rest der Bevölkerung wird als „unverdächtig" eingestuft.

Big-Data-Logik

Das wichtigste Credo im Big-Data-Business lautet: Mehr Daten sind immer besser als weniger. Die größere Menge verspricht genauere Entscheidungen. Sinnvolle Schlussfolgerungen aus der Datenmenge zu ziehen ist insbesondere in Pakistan eine echte Herausforderung. Dort werden über 80 Einzelmerkmale auf ein Land mit 190 Millionen Einwohnern angewandt. Zwar kann die NSA statistische Verteilungen plotten; den riesigen Datenhaufen von Hand nach Mustern zu durchforsten, ergibt aber wenig Sinn. An die Stelle menschlicher Analysten tritt in Skynet daher eine Bewertung durch überwachtes maschinelles Lernen: Durch Training mit den Referenzdaten soll der Computer automatisch Muster in den Daten finden, die auf mutmaßliche „Terroristen" hindeuten.

Dazu bekommt ein lernender Algorithmus die Referenzdaten vorgesetzt, in denen alle Personen als „Terrorist" oder „unverdächtig" markiert sind. Anhand der jeweils mit ihnen verknüpften Merkmale („viel unterwegs", „spät am Telefon" etc.) errechnet er einen Klassifizierungsalgorithmus, der anhand dieser Merkmale selbst Zuweisungen treffen kann und jedem Individuum einen numerischen Wert zuordnet. Je stärker die Merkmale denen eines „Terroristen" gleichen, desto höher fällt deren Wert im Idealfall aus. Der lernende Algorithmus, der für die NSA dieses Klassifikationsverfahren berechnen soll, ist vom Typ „Random Forest". Dieser toleriert irrelevante Merkmale in den Eingangsdaten besonders gut und kann daher einfach auf eine Masse mit beliebigen Merkmalen losgelassen werden.

Der Algorithmus bei Random Forest lernt auf der Basis von Entscheidungsbäumen, die aus zufälligen Untermengen der Trainingsdaten konstruiert werden. Bei der Auswertung von Entscheidungsbäumen werden hierarchisch aufeinander aufbauende Wenn-Dann-Beziehungen abgearbeitet, beispielsweise: Telefoniert jemand sehr selten, dann aber nachts und wechselt danach stets den Standort,

weicht dieses Verhalten signifikant von dem anderer Nutzer ab und er wird als „Terrorist“ markiert, sonst nicht. Durch das Mitteln Dutzender oder Hunderter dieser Entscheidungsbäume – und damit das Durchspielen verschiedener Entscheidungspfade – wird versucht, die Genauigkeit der Vorhersage zu verbessern.

Welche Algorithmen bei der NSA die Entscheidungsbäume für Skynet erlernen, geht aus der veröffentlichten Präsentation nicht hervor. Klar ist jedoch, dass die Entscheidungsbäume nach dem Random-Forest-Verfahren kombiniert werden, um schließlich einen Algorithmus zu erhalten, der das Datenfutter mit den 80 unterschiedlichen Merkmalen verarbeitet und numerische Werte für alle Bürger ausspuckt.

Damit der Algorithmus zwischen „Terrorist“ und „harmlos“ unterscheiden kann, muss ein Schwellenwert festgesetzt werden. Bürger, deren Wert über der Schwelle liegt, sind dann „Terroristen“. Diejenigen, deren Wert darunter bleibt, werden als „unverdächtig“ eingestuft. Für ein bekanntes Referenzdatenset kann die NSA den Schwellenwert so festlegen, dass auf jeden Fall ein bestimmter Prozentsatz der von der NSA schon zuvor als Terroristen eingestuften Personen im Raster hängen bleibt. In den vorliegenden Dokumenten setzt die NSA 50 Prozent als Erfolgsquote fest; der Schwellenwert wird also so festgelegt, dass das System jeden zweiten „Terroristen“ findet – die andere Hälfte dann aber eben nicht. Das sind die sogenannten „False Negatives“. Denen stehen „False Positives“ gegenüber: Unschuldige, die fälschlicherweise als „Terroristen“ gekennzeichnet werden. Deren Zahl würde zwar sinken, wenn der Schwellenwert erhöht würde, dann würde sich aber auch die Erkennungsrate von „Terroristen“ verschlechtern – ein Preis, den die NSA nicht zu zahlen bereit scheint.

Nach Abschluss der Kalibrierung füttert die NSA das Auswertungsprogramm mit den Daten der zu analysierenden Bevölkerung. Im Falle Pakistans waren das im Jahr 2007 rund 55 Millionen Telefonnutzer. Diese Zahl wird in der Präsentation genannt, die aus dem Jahr 2012 stammt. Inzwischen dürften es deutlich mehr sein. Die NSA nutzt das bei der Verarbeitung von Cloud-Daten gebräuchliche Big-Data-Werkzeug „Map Reduce“.

Rücksichtslose Berechnungen

Um ihr System zu überprüfen, nutzt die NSA zwei Quellen. Laut der uns vorliegenden Darstellung wird unter die Daten von 100 000 zufällig ausgewählten Bürgern eine Gruppe von sieben Personen gemischt, die die NSA bereits auf der Basis anderer Informationen als „Terroristen“ klassifiziert hat. Der lernende Algorithmus wird an sechs davon trainiert und soll dann den siebten „Terroristen“ finden.

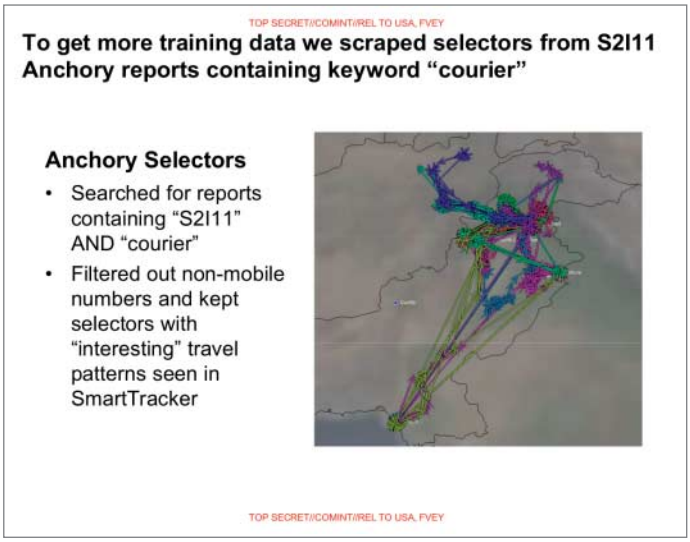
Bei diesem Datensatz kennt die NSA also die – weil bereits identifizierten – Terroristen und kann ausrechnen, wie hoch der Prozentsatz von „False Positives“ ist. Außerdem wird der Algorithmus auf die Datenpunkte von 55 Millionen Bürgern in Pakistan losgelassen, um die Skalierbarkeit zu testen und zu sehen, wie gut der Algorithmus bereits bekannte Verdächtige erkennt. Unklar lassen die geleakten Daten, ob der große Datensatz auch die zuvor fürs Training eingesetzten Daten der 100 000 Bürger enthält.

In jedem Fall verletzt die NSA mit dieser Methodik wissenschaftliche Standards zur Evaluierung von maschinellem Lernen, die mit gutem Grund aufgestellt wurden. Die hier stattfindende Verallgemeinerung der Ergebnisse ist nicht zulässig. So wurden die 100 000 Bürger zwar zufällig ausgewählt, die sieben „Terroristen“ stammen jedoch aus einer bekannten Gruppe. Die Einschränkung auf eine kleine zufällige Untermenge (weniger als 0,1 Prozent der Bevölkerung) reduziert die Dichte des sozialen Netzes der Bürger, wohingegen die „Terroristen“ engmaschig verknüpft bleiben. Die statistische Analyse wäre aber nur dann wissenschaftlich und damit belastbar , wenn die „Terroristen“ über dasselbe Auswahlverfahren in der Gesamtmenge gelandet wären. Das ist aufgrund der geringen Anzahl jedoch nicht praktikabel. Für die Gesamtdaten der damals 55 Millionen überwachten Pakistaner wäre dieses Problem der zufälligen Untermenge nicht so gravierend. Für solch einen Testlauf gibt die NSA-Präsentation aber keinen Prozentsatz der „false positives“ an. Die Erkennungsrate dürfte sich bei einer größeren Gesamtmenge aber verringern.

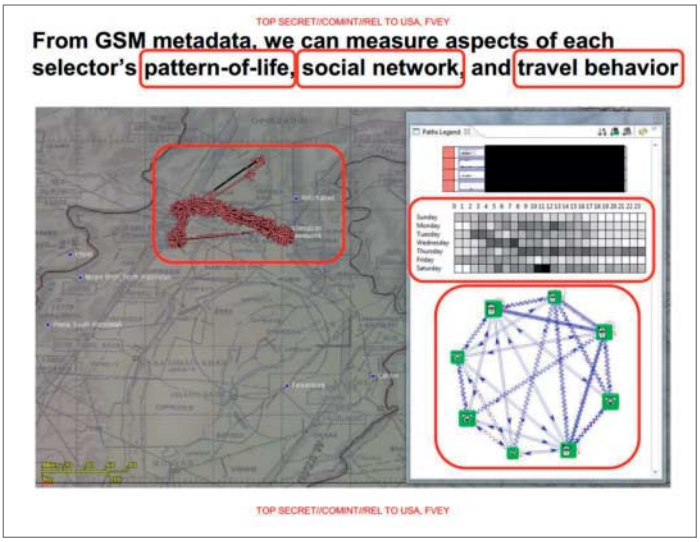
Was nach akademischer Spitzfindigkeit klingen mag, hat erheblichen Einfluss auf die Qualität der Ergebnisse – also darauf, wie Bürger als Terroristen klassifiziert und verfolgt werden. Darüber hinaus ist die Datenanalyse der NSA auch deswegen problematisch, weil die Random-Forest-Methode dazu neigt, sich zu stark an das Training-Set anzupassen. Die Analyse der NSA liefert daher übertrieben optimistische Ergebnisse und erbringt keinen Nachweis für die Qualität des gewählten Verfahrens.

In einer der Grafiken zeigt die NSA das Verhältnis zwischen „False Positives“ (x-Achse) und „False Negatives“ (y-Achse) für verschiedene Schwellenwerte und unterschiedliche Varianten der lernenden Algorithmen.

Das Endergebnis ist bemerkenswert: Der von der NSA gewählte Algorithmus produziert False Positives mit einer Rate von 0,18 Prozent. Dabei fallen allerdings 50 Prozent der mutmaßlichen Terroristen aus dem Raster. In einer optimierten Version fällt die Fehlerrate sogar auf 0,008 Prozent. Wegen der beschriebenen methodischen Fehler bleibt diese Evaluierung jedoch wissenschaftlich höchst fragwürdig. Die niedrigen Prozentsätze bei den „false positives“ sind vermutlich das Ergebnis dieses Fehlers.



Die NSA trainiert ihre Algorithmen anhand von Bewegungsmustern mutmaßlicher Kuriere für Terrororganisationen.



Eine NSA-Folie zeigt, was der Geheimdienst aus den Bewegungsmustern errechnet: Lebensmuster, soziale Netze und Reiseverhalten.

Die US-Angriffe werden unter anderem von Drohnen des Typs General Atomics MQ-9 Reaper geflogen.

Allerdings prahlt die Präsentation damit, dass sechs bereits bekannte Zielpersonen innerhalb der Top 100 der Auswertung aufgetaucht seien. Bei den Top 500 sind es 21. Die als „tasked selectors" bezeichneten Ziele sind Bürger, die bereits unabhängig von Skynet zur Überwachung durch die NSA markiert wurden.

Die optimistische Fehlerrate der NSA von 0,008 Prozent schlägt alle sonst von Big Data bekannten Erfolgsquoten. Eine solche Rate wäre mehr als akzeptabel für Unternehmen, die damit lediglich riskieren, Werbung nicht zielgruppengerecht abzuliefern oder der falschen Person einen Premiumpreis anzubieten. Doch 0,008 Prozent der pakistanischen Bevölkerung sind 15 000 unschuldige Bürger, die mit dem Label „Terrorist" versehen wurden. Gleichzeitig identifiziert das System bei dieser Fehlerquote nur jeden zweiten bereits als solchen bekannten „Terroristen".

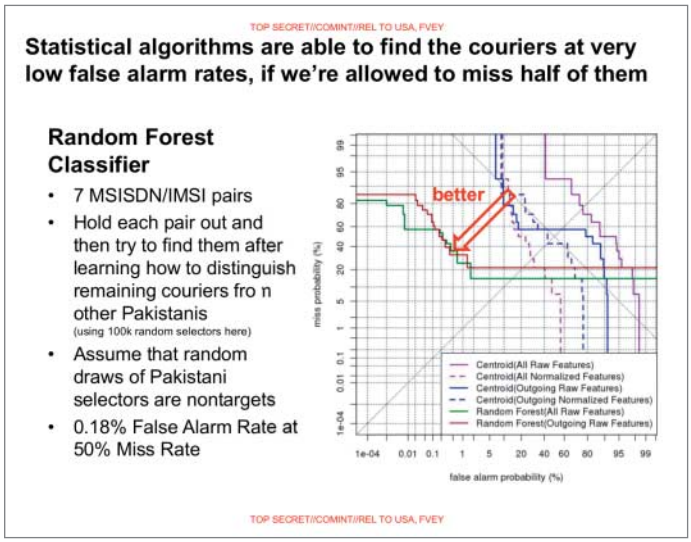
Nach Informationen des Bureau of Investigative Journalism wurden in der Zeit zwischen 2004 und 2015 insgesamt 2500 bis 4000 Menschen durch Drohnenangriffe in Pakistan getötet. Die meisten wurden von der US-Regierung als „Extremisten" eingestuft. Unbekannt ist, wie stark die NSA für ihre Klassifizierung und Zielsuche auf Skynet zurückgegriffen hat. Allerdings ist Skynet im Moment das einzige bekannte Programm zur Verarbeitung von Metadaten aus Pakistan.

Algorithmus statt Rechtsstaatlichkeit

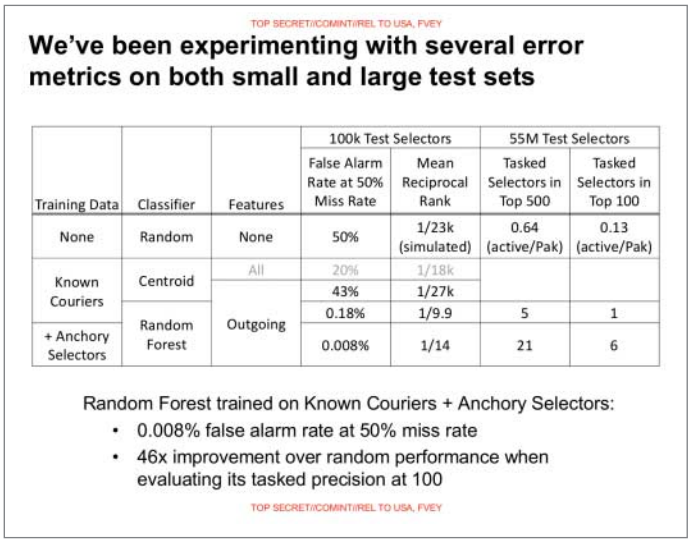
Statistische Datenanalyse und maschinelles Lernen sind die bevorzugten Werkzeuge, um großen Datenbergen zu Leibe zu rücken. Um die Genauigkeit der statistischen Methoden abzusichern, müssen wissenschaftliche Standards eingehalten werden. Nur dann lassen sich sinnvolle Ergebnisse erzielen. Eine besonders hohe Genauigkeit ist etwa bei der Wahl der Behandlung seltener Krankheiten erforderlich, um Fehldiagnosen zu vermeiden. Auf einen Bereich, in dem es nicht um die Rettung von Leben sondern um die Verurteilung und Tötung von Menschen geht, sollten diese Werkzeuge nicht übertragen werden.

Selbst bei einer theoretisch hohen Trefferquote, wie sie die NSA angesichts ihres offenbar laxen Umgangs mit wissenschaftlichen Methoden wohl kaum erreicht, ist Data Mining zur Selektion von Drohnenzielen auf keinen Fall akzeptabel. Abgesehen von Fragen der Rechtsstaatlichkeit und dem Menschenrecht auf ein faires Verfahren ist die Zahl der einkalkulierten unschuldig Getöteten nicht zu rechtfertigen. Maschinelles Lernen als Methode, um „Todeskandidaten" samt erwarteter ziviler Opfer für ein Attentatsprogramm zu selektieren, stellt national und international verbriefte Menschenrechte auf den Kopf. (mho@ct.de)

c't Weiterführendes: ct.de/ypzx



Diese NSA-Folie visualisiert die Erfolgsquote: Je weiter unten links die Linie gezeichnet wird, desto genauer werden nur wirkliche „Terroristen" als solche markiert.



Training Data	Classifier	Features	100k Test Selectors: False Alarm Rate at 50% Miss Rate	100k Test Selectors: Mean Reciprocal Rank	55M Test Selectors: Tasked Selectors in Top 500	55M Test Selectors: Tasked Selectors in Top 100
None	Random	None	50%	1/23k (simulated)	0.64 (active/Pak)	0.13 (active/Pak)
Known Couriers	Centroid	All	20%	1/18k		
		Outgoing	43%	1/27k		
	Random Forest		0.18%	1/9.9	5	1
+ Anchory Selectors			0.008%	1/14	21	6

Die NSA erreicht angeblich bemerkenswerte Ergebnisse bei der automatisierten Datenauswertung.