Die Krux mit Machine Learning: Fehler in der Anwendung mit fataler Auswirkung

Machine Learning ist groß, wirklich groß. Machine Learning ist eine Technologie, die alles verändern könnte. Sie hat sich bereits als bahnbrechende Technologie für so unterschiedliche Aufgaben wie das Aufspüren von Bankbetrug, das Fahren von Autos, das Verstehen menschlicher Sprache und das Erkennen von Malware bewährt.

In naher Zukunft werden Unternehmen viele Millionen investieren und Menschen werden Machine-Learning-Lösungen entwickeln und einsetzen. Doch neue Technologien führen oft zu neuen Fehlerquellen und Möglichkeiten für Cyberangriffe. Neben den Machenschaften der Hacker ist im jungen Stadium des Machine Learning sicherlich der Entwickler und Anwender die größte Gefahr. Machine Learning ist neu und komplex und das Potenzial für selbst verschuldetes Versagen der Technologie ist hoch.

Definition Machine Learning und Abgrenzung zu herkömmlicher Software
Traditionelle Software ist im Wesentlichen ein Regelwerk das festlegt, wonach sich ein Computer in einem bestimmten Kontext verhalten soll. Sie ist hervorragend im Umgang mit strukturierten Daten und insbesondere mit Dingen, in denen wir schlecht sind, namlich darin, hochkomplexe Befehlssätze perfekt und mit enormer Geschwindigkeit auszuführen.

Machine Learning hingegen ist ein Zweig der Künstlichen Intelligenz (KI). Hier werden Softwaremodelle verwendet, die anhand von Beispielen trainiert werden und ihre eigenen Regeln erstellen. Der Computerpionier Arthur Lee Samuel hält das maschinelle Lernen für einen „Forschungsbereich, der Computern die Möglichkeit gibt zu lernen, ohne explizit programmiert zu werden“. Während sich herkömmliche Software durch Transparenz und nachweislich korrektes Verhalten auszeichnet, ist Machine Learning unscharf, flexibel und undurchsichtig.

Gutes Füttern ist Grundvoraussetzung
Modernes Machine Learning funktionieren gut, weil es komplexe Zusammenhänge aus Trainingsdaten lernen kann. Auf diese Weise erkennt es Dinge, seien es Gesichter, Betrugsmuster oder Spam, mit denen menschliche Programmierer nicht mithalten können.
Aber diese Fähigkeit kann auf unerwartete Weise zum Bumerang werden. Wenn die Trainingsdaten eine falsche oder nicht existente Korrelation enthalten, kann Machine Learning leicht die falschen Lektionen lernen. Eine fatale Situation denn vielfach wird Machine Learning mit Big Data gefüttert, Daten, die voll von falschen Korrelationen sind.

Ein Beispiel: Man stelle sich ein Machine-Learning-Modell vor, um Spam-E-Mails zu erkennen. Die Trainingsdaten sind eine Datenbank mit E-Mails, die von Menschen gewissenhaft als "Ham" (E-Mails die wir mögen) oder "Spam"(E-Mails die wir nicht haben wollen) gekennzeichnet sind. Nun passiert es, dass die Trainingsdaten eine plausible aber falsche Korrelation enthalten: Zufällig landet jede E-Mail mit einem Bildanhang, die von einer IP-Adresse mit der Endung 12 stammt, im Spam-Stapel. Und damit ist es geschehen: obwohl alles mit größter Sorgfalt gepflegt wurde, erhält das Machine-Learning-Modell Datenmüll – mit Folgen. Das komplexe Modell kann man daraus schließen, dass das Vorhandensein der IP-Adresse eines Absenders, die mit 12 endet, in einer E-Mail mit einem Bildanhang ein sicherer Indikator für Spam ist, obwohl dies außerhalb unserer Trainingsdaten nicht der Fall ist. Wenn

dieses Modell in der Security eingesetzt würde, blockiert die Anti-Spam-Engine eine Menge korrekter E-Mails von Personen, deren IP-Adresse mit einer 12 endet.

Tipps für das richtige Füttern von Machine Learning
Es gibt keine einfache Lösung für das Machine Learning. Aber es gibt sechs Punkte auf die man achten sollte, um möglichst viele Probleme im Vorfeld auszuschließen:

- Verwendung von guten Daten. Es ist wichtig, das Modell mit vielen gut beschrifteten Daten (Labels) aus Quellen zu füttern, die ein reales Bild repräsentieren.
- Daten müssen bereinigt werden. Es ist mühevoll Daten zu reinigen, zu beschriften oder zu ändern, um Fehler zu minimieren. Aber der Aufwand lohnt.
- Das Modell sollte nicht zu stark trainiert werden, um es nicht zu überfrachten. Denn es geht nicht darum, die Trainingsdaten mit perfekter Klarheit zu erkennen, sondern darum, Dinge zu erkennen, die Ähnlichkeiten mit den Trainingsdaten aufweisen.
- False Positives und False Negatives sollten beim Test genau beachtet werden und das Modell sollte auch bei der Bereinigung der Daten unterstützen. Achtung: manchmal sind nur die Label falsch und das Modell ist richtig.
- Deep Learning sollte bevorzugte Methode des maschinellen Lernens sein. Untersuchungen haben ergeben, dass es besser ist, mit vielen unterschiedlichen Labels umzugehen, als mit flacheren Lernmethoden.

Eine ausführlichere Abhandlung zum Thema Machine Learning ist auf Naked Security unter: https://nakedsecurity.sophos.com/2018/03/01/machine-learning-self-defence-how-to-not-shoot-yourself-in-the-foot/

Über Sophos
Mehr als 100 Millionen Anwender in 150 Ländern verlassen sich auf Sophos‘ Complete-Security-Lösungen als den besten Schutz vor komplexen IT-Bedrohungen und Datenverlust. Sophos bietet dafür preisgekrönte Verschlüsselungs-, Endpoint-Security-, Web-, Email-, Mobile- und Network Security-Lösungen an, die einfach zu verwalten, zu installieren und einzusetzen sind. Das Angebot wird von einem weltweiten Netzwerk eigener Analysezentren, den SophosLabs, unterstützt.
Sophos hat seinen Hauptsitz in Boston, USA, und Oxford, Großbritannien. In Deutschland hat das Unternehmen seinen Hauptsitz in Wiesbaden und ist in Österreich und der Schweiz je an einem Standort vertreten. Weitere Informationen unter www.sophos.de.

Pressekontakt:

Sophos
Jörg Schindler, PR Manager CEEMEA
joerg.schindler@sophos.com, +49-721-25516-263

TC Communications
Arno Lücht, +49-8081-954619
Thilo Christ, +49-8081-954617
Ulrike Masztalerz, +49-30-55248198
Ariane Wendt +49-172-4536839
sophos@tc-communications.de