Machine Unlearning: Algorithmen können nichts vergessen

Originalartikel

Backup

<html> <p class=„printversionback-to-article printversion–hide“>zurück zum Artikel <figure class=„printversionlogo“></figure><figure class=„aufmacherbild“><figcaption class=„akwa-caption“><p class=„caption akwa-captiontext“>Prof. Nicolas Papernot während seines Vortrags auf der Usenix Enigma 2020. <p class=„source akwa-captionsource“>(Bild: Daniel AJ Sokolov) </figcaption></figure> Weil wir nicht wissen, wie die Maschine genau denkt, können wir sie nur indirekt vergessen lassen. Das wurde auf der Konferenz Usenix Enigma deutlich. Daten zu löschen ist in Zeiten des Machine Learning alles andere als einfach. Wurde ein Algorithmus einmal anhand großer Datenmengen trainiert, kann man ihn nicht einfach etwas vergessen machen. Daher befasst sich Professor Nicolas Papernot von der Universität Toronto mit dem jungen Thema des Machine Unlearning. Er empfiehlt, Löscherfordernisse schon in der Vorbereitung des Algorithmen-Trainings zu berücksichtigen. Zwei Ansätze verfolgt der Forscher dafür: Erstens lassen sich statt eines großen Modells, das mit allen verfügbaren Daten trainiert wurde, mehrere kleinere Modelle erstellen, die jeweils auf einen Teil der Daten zurückgreifen. Im Anwendungsfall lässt man dann alle Modelle laufen und fasst ihre Ergebnisse zusammen, beispielsweise in Form einer Mehrheitsabstimmung. Hier kann es aber unter Umständen zu zusätzlichem Aufwand beim Einsatz der Algorithmen kommen. Der Betreiber muss also vorher abschätzen, wie oft er seine Modelle verwenden und wie oft er Daten zu löschen haben wird. Zweitens kann man zunächst ein Modell mit einem Teil der Daten trainieren, beispielsweise zehn Prozent, und das Ergebnis speichern. Im nächsten Schritt wird das Modell genommen und mit einem weiteren Datenteil weiterentwickelt. Das Prozedere wird wiederholt, bis es ein mit allen Daten gefüttertes Modell gibt. Muss man später bestimmte Daten löschen, greift man auf jenen Zwischenstand zurück, der diese Datensätze noch nicht kannte, und trainiert nur von dort an neu. Sofern die zu löschenden Informationen nicht gerade im allerersten Modell enthalten sind, spart das später Zeit und Geld. <h3 class=„subheading“ id=„nav_weil_wir_sie0“>Weil wir sie nicht verstehen</h3> „Wir müssten nicht auf diese Lösungen zurückgreifen, wenn wir verstünden, wie Machine-Learning-Modelle ihre Vorhersagen treffen“, erläuterte Papernot heise online am Montag auf der Konferenz Usenix

Enigma in San Francisco. „Wir wissen nicht, welche Datensätze genau in eine bestimmte Berechnung eingeflossen sind. Daher können wir nicht garantieren, dass ein bestehender Algorithmus sich nicht mehr auf einen bestimmten Datensatz bezieht. Dieses Problem ist nicht gelöst.“ <div class=„collapse-boxtarget collapse-boxcontent a-inline-textboxcontent a-inline-textboxcontent–horizontal-layout“ data-collapse-target=„“> <figure class=„a-inline-textboximage-container“></figure> Keine News verpassen! Mit unserem täglichen Newsletter erhalten Sie jeden Morgen alle Nachrichten von heise online der vergangenen 24 Stunden. <ul class=„a-inline-textboxlist“><li class=„a-inline-textboxitem“>Newsletter jetzt abonnieren [1] Und einen Algorithmus von der Pike an neu zu schaffen kostet Zeit und Geld. Ein Paper zum Thema Machine Unlearning, zu dessen Autoren Papernot zählt, ist als Vorabdruck erhältlich [2]. Es befindet sich gerade in der Peer Review. <h3 class=„subheading“ id=„nav_warum_wir1“>Warum wir löschen müssen</h3> Gründe für notwendige Datenlöschung kann es mehrere geben: Einerseits könnten Bürger ihnen zustehende Datenlöschungen fordern, andererseits könnte ein Angreifer manipulierte Daten in die hinzugezogenen Datensätze eingeschleust haben. Auch Lizenzprobleme sind denkbar, beispielsweise wenn beim Machine Learning urheberrechtlich geschützte Werke ausgewertet wurden und eine entsprechende Lizenz abläuft oder sich im Nachhinein als ungültig erweist. <article class=„a-article-teaser a-article-teaser–horizontal-layout-small a-u-no-margin-bottom articleboxarticle-teaser“> <figure class=„a-article-teaserimage-container“> <noscript> </noscript> </figure> [3]</article> Schließlich könnten sich auch Zweck oder Parameter des Algorithmuseinsatzes ändern. Als Beispiel zieht Papernot eine medizinische Anwendung heran, bei der sich später herausstellt, dass Daten von Patienten vor einem bestimmten Geburtsjahr nicht länger relevant sind. In solchen und ähnlichen Fällen hilft es dem Betreiber, wenn er sein Machine-Learning-Modell nicht ganz von vorne neu trainieren muss. Wurden hingegen ganz einfach falsche Daten zum Training genutzt, ist es in der Regel möglich, dem Algorithmus dieses neu gewonnene Wissen nachträglich beizubringen, sagte Papernot. Das ist ein deutlich geringerer Aufwand als echte Machine Unlearning, bei dem einbezogene Daten verlässlich ausgenommen werden müssen. <h3 class=„subheading“ id=„nav_usenix_enigma2“>Usenix Enigma 2020</h3> Usenix Enigma ist eine jährliche Konferenz zu IT-Sicherheit und Datenschutz, die sich mit gegenwärtigen sowie sich

anbahnenden Bedrohungen an der Schnittstelle von Gesellschaft und Technik befasst. Sie findet diese Woche mit zirka 450 Teilnehmern in San Francisco statt. Es ist die fünfte Auflage der Veranstaltung. ()<br class=„clear“/> <hr/> URL dieses Artikels:
<small><code>http://www.heise.de/-4646747</code></small> Links in diesem Artikel:
<small><code>[1] https://www.heise.de/newsletter/manage/ho?wt_mc=nl.red.ho.daily.meldung.link.link</code></small>
<small><code>[2] https://arxiv.org/abs/1912.03817</code></small>
<small><code>[3] https://www.heise.de/meldung/BDI-Datenschutzvorgaben-sind-toxisch-fuer-Kuenstliche-Intelligenz-4538401.html</code></small>
<small><code>[4] mailto:ds@heise.de</code></small>
 Copyright © 2020 Heise Medien </html>