

# Machine Unlearning: Algorithmen können nichts vergessen

Originalartikel

Backup

<html> <p class=„printversionback-to-article printversion-hide“><a href=„<https://www.heise.de/newsticker/meldung/Machine-Unlearning-Algorithmen-koennen-nichts-vergessen-4646747.html>“>zur&#252;ck zum Artikel</a></p> <figure class=„printversionlogo“><img src=„<https://1.f. ix.de/icons/svg/logos/svg/heiseonline.svg>“ alt=„heise online“ width=„180“ heighth=„40“></figure> <figure class=„aufmacherbild“><img src=„[https://heise.cloudimg.io/width/700/q75.png-lossy-75.webp-lossy-75.foil1/\\_www-heise-de\\_/imgs/18/2/8/3/1/7/2/5/Papernot\\_Useix\\_E nigma-414b788e3b479e98.jpeg](https://heise.cloudimg.io/width/700/q75.png-lossy-75.webp-lossy-75.foil1/_www-heise-de_/imgs/18/2/8/3/1/7/2/5/Papernot_Useix_E nigma-414b788e3b479e98.jpeg)“ srcset=„[https://heise.cloudimg.io/width/700/q75.png-lossy-75.webp-lossy-75.foil1/\\_www-heise-de\\_/imgs/18/2/8/3/1/7/2/5/Papernot\\_Useix\\_E nigma-414b788e3b479e98.jpeg](https://heise.cloudimg.io/width/700/q75.png-lossy-75.webp-lossy-75.foil1/_www-heise-de_/imgs/18/2/8/3/1/7/2/5/Papernot_Useix_E nigma-414b788e3b479e98.jpeg)“ 700w, [https://heise.cloudimg.io/width/1050/q75.png-lossy-75.webp-lossy-75.foil1/\\_www-heise-de\\_/imgs/18/2/8/3/1/7/2/5/Papernot\\_Useix\\_E nigma-414b788e3b479e98.jpeg](https://heise.cloudimg.io/width/1050/q75.png-lossy-75.webp-lossy-75.foil1/_www-heise-de_/imgs/18/2/8/3/1/7/2/5/Papernot_Useix_E nigma-414b788e3b479e98.jpeg) 1050w, [https://heise.cloudimg.io/width/1500/q75.png-lossy-75.webp-lossy-75.foil1/\\_www-heise-de\\_/imgs/18/2/8/3/1/7/2/5/Papernot\\_Useix\\_E nigma-414b788e3b479e98.jpeg](https://heise.cloudimg.io/width/1500/q75.png-lossy-75.webp-lossy-75.foil1/_www-heise-de_/imgs/18/2/8/3/1/7/2/5/Papernot_Useix_E nigma-414b788e3b479e98.jpeg) 1500w, [https://heise.cloudimg.io/width/1600/q75.png-lossy-75.webp-lossy-75.foil1/\\_www-heise-de\\_/imgs/18/2/8/3/1/7/2/5/Papernot\\_Useix\\_E nigma-414b788e3b479e98.jpeg](https://heise.cloudimg.io/width/1600/q75.png-lossy-75.webp-lossy-75.foil1/_www-heise-de_/imgs/18/2/8/3/1/7/2/5/Papernot_Useix_E nigma-414b788e3b479e98.jpeg) 1600w“ sizes=„(min-width: 80em) 43.75em, (min-width: 64em) 66.66vw, 100vw“ alt=„Schlanker wei&#223;er Mann in hellem Pullover an Rednerpult, dahinter hellblaue Wand mit Schriftzug „ENIGMA““ class=„img-responsive“><figcaption class=„akwa-caption“><p class=„caption akwa-captiontext“>Prof. Nicolas Papernot w&#228;rend seines Vortrags auf der Usenix Enigma 2020.</p> <p class=„source akwa-captionsource“>(Bild:&#160;Daniel AJ Sokolov)</p> </figcaption></figure><p><strong>Weil wir nicht wissen, wie die Maschine genau denkt, k&#246;nnen wir sie nur indirekt vergessen lassen. Das wurde auf der Konferenz Usenix Enigma deutlich.</strong></p> <p>Daten zu I&#246;schen ist in Zeiten des Machine Learning alles andere als einfach. Wurde ein Algorithmus einmal anhand gro&#223;er Datenmengen trainiert, kann man ihn nicht einfach etwas vergessen machen. Daher befasst sich Professor Nicolas Papernot von der Universit&#228;t Toronto mit dem jungen Thema des Machine Unlearning. Er empfiehlt, L&#246;scherfordernisse schon in der Vorbereitung des Algorithmen-Trainings zu ber&#252;cksichtigen.</p> <p>Zwei Ans&#228;tze verfolgt der Forscher daf&#252;r: Erstens lassen sich statt eines gro&#223;en Modells, das mit allen verf&#252;gbaren Daten trainiert wurde, mehrere kleinere Modelle erstellen, die jeweils auf einen Teil der Daten zur&#252;ckgreifen. Im Anwendungsfall l&#228;sst man dann alle Modelle laufen und fasst ihre Ergebnisse zusammen, beispielsweise in Form einer Mehrheitsabstimmung. Hier kann es aber unter Umst&#228;nden zu zus&#228;tzlichem Aufwand beim Einsatz der Algorithmen kommen. Der Betreiber muss also vorher absch&#228;tzen, wie oft er seine Modelle verwenden und wie oft er Daten zu I&#246;schen haben wird.</p> <div class=„inread“> <p>Zweitens kann man zun&#228;chst ein Modell mit einem Teil der Daten trainieren, beispielsweise zehn Prozent, und das Ergebnis speichern. Im n&#228;chsten Schritt wird das Modell genommen und mit einem weiteren Datenteil weiterentwickelt. Das Prozedere wird wiederholt, bis es ein mit allen Daten gef&#252;ttertes Modell gibt. Muss man sp&#228;ter bestimmte Daten I&#246;schen, greift man auf jenen Zwischenstand zur&#252;ck, der diese Datens&#228;tze noch nicht kannte, und trainiert nur von dort an neu. Sofern die zu I&#246;schen Informationen nicht gerade im allerersten Modell enthalten sind, spart das sp&#228;ter Zeit und Geld.</p> <h3 class=„subheading“ id=„nav\_weil\_wir\_sie0“>Weil wir sie nicht verstehen</h3> <p>„Wir m&#252;ssten nicht auf diese L&#246;sungen zur&#252;ckgreifen, wenn wir verst&#252;nden, wie Machine-Learning-Modelle ihre Vorhersagen treffen“, erl&#228;uterte Papernot heise online am Montag auf der Konferenz Usenix

Enigma in San Francisco. „Wir wissen nicht, welche Datens<#228;tze genau in eine bestimmte Berechnung eingeflossen sind. Daher k<#246;nnen wir nicht garantieren, dass ein bestehender Algorithmus sich nicht mehr auf einen bestimmten Datensatz bezieht. Dieses Problem ist nicht gel<#246;st.“

</p> <div class=„collapse-boxtarget collapse-boxcontent a-inline-textboxcontent a inline-textboxcontent-horizontal-layout“ data-collapse-target=„“> <figure class=„a-inline-textboximage-container“><img alt=„src=„https://heise.cloudimg.io/width/1600/q50.png-lossy-50.webp-lossy-50.foil1/\_www-heise-de/\_imgs/71/2/8/1/2/5/7/6/daily\_grafik-b53d51b363224691.png“ srcset=„https://heise.cloudimg.io/width/3200/q30.png-lossy-30.webp-lossy-30.foil1/\_www-heise-de/\_imgs/71/2/8/1/2/5/7/6/daily\_grafik-b53d51b363224691.png 2x“ class=„c1“></figure><div class=„a-inline-textboxcontent-container“> <p class=„a-inline-textboxsynopsis“>Keine News verpassen! Mit unserem täglichlichen Newsletter erhalten Sie jeden Morgen alle Nachrichten von heise online der vergangenen 24 Stunden.</p> <ul class=„a-inline-textboxlist“><li class=„a-inline-textboxitem“><a class=„a-inline-textboxtext“ href=„https://www.heise.de/newsletter/manage/ho?wt\_mc=nl.red.ho.daily.meldung.link.link“ title=„Newsletter jetzt abonnieren“><strong>Newsletter jetzt abonnieren [1]</strong></a></li></ul></div> <p>Und einen Algorithmus von der Pike an neu zu schaffen kostet Zeit und Geld. Ein Paper zum Thema Machine Unlearning, zu dessen Autoren Papernot z<#228;hlt, ist <a href=„https://arxiv.org/abs/1912.03817“ rel=„external noopener“ target=„\_blank“><strong>als Vorabdruck erh<#228;ltlich [2]</strong></a>. Es befindet sich gerade in der Peer Review.</p> <h3 class=„subheading“ id=„nav\_warum\_wir1“>Warum wir l<#246;schen m<#252;ssen</h3> <p>Gr<#252;nde f<#252;r notwendige Datenl<#246;schung kann es mehrere geben: Einerseits k<#246;nnten B<#252;rger ihnen zustehende Datenl<#246;schungen fordern, andererseits k<#246;nnte ein Angreifer manipulierte Daten in die hinzugezogenen Datens<#228;tze eingeschleust haben. Auch Lizenzprobleme sind denkbar, beispielsweise wenn beim Machine Learning urheberrechtlich gesch<#252;tzte Werke ausgewertet wurden und eine entsprechende Lizenz abl<#228;uft oder sich im Nachhinein als ung<#252;ltig erweist.</p> <div class=„articlebox“> <article class=„a-article-teaser a-article-teaser-horizontal-layout-small a-u-no-margin-bottom articleboxarticle-teaser“><a class=„a-article-teaserlink“ href=„https://www.heise.de/meldung/BDI-Datenschutzvorgaben-sind-toxisch-fuer-Kuenstliche-Intelligenz-4538401.html“ name=„...1“ title=„BDI: Datenschutzvorgaben sind "toxisch f<#252;r K<#252;nstliche Intelligenz"“> <figure class=„a-article-teaserimage-container“><div><strong><noscript> <p><img alt=„BDI: Datenschutzvorgaben sind &quot;toxisch f<#252;r K<#252;nstliche Intelligenz&quot;“ class=„c1“ src=„https://heise.cloudimg.io/width/200/q50.png-lossy-50.webp-lossy-50.foil1/\_www-heise-de/\_imgs/18/2/8/3/1/7/2/5/DSC\_2131-07d1bc83f904041e.jpeg“ srcset=„https://heise.cloudimg.io/width/200/q30.png-lossy-30.webp-lossy-30.foil1/\_www-heise-de/\_imgs/18/2/8/3/1/7/2/5/DSC\_2131-07d1bc83f904041e.jpeg 2x“></p></noscript></strong></div></figure><div class=„a-article-teasercontent-container“> </div> [3]</a></article></div> <p>Schlie<#223;lich k<#246;nnten sich auch Zweck oder Parameter des Algorithmuseinsatzes <#228;ndern. Als Beispiel zieht Papernot eine medizinische Anwendung heran, bei der sich sp<#228;ter herausstellt, dass Daten von Patienten vor einem bestimmten Geburtsjahr nicht l<#228;nger relevant sind. In solchen und <#228;hnlichen F<#228;llen hilft es dem Betreiber, wenn er sein Machine-Learning-Modell nicht ganz von vorne neu trainieren muss.</p> <p>Wurden hingegen ganz einfach falsche Daten zum Training genutzt, ist es in der Regel m<#246;glich, dem Algorithmus dieses neu gewonnene Wissen nachtr<#228;glich beizubringen, sagte Papernot. Das ist ein deutlich geringerer Aufwand als echte Machine Unlearning, bei dem einbezogene Daten verl<#228;sslich ausgenommen werden m<#252;ssen.</p> <h3 class=„subheading“ id=„nav\_usenix\_enigma2“>Usenix Enigma 2020</h3> <p>Usenix Enigma ist eine j<#228;hrliche Konferenz zu IT-Sicherheit und Datenschutz, die sich mit gegenw<#228;rtigen 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 f&#252;nfte Auflage der Veranstaltung. ()  
<br class=„clear“/></p> <hr/><p><strong>URL dieses Artikels:</strong><br/><small><code><http://www.heise.de/-4646747></code></small></p><p><strong>Links in diesem Artikel:</strong><br/><small><code><strong>[1]</strong>&#160;[https://www.heise.de/newsletter/manage/ho?wt\\_mc=nl.red.ho.daily.meldung.link.link](https://www.heise.de/newsletter/manage/ho?wt_mc=nl.red.ho.daily.meldung.link.link)</code></small><br/><small><code><strong>[2]</strong>&#160;<https://arxiv.org/abs/1912.03817></code></small><br/><small><code><strong>[3]</strong>&#160;<https://www.heise.de/meldung/BDI-Datenschutzvorgaben-sind-toxisch-fuer-Kuenstliche-Intelligenz-4538401.html></code></small><br/><small><code><strong>[4]</strong>&#160;<mailto:o:ds@heise.de></code></small><br/></p> <p class=„printversioncopyright“><em>Copyright &#169; 2020 Heise Medien</em></p> </html>

From:

<https://schnipsl.qgelm.de/> - Qgelm

Permanent link:

[https://schnipsl.qgelm.de/doku.php?id=wallabag:machine-unlearning\\_-algorithmen-knnen-nichts-vergessen](https://schnipsl.qgelm.de/doku.php?id=wallabag:machine-unlearning_-algorithmen-knnen-nichts-vergessen)

Last update: 2021/12/06 15:24

