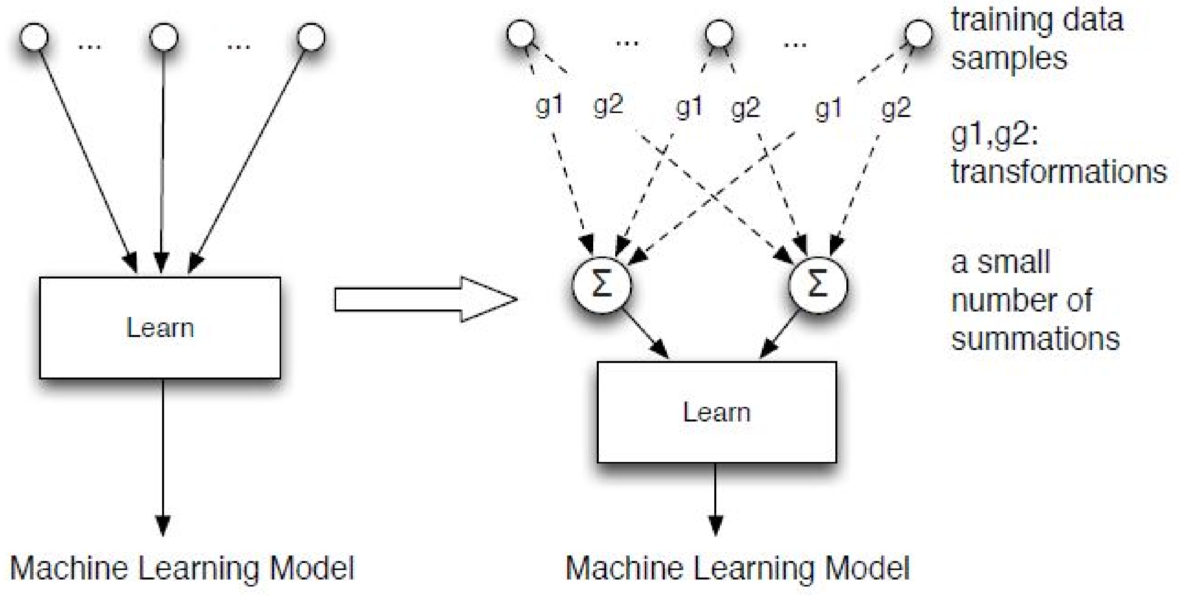
PHYS.ORG

# New 'machine unlearning' technique wipes out unwanted data quickly and completely

March 14 2016



The novel approach to making systems forget data is called "machine unlearning" by the two researchers who are pioneering the concept. Instead of making a model directly dependon each training data sample (left), they convert the learning algorithm into a summation form (right) - a process that is much easier and faster than retraining the system from scratch. Credit: Yinzhi Cao and Junfeng Yang

Machine learning systems are everywhere. Computer software in these machines predict the weather, forecast earthquakes, provide recommendations based on the books and movies we like and, even, apply the brakes on our cars when we are not paying attention.

To do this, computer systems are programmed to find predictive relationships calculated from the massive amounts of data we supply to them. Machine learning systems use advanced algorithms—a set of rules for solving math problems—to identify these predictive relationships using "training data." This data is then used to construct the models and features within a system that enables it to correctly predict your desire to read the latest best-seller, or the likelihood of rain next week.

This intricate learning process means that a piece of raw data often goes through a series of computations in a given system. The data, computations and information derived by the system from that data together form a complex propagation network called the data's "lineage." The term was coined by researchers Yinzhi Cao of Lehigh University and Junfeng Yang of Columbia University who are pioneering a novel approach toward making such learning systems forget.

Considering how important this concept is to increasing security and protecting privacy, Cao and Yang believe that easy adoption of forgetting systems will be increasingly in demand. The pair has developed a way to do it faster and more effectively than what is currently available.

Their concept, called "machine unlearning," is so promising that the duo have been awarded a four-year, $1.2 million National Science Foundation grant—split between Lehigh and Columbia—to develop the approach.

"Effective forgetting systems must be able to let users specify the data to forget with different levels of granularity," said Yinzhi Cao, Assistant Professor of Computer Science and Engineering at Lehigh University's

P.C. Rossin College of Engineering & Applied Science and a Principal Investigator on the project. "These systems must remove the data and undo its effects so that all future operations run as if the data never existed."

## Increasing security & privacy protections

There are a number of reasons why an individual user or service provider might want a system to forget data and its complete lineage. Privacy is one.

After Facebook changed its privacy policy, many users deleted their accounts and the associated data. The iCloud photo hacking incident in 2014—in which hundreds of celebrities' private photos were accessed via Apple's cloud services suite—led to online articles teaching users how to completely delete iOS photos including the backups. New research has revealed that machine learning models for personalized medicine dosing leak patients' genetic markers. Only a small set of statistics on genetics and diseases are enough for hackers to identify specific individuals, despite cloaking mechanism.

Naturally, users unhappy with these newfound risks want their data and its influence on the models and statistics to be completely forgotten.

Security is another reason. Consider anomaly-based intrusion detection systems used to detect malicious software. In order to positively identify an attack, the system must be taught to recognize normal system activity. Therefore the security of these systems hinges on the model of normal behaviors extracted from the training data. By polluting the training data, attackers pollute the model and compromise security. Once the polluted data is identified, the system must completely forget the data and its lineage in order to regain security.

Widely-used learning systems such as Google Search are, for the most part, only able to forget a user's raw data upon request and not that data's lineage. While this is obviously problematic for users who wish to ensure that any trace of unwanted data is removed completely, this limitation is also a major challenge for service providers who have strong incentives to fulfill data removal requests, including the retention of customer trust.

Service providers will increasingly need to be able to remove data and its lineage completely to comply with laws governing user data privacy, such as the "right to be forgotten" ruling issued in 2014 by the European Union's top court. In October 2014 Google removed more than 170,000 links to comply with the ruling that affirmed an individual's right to control what appears when their name is searched online. In July 2015, Google said it had received more than a quarter-million requests.

## Breaking down dependencies

Building on their previous work that was revealed at a 2015 IEEE Symposium and then published, Cao's and Yang's "machine unlearning" method is based on the fact that most learning systems can be converted into a form that can be updated incrementally without costly retraining from scratch.

Their approach introduces a layer of a small number of summations between the learning algorithm and the training data to eliminate dependency on each other. So, the learning algorithms depend only on the summations and not on individual data. Using this method, unlearning a piece of data and its lineage would no longer require re-building the models and features that predict relationships between pieces of data. Simply re-computing a small number of summations would remove the data and its lineage completely—and much faster than through retraining the system from scratch.

Cao says he believes they are the first to establish the connection between unlearning and the summation form.

And, it works. Cao and Yang evaluated their unlearning approach by testing it out on four real-world systems. The diverse set of programs serves as a representative benchmark for their method and included LensKit, an open-source recommendation system; Zozzle, a closed-source JavaScript malware detector; an open-source OSN spam filter and PJScan, an open-source PDF malware detector.

The success they achieved during these initial evaluations have set the stage for the next phases of the project, which include adapting the technique to other systems and creating verifiable machine unlearning to statistically test whether unlearning has indeed repaired a system or completely wiped out unwanted data.

In their paper's introduction, Cao and Yang look ahead to what's next for Big Data and are convinced that "machine unlearning" could play a key role in enhancing security and privacy and in our economic future:

"We foresee easy adoption of forgetting systems because they benefit both users and service providers. With the flexibility to request that systems forget data, users have more control over their data, so they are more willing to share data with the systems. More data also benefit the service providers, because they have more profit opportunities and fewer legal risks."

They add: "...we envision forgetting systems playing a crucial role in emerging data markets where users trade data for money, services, or other data because the mechanism of forgetting enables a user to cleanly cancel a data transaction or rent out the use rights of her data without giving up the ownership."

**More information:** dx.doi.org/10.1109/SP.2015.35

Provided by Lehigh University

Citation: New 'machine unlearning' technique wipes out unwanted data quickly and completely (2016, March 14) retrieved 26 April 2024 from https://phys.org/news/2016-03-machine-unlearning-technique-unwanted-quickly.html