ISSN: 0975-766X
CODEN: IJPTFI
Available Online through *Research Article*
www.ijptonline.com

MACHINE UNLEARNING- AN APPROACH TO MAKE MACHINE LEARNING ALGORITHMS FORGET DATAAND AUTHENTICATES ML LEVEL SECURITY

[1]Ashish Gupta, [2]Nancy Victor
School of Information Technology and Engineering, VIT University, Vellore, TN, India.
Email: ashishgupta1v@gmail.com

Received on 25-10-2016 *Accepted on 02-11-2016*

Abstract

In this age of big data, system produce massive amounts of data, that further derives more data, and even propagated to different clouds, forming a data propagation network, that is difficult to analyze, known as Data's lineage. Data's lineage is the derived data, which sometimes users wants to delete. There may exist several reasons that users want systems to delete certain data with its lineage. Removing noise and incorrect entries from lineage allows the recommender system to give only useful recommendations. Also, one may want the recommender systems to forget their data, to avoid privacy risks. This paper focuses on transforming the machine learning algorithms to forget the data (or say its lineage) and also Machine learning level user authentication to provide users better security mechanisms and ensure learning algorithms to maintain privacy and hassle free data. We present a comprehensive approach of transforming learning algorithms to a special form i.e. called Statistic Query Learning. This approach uses small number of aggregation, i.e. faster than retraining recommender system from scratch. And also this approach can be applied to all levels of machine learning algorithms.

Keywords: Lineage, Summation, Recommender systems, Statistic Query Learning.

I. Introduction

A. Need to remove data from Recommender systems

In this age of Big Data, systems are producing a rapidly increasing amount of data. That data may be from Internet Sources, Black box, Satellites data, or social media, etc. This data is going through countless computations which derives even more data. The derived data may be due to Machine Learning algorithms which extracts models and properties from the training data using some advanced machine learning algorithms. So, data recursively derives more data, as like

recommender system predicting sentiments analysis of tweets based on twitter data. In short, data often goes through series of computations, growing in a complex network in different places, residing in many forms.

For instance, based on our experience, if a man search for women's clothing in amazon, Amazon's Recommender system will get to know that you need that clothes, and offers you similar variety of products (based on User Ratings) available in market. And you start getting recommendations of the same product in our cell phone, when there is any price updates or sale for that type products for some period of time. Now based on the machine learning algorithms, recommender systems learn that you need a product.

Now, if the user doesn't want that similar products as it not required any more, he deletes the Google 's Search history, and blackout all the search results. Now, even after removing the actual data i.e. women clothing, user got to see some recommendations. The reason is it's not just the data, i.e. clothes, but it also the lineage, which the user is unable to delete.

So, in this era of Big data, we have original data and the data which is generated from the actual data. That is further amplified and is propagated to different clouds, and is going to generate additional more data. This is our primary reason; we want our system to forget i.e. usability. Second Reason i.e. noise or incorrect entries in analytics that can seriously degrade the quality of recommendation[5].System operators or service providers have strong incentives, to honor users requests to forget data, both to keep users happy and to comply with the law [8].

Privacy is the third reason, where user got inference attacks due to Item-Item Matrix. iCloud Data Breach: Hacking Celebrity Photos and sharing to imgur, twitter and social networking sites [7] led to completely delete iOS photos including the backups [10].

So, there's a need that user wants the system to forget data, not only the original data but also data lineage. The ability to remove a single thread of data from the larger set of data has multiple potential benefits. It provides an ability to user to remove their own sensitive personal data from machine. For instance, Google had removed 171,183 links, or 41.8 percent of all requests across the EU under the "right to be forgotten" by October 2014 [13].

B. Attacks

Further inspiring the need for machine unlearning, we prescribed some of the attacks that target learning systems.

1) System Inference Attacks

Training data sets often contains some private data, and this private data go through the machine learning system, including feature set and training models, which can lead to Inference attacks where an attacker can feed variety of samples to the model, getting the opportunity to steal the private information from result set, exploiting its lineage and training dataset. Such an attack is calleda system inference attack [9]. Recommender system uses various collaborative filtering methods, user-user collaborative filtering is one, which learns item-item similarities from the pool of users and recommends the user the similar types of items or data he actually have gone through. Consider a user's friendgot some disease and user search for the respective disease, the recommender system will recommend medicines related to the respective disease, also this system infers that the user have a disease, this is due to item-item matrix. So, next time when the user search for something, recommender system recommends medicines, as it comes to know, user have a particular disease. So, we can remove actual data, but not its lineage, leading to inference attacks.

Some Statistics show that, 15 million users from US and approximately 5.5 million Canadians removed themselves from Facebook, due to privacy concerns [11].

2) Training Data Pollution Attacks: An attacker use training data and carefully injects the polluted data samples, into the learning system, leading the algorithms to produce false positive results basedon computation of incorrectfeature set and model, thus exploiting lineage. Subsequently, algorithm produces a large number of malicious results, misleading the systems that rely on features and these true malicious samples evade detection. For instance, three researchers, Battista Biggio, Blaine Nelson (Italy) and Pavel Laskov (Germany), have found a way to feed an SVM with polluted data specially designed to increase the error rate of the learning system as much as possible [2].

II. Background

A. Machine learning algorithm

Figure 1 shows, ageneral machine learning system with three processing stages:

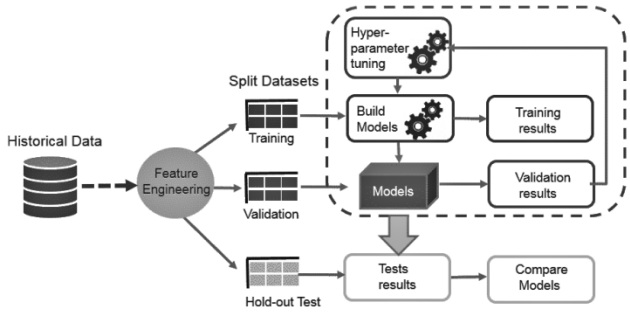

Fig. 1: Machine Learning System.

Given set of training data which includes both benign (+) and malign (-) samples, this learning system first selects the set of features that are most important for classifying data. Then it uses the training dataset to build the model. For processing an unknown sample data, the system examines thefeatures(attributes) in the sample data set, and uses the model to predict whether the sample data is benign or malign. Thus the lineage of the training data passes to the feature set, model and the prediction results, which results out in training data pollution attack. Also this is not only the data, but the lineage that flows through different processes. So, to remove the lineage, we need to build the model from scratch, which is a very slow process.

Feature selection: In this stage, the system selects the set of features, from all the features of the training data, that are most crucial for classifying data. This selected feature set is typically small, so as to make further stages more accurate and efficient. Feature selection can bemanual where programmers carefully craft the featuresor can beautomatic where somelearning algorithms are used to select the most crucial features.

Model Training:From each training data, the system extracts the data of the selected feature into a feature vector. It feeds the feature vector and the benign or malignlabels of training results into machine learning algorithms to construct a precise model.

Prediction: When the system receives data sample, it extracts the data sample's feature vector and use the model to make predictions, whether the sample is benign or malign.

This is a general machine learning approach we have used,as it matches many of the machine learning systems, so we are trying to transform this machine learning algorithm into some special form where we can make predictions and remove the incorrect data and maintain privacy by having a machine learning level user authentication.

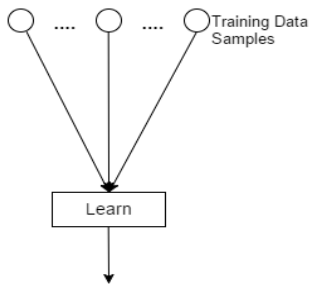

Fig. 2: Basic Machine Learning Model.

Given model is directly dependent on the given training data sample. So, this learning model will predict the result according to given training data.

III. Recommendation System

Recommendation systems are just an automated form of a "shop counter guy" who not only shows that product which you asked, but also the related ones, which you can buy. They know how to cross sell and up sell. So, does our recommendation systems. These systems have the ability to recommend personalized content to user, based on past behavior. For instance, if user-A likes the Items 1, 2, 4, 5 and user-B likes items 1, 2, 5, 6, then they have similar interests, so user-A will be recommended item 6, corresponding user-B will be recommended item 4. Collaborative filtering is tailor made for solving these recommendation problems. Using item look alike matrix, algorithm will recommend alike items to the user, who have done some transaction or purchased any item from store. This item-item collaborative filtering works efficient as the product-product look alike matrix is fixed. So, for considering iisthe item, for which we have to find similar items$r_{xi} = \frac{\sum_{j \epsilon N(i;x)} s_{ij} \cdot r_{xj}}{\sum_{j \epsilon N(i;x)} s_{ij}}$, here$s_{ij}$is a similarity matrix of items i and j.$r_{xj}$is rating of user x on item j. $N(i;x)$is set of items most similar to i, rated by x. Estimate rating r_{xi}as the weighted average.

This item-item collaborative filtering learning algorithm doesn't need any feature selection, that make it ideal, but this algorithm needs enough users in the system, to find the similar product, we can say, it cannot recommend any unrated item, and tends to recommend popular items.

So, we combine 2 learning algorithms, content-based and collaborative filtering, which combines prediction of both algorithms, providing demographics to deal with new user problem.This is done by using the global baseline added with collaborative filtering. Baseline estimated forr_{xi}as:

$$b_{xi} = \mu + b_x + b_i$$

where, μ is overall mean of product rating, b_x is (avg. rating of user x) – μ, b_i is rating deviation of product i. So, estimated Rating r_{xi}as the weighted average:

$$r_{xi} = b_{xi} + \frac{\sum_{j \epsilon N(i;x)} s_{ij} \cdot r_{xj}}{\sum_{j \epsilon N(i;x)} s_{ij}}$$

IV. Machine Unlearning

This approach as prescribed by the Yinzhi Caowho introduces a layer between the learning algorithm and the training data. This layer is of small number of summations (∑) so to break down the dependencies [5]. Now this unlearning

approach entirely depends on the summations, that efficiently computes transformations of training data sample. Many machine learning algorithms including Naïve Bayes, k-means clustering, support vector machines can be represented in this form and further this summation passes through the channel of encryption where this machine learning algorithm authenticates the user. If suppose a user wants the machine learning algorithm to forget the data, at initial stage the learning mechanism first authenticates the user's ability to delete the incorrect entries or noise from the training data set. This enhances the usability of unlearning algorithm, making the attackers to not to mislead the training data set by feeding different samples to the model, which leads to exploiting data and steals private information. And also the attackers are unable to inject polluted data into a learning system, which leadthe algorithm to produce false positive results.

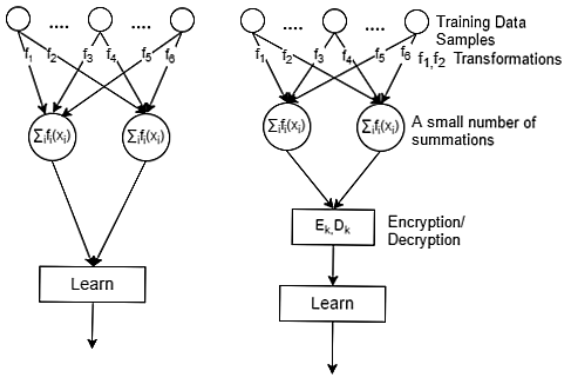

Fig. 3: Machine Unlearning Model with machine learning level user authentication.

This Unlearning algorithm prescribed by the Yinzhi cao converts the learning algorithms into summation form(left), here each summation $\left(\sum_i f_i\ (x_i)\right)$ is the sum of transformed data sample, having each transformation function fi efficiently computable. Now this learning algorithm depends upon summations, where each summation contains set of training data sets. Now on the right, we provided machine learning level user authentication where the learning algorithm authenticates whether the user must not be an attacker. After successful authentication, if the user wants the data to be forgotten including its lineage, simply updatethe summation and compute the updated model. For this, recommender system reverts the effects of data on the crafted features and models.This process of reverting back the piece of data, we call as machine unlearning.

If a feature is added in training data, we update the summations automatically, enabling the feature to compute its data within the model. Same time, if feature is added, data is to be computed into the model using summations.

To remove a piece of data, we simply delete the transformationsfrom the summations that depend on that sample. This have a complexity of O(1), and compute the updated model asymptotically faster, than retraining it from scratch.

Statistical Query Learning is used for these summations. SQ learning algorithm allows the algorithm to query only statistics of the train data using some structured database, this does not allow to query individual data sample. More precisely, this algorithm sends a function $f(x, l_x)$, here x is a training data sample, lx is the corresponding label of that sample and f is the efficiently computable functions. Subsequently, oracle database results in an estimated expectation of $f(x, l_x)$overthe training data set.

V. STATISTICAL QUERY LEARNING WITH ENCRYPTION

Many learning algorithms can be converted to statistical query learning with exactly the same precision, these algorithmsinclude Naïve Bayes, K-means clustering, Support vector machines.

Let's say, we have all the training data as $x1, x2, x3, x4$ ... and based on the training data, we are going to generate its summation form$\left(\sum_i f_i\ (x_i\ , l_{xi})\right)$, i.e. adding multiple training data sets together using fifunctions. This fiis the very efficiently computed functions. We have limit to the number of summations based on number of fifunction, we have $\left(\sum_i f_1\ (x_i\ , l_{xi}), \sum_i f_2\ (x_i\ , l_{xi}) \dots \dots \dots, \sum_i f_n\ (x_i\ , l_{xi})\right)$Then we are going to compute this machine learning model purely on the summations instead of the initial training data. Likewise

$$\text{learn}\left(\sum_i f_1\ (x_i\ , l_{xi}), \sum_i f_2\ (x_i\ , l_{xi}) \dots \dots \dots, \sum_i f_n\ (x_i\ , l_{xi})\right)$$

herex is a training data sample, lx is the corresponding label.

Now let $F_k$be$\sum f_k\ (x_i\ , l_{xi})$. All $F_k$sare savedwith the learning model.To unlearn the data sample xa, we just subtract the data sample xafrom each summation, subsequently update the model. As illustrated in Figure3, deleting a data item now requires re-computing only on a small number of summation i.e. F_k.

so, we computeF'_kas $F_k - f_k(x_a, l_{xa})$.

The updated model is thus

$$\text{learn}\left(F_1 - f_1(x_a, l_{xa}), F_2 - f_2(x_a, l_{xa}), \dots \dots, F_n - f_n(x_a, l_{xa})\right)$$

But before updating the records, machine learning algorithm authenticates the user based on a proposed encryption algorithm. This algorithm works on the principle of num%9, which generates a new sequence after every 9 digits it

comes across. So making the intruder hard to find out the key, generating every time new sequence out of 128 ASCII characters. So, this proposed algorithm having the time complexity of O(n), is described as follows:

Algorithm 1: Encryption

Input: Lineage, Machine Learning Algorithm wants to delete.

Process:

Step1: Read user input.

Step2: Set the value of k as following condition:

If number is divisible by 9, then set k=9

else

Set k= num%9.

end if

Step3: count number of digits in the given number and

set t= num_of_digit(num).

Step4: Calculate s as following:

s=10t-1

Step5: set i=s and g=0

Step6: Repeat:

If k=9 and i%9=0 then

Increment g by 1

Else if i%9= k then

increment g by1

end if

until: i<=num

Step7: calculate z as following:

Z=(100*g)+(t*10) +k

Step8: send z as encrypted number.

Algorithm 2: Decryption

Process:

Step1: Read encrypted number.

Step2: Calculate sum, count, pos and s:

Sum=num%10;

Count=(num/10)%10;

pos= num/100;

s= 10count-1

Step3: set i= s and g=0

Step4: Repeat:

If sum=9 and i%9=0 then

Increment g by 1

Else if i%9= sum then

increment g by1

end if

increment I by 1.

until: g<pos

Step7: set z=i-1;

Step8: send z as decrypted number.

Unlearning on Statistic query learning algorithm is completedas this updated model is similar to

$$\text{learn}\left(\sum_{i\neq a} f_1(x_a, l_{xa}), \sum_{i\neq a} f_2(x_a, l_{xa}), \ldots\ldots, \sum_{i\neq a} f_n(x_a, l_{xa})\right)$$

So, this model is computed based on the retrained data, i.e. excluding xa. This approach is very fast compared to retraining from scratch as computation is based on limited number of summations and F'_k is easily computed by subtracting $f_k(x_a, l_{xa})$fromF_k. Therefore, the whole model is computed very efficiently without having effect on the training data sets and maintaining the security by preventing attackers to pollute the data as there is authentication, which user has to pass through.

VI. Result Summary

We have applied this unlearning algorithm in 3 of the machine learning algorithms i.e. Naïve Bayes, K-means, Support vector machines.

All of them are successfully implementing this machine unlearning approach with machine learning level authentication with exactly the same precision i.e. not affecting the time as much and as this approach is based on number of summations which are very less.

This result set leads to the conclusion that the unlearning process including the data encryption is not much far, it's a faster approach, and have compatibility with most of the algorithms. So this techniqueprovides better usability, privacy of data and security from attackers.

VII. Future Perspective

Machine unlearning provides benefits to both the users and the service providers, with having more flexibility to data, users have more privacy and control over their data, so they can make changes in their existing data and are willing to share the data of their choice with the systems. This not only help users managing their data, but the business analyst, who make out decision based on customer choices, leading to current market trends. If data is more, service providers will also be benefitted because they'll have more profit opportunities by having useful insights of the data, and have few legal risks to their organizations.This enables the user to remove their personal or sensitive data, leading no residual lineage of data.Also enables the analysts, to help remove noise and incorrect entries by the user itself completely, ensuring the recommender systems to give useful results.It prevents the training data set and model from polluting by attackers and maintains user's security.

References:

1. Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani Witten, "An Introduction to Statistical Learning" ISSN 1431-875X, ISBN 978-1-4614-7137-0 –Springer.
2. Alex Armstrong: Poison Attacks Against Machine Learning, http://www.i-programmer.info/news/105-artificial-intelligence/4526-poison-attacks-against-machine-learning.html.

3. Technique that wipes out the unwanted data quickly and completely http://phys.org/news/2016-03-machine-unlearning-technique-unwanted-quickly.html.
4. "A CEO's Guide to BigData Security", http://www.infosecurity-magazine.com/magazine-features/a-ceos-guide-to-big-data-security/.
5. Yinzhi Cao and Junfeng Yang. "Towards Making Systems Forget with Machine Unlearning", Columbia University. In Proceedings of IEEE Symposium, September 2015.
6. Elisa Costante, Milan Petkovi´c, Yuanhao Sun, Jerry den Hartog, "A Machine Learning Solution to Assess Privacy PolicyCompleteness" WPES 2012: 91-96.
7. Dave Lewis. iCloud security questioned over celebrity photo leak 2014: Apple officially launches result of investigation over hacking.http://www.forbes.com/sites/davelewis/2014/09/02/icloud-data-breach-hacking-and-nude-celebrity-photos/#a00ac7b3f693.
8. The Editorial Board of the New York Times. "Ordering google to forget", https://www.nytimes.com/2014/05/14/opinion/ordering-googleto-forget.html? r=0.
9. Joseph A. Calandrino Ann Kilzer, Arvind Narayanan, E. W. Felten, and Vitaly Shmatikov. You might also like: Privacy risks of collaborative filtering. In Proceedings of 20th IEEE symposium on Security and Privacy, May 2011.
10. V. Woollaston. How to delete your photos from iCloud: Simple step by step guide to stop images going in wrong hands. http://www.dailymail.co.uk/sciencetech/article-2740607/Howdelete-YOUR-photos-iCloud-stop-getting-wrong-hands.html.
11. J. Sutter. Some quitting Facebook as privacy concerns escalate. http://www.cnn.com/2010/TECH/05/13/facebook.delete.privacy/, http://sickfacebook.com/15-million-us-facebook-users-have-deleted-their-accounts.
12. Vitaly Feldman, Varun Kanade "Computational Bounds on Statistical Query Learning", 25th Annual Conference on Learning Theory, COLT 2012.
13. By Vanessa Allen and Sian Boyle - Google axes 170,000 'right to be forgotten' links. http://www.dailymail.co.uk/news/article-2790326/google-deletes-18-000-uk-links-right-forgotten-laws-just-month-60-europe-wide-requests-come-fraudsters-criminals-sex-offenders.html.

14. Vitaly Feldman, IBM Research – Almaden. "Statistical Query Learning" (1993; Kearns).
15. Machine Unlearning: The Value of Imperfect Models, https://www.mapr.com/blog/machine-unlearning-value-imperfect-models.
16. Gert Cauwenberghs and Tomaso Poggio "Incremental and Decremental Support Vector Machine Learning", MIT press 2001.
17. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, An Introduction to Statistical Learning with Applications in R. Chapter 2, 6th printing (October 2015), Springer.
18. M. S. Hwang and Chi-yu Liu, "Authenticated encryption schemes: current status and key issues", International Journal of Network Security, Vol. 1, no. 2, pp. 61-73, 2005.