Privacy Aware Machine Learning and the "Right to be Forgotten"

by Bernd Malle, Peter Kieseberg (SBA Research), Sebastian Schrittwieser (JRC TARGET, St. Poelten University of Applied Sciences), and Andreas Holzinger (Graz University of Technology)

While machine learning is one of the fastest growing technologies in the area of computer science, the goal of analysing large amounts of data for information extraction collides with the privacy of individuals. Hence, in order to protect sensitive information, the effects of the right to be forgotten on machine learning algorithms need to be studied more extensively.

Data driven economy (and related concepts like Industry 4.0) and data driven science, as well as big data are the keywords most often heard in discussions on the future of high-profile industries and on the upcoming revolutions in the economic world. With the integration of modern information technology into "classical" industrial environments or services, many new opportunities can be envisioned, e.g., in the optimisation of supply chains or in on-demand production of specifically tailored goods, but even in governmental areas like health environments, where P4-medicine (predictive, preventive, personalised, participatory) is seen as a new paradigm that could revolutionise health care. With all these new opportunities, the challenges were traditionally located in the technical area, especially regarding technologies for enabling the efficient and correct analysis of the large amounts of data produced by factories and large sensor networks. In recent years, the area of machine learning has seen a surge in new technologies developed and brought to the market. In combination with the ever increasing amount of computational power and storage that is available for a relatively reasonable price, many of these applications can now be applied in real life environments.

While many of the technological issues have been apparently solved, the legal aspects of the collection and processing of vast amounts of data using machine learning algorithms has been neglected (see [1]). The "right to be forgotten" has been recently discussed with a particular focus on removing personal data and sensitive (personal) information from automated analysis if requested by an individual. This brings up technical as well as ethical questions. Especially in the European Union, the right to be forgotten has remained in political discussion, especially fed by a legal base for protection of personal information on the Internet by the European Commission (see [L1]), with the draft European Data Protection Regulation Article 17 (see [2]). While the main focus in the current discussion is leaning mostly towards the removal of information from search indexes of prominent search engines like Google, the underlying technological challenges run much deeper and touch fundamental aspects of machine learning and its application in the industry.

One of the major questions are the effects of removing information from knowledge bases on machine learning algorithms. This is especially important for algorithms or analytical systems that

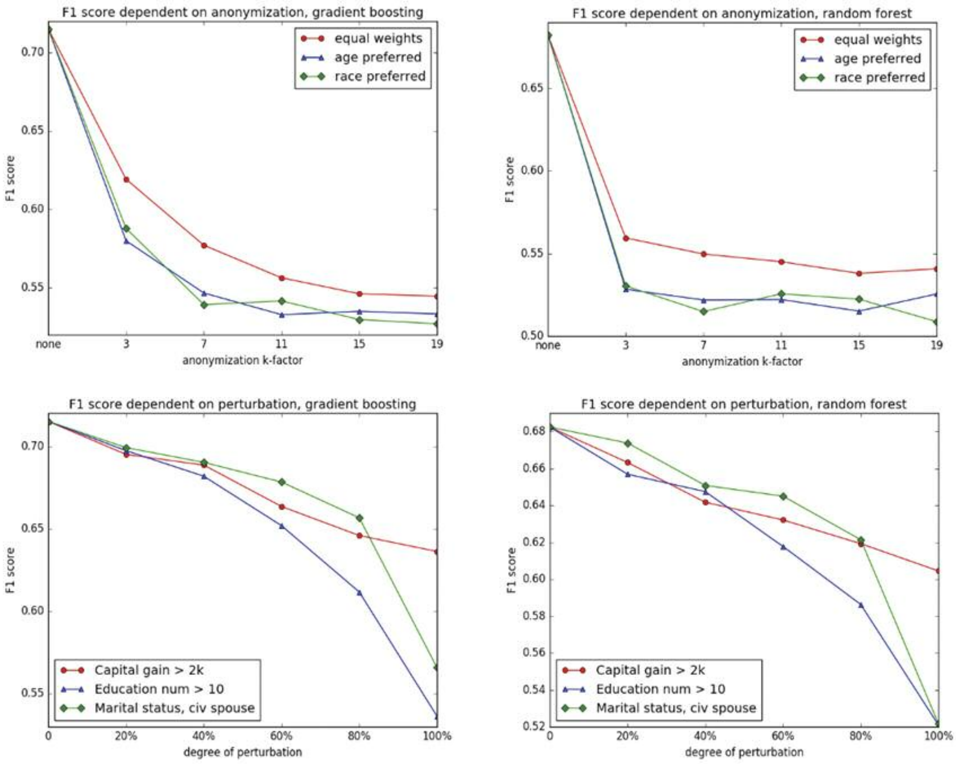

Figure 1: Effects of information removal on selected algorithms.

rely on a large knowledge base containing previously analysed data to learn from or which work by iteratively enhancing the global analytical result by continuously increasing the amount of information. Deletion of existing data thus tends to worsen the quality of the existing results, putting the organisations operating under such law at a serious disadvantage compared to organisations not subject to those restrictions. For instance, in developing clinical decision support software, companies not bound by the right to be forgotten can train their algorithms on a more comprehensive data set, giving them a worldwide advantage in marketing their product. In a project we studied the effects of selective deletion of valuable data items (in terms of their contribution to a classifier accuracy) as well as different levels of anonymization of a whole data set on machine learning algorithms [3]. In those experiments, which were based on data from the 1994 US-census with around 32,000 individual data records, the first phase tested four different classifiers (gradient boosting, linear SVC, logistic regression, random forest) with respect to precision, recall and F1-score. Subsequently, increasing fractions of valuable data were removed from the data set, which resulted in significant loss of classifier performance.

The other major research topic concerned the design of machine learning algorithms and applications that can cope with anonymized or otherwise generalised information. The fundamental idea is that the sensitive information is encoded by some privacy protecting means, analysed using machine learning algorithms and then prepared for inspection. Although several approaches exist for this strategy, they are currently not practical either due to their impact on the quality of the results, or due to the additional costs introduced:

- *Trusted environments*. While being the most popular strategy, the main issue of using a trusted environment lies in the large amount of resources that have to be reserved for this task, even when the analysis is only done very infrequently; e.g., using shared environments like the Cloud is not possible with this methodology. Furthermore, the environment needs to be set up with high security standards and a lot of audit and control mechanisms, as mechanisms for thwarting insider attacks must be introduced, including fully trusted human operators.
- *Anonymization*. The data set is transformed into a derived set that blurs the sensitive attributes without actually removing them altogether. Popular techniques work by generalising records until a certain minimal amount of them form an equivalence group (are indistinguishable).
- *Pseudonymization*. Related to anonymization, Pseudonymization works by removing sensitive attributes and replacing them with a placeholder, while keeping the internal logical structure of the data set, i.e., records with the same sensitive attributes get assigned the same pseudonym.
- *Functional Encryption*. Functional encryption allows the calculation of certain mathematical operations on the encrypted values, e.g., let $F(x)$ be the encryption function of x and F^{-1} (y) the decryption routing, then it holds true that $x+y=F^{-1}$ $(F(x)+F(y))$. While this works well in theory, currently available algorithms are very slow and thus cannot be used on close to all real life scenarios.

Our experiments to date (see [3] and [L2]) have focused on the loss of classifier performance when applied to anonymised knowledge bases. We used the same data set as described above and anonymised it using the k-anonymity criterion. The classifiers were then re-applied on a series of increasingly anonymised data sets (by increasing the k-factor), again resulting in significant losses of classifier performance. A noteworthy difference between selective deletion and anonymisation of data is that classifier performance on reduced data decreased rather slowly at the beginning and became more drastic with increased fractions of data removed, whereas for anonymised data sets the greatest loss occurred instantly and subsequently mellowed with increasing factors of k-anonymity (see Figure 1).

In conclusion, we can see that the effects of introducing the right to be forgotten to machine learning predictably results in losses of algorithmic performance. However, our experiments so far have only considered classification. A next logical step in our efforts would be the inclusion of predictors as well as unsupervised learning methods (clustering for automatic label provision, pattern / preference recognition for product design etc.). Even though in reality the effects might not be as drastic as produced by our initial experimental setting, even a few percentage points in ML performance could make a significant difference in crucial areas of application or highly competitive market environments. To sum up, we believe a lot of additional future research is needed in order to:

- fully understand the effects of the right to be forgotten on machine learning environments;
- be able to design algorithms more resilient to changes in the knowledge base;
- understand the effect of perturbing other forms of knowledge bases, e.g., graph based data sets, in which distance is derived from node feature vectors as well as associations.

Links:

[L1] http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf

[L2] http://www.hci-kdd.org/

References:

[1] P. Kieseberg, H. Hobel, S. Schrittwieser et al: “Protecting Anonymity in Data-Driven Biomedical Science”, pp. 301-316, 2014.

[2] P. De Hert, V. Papakonstantinou : “The proposed data protection Regulation replacing Directive 95/46/EC: A sound system for the protection of individuals”, Computer Law & Security Review 28, no. 2 (2012): 130-142, 2012.

[3] B. Malle, P. Kieseberg, E. Weippl, A. Holzinger: “The Right to Be Forgotten: Towards Machine Learning on Perturbed Knowledge Bases”, Workshop on Privacy Aware Machine Learning (PAML), August 2016

Please contact:

Peter Kieseberg
SBA Research, Vienna, Austria
pkieseberg@sba-research.org