Approximate K-Means on Encrypted Data

Jonathan Lennartz, Simon Schmitz, Priya Tomar, and Michael Nüsken

b-it, University of Bonn
Friedrich-Hierzebruch-Allee 6, D-53115 Bonn
nuesken@bit.uni-bonn.de
http://crypto.uni-bonn.de/~nuesken/

32th Crypto Day, 15 January 2021

Fully homomorphic encryption schemes allow arithmetic operations on ciphertexts without decrypting them first. The security notion of homomorphic encryption depends on the hardness of the ring learning with errors problem. As a result, the encryption works on approximate values with a certain error rather than the exact numbers. In highly regulated environments such schemes can be used to involve third parties in calculation processes without making sensible information visible to the system. While this allows applications in which data protection can be fully ensured, it is currently a resource intensive task. Fortunately, recent developments in the field allow for increasingly efficient procedures which we draw on to realize methods from the machine learning domain. Specifically, we aim to contribute by implementing and analyzing a fuzzy variation of classical k-means with reasonable runtime. Clustering methods have been implemented in a similar fashion before and especially the work done by Cheon, Kim & Park (2019) served as a strong reference for our group. Consequentially, we leverage the open-source HEAAN library for implementations and aim to extend it with additional atomic operations. While HEAAN can operate faster than comparable libraries, operations are restricted to polynomial calculations with a run time depending largely on the choice of security parameters. To achieve satisfactory results both in terms of security and speed, efficient polynomial approximation of more complex operations as well as extensive vectorization is important. Ultimately, our goal is to implement a clustering method that is comparable to Jäschke & Armknecht (2018) in performance and Cheon *et al.* (2019) in speed. For convenient testing, we re-implemented the HEAAN library for plaintext vectors and recently finished a first implementation of our clustering approach that accepts arbitrarily long inputs. Experiments on artificial data delivered promising results and we are currently revisiting some design decisions to justify them formally. This includes the analysis of alternative objective functions that can be directly realized without polynomial detours. Once this process reaches a satisfactory state, we plan to port the implementation to the actual HEAAN library and benchmark our results against data sets from the literature.

crypto day matters 32 (2021)
Beiträge vom 32. Kryptotag, 15.01.2021, online.
doi:10.18420/cdm-2021-32-41

References

Jung Hee Cheon, Duhyeong Kim & Jai Hyun Park (2019). Towards a Practical Cluster Analysis over Encrypted Data. 227–249. URL https://doi.org/10.1007/978-3-030-38471-5_10. See also https://ia.cr/2019/465.

Angela Jäschke & Frederik Armknecht (2018). Unsupervised Machine Learning on Encrypted Data. Number 2018/411 in Cryptology ePrint Archive. URL https://ia.cr/2018/411.