Towards Scalable, Efficient and Privacy Preserving Machine Learning
Rania Talbi, Sara Bouchenak

To cite this version:
Rania Talbi, Sara Bouchenak. Towards Scalable, Efficient and Privacy Preserving Machine Learning, Middleware ’18 Doctoral Symposium, Dec 2018, Rennes, France. <hal-01956155>

HAL Id: hal-01956155
https://hal.archives-ouvertes.fr/hal-01956155
Submitted on 14 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Minimize the computational costs incurred by privacy preservation.
- Provide an end-to-end privacy preserving outsourced data classification service.
- Enable a set of mutually untrusted data owners to have a global vision on the union of their data without breaching the privacy of each one of them.
- Enable dynamic data model updates when new training data samples are available.

Related work

Different ML algorithms
- Clustering
- Classification
- Association Rule Mining

Different Privacy-preservation objectives
- ML output protection
- Data protection

Privacy Preservation techniques
- Cryptographic techniques (SMC/HE, GC, OT)
- Non-cryptographic techniques (PP-Data Publishing techniques)

Design principles
- Decent privacy and utility levels
- Efficient runtime
- Entirely outsourced ML computations over encrypted data

We implemented the VFDT incremental decision tree learning algorithm [7]

Naive approach: a combination of low level PPML building blocks
1st optimization: use inline building blocks
2nd optimization: Parallel computing

References

4. [H. Wu et al.]: Privacy-Preserving SVM Classification on Vertically Partitioned Data. PAKDD 2006: 647-656.
5. [Tu et al.]: Outsourced privacy-preserving classification service over encrypted data. 1 Networks and Computer Applications 106: 100-110 (2018).
7. [M. Domingo et al.]: Mining high speed data streams. 42003: 71-80.