Student: Andrei Titu (4909941),
Responsible professor: Dr. Kaitai Lang
Supervisor: Rui Wang

Privacy-preserving vertical federated learning: A Literature Study

TU Delft

Background

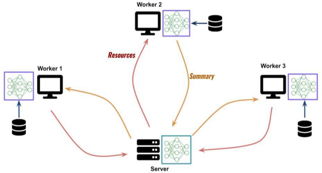
★ Defeating ***data silos***

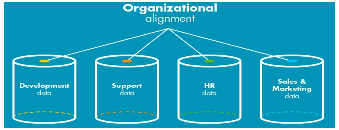
★ Empowering ***ownership*** of data

- ***Federated learning***: distributed machine learning technique based on moving resources to where data lies instead of transfering the data

- ***Vertical federated learning***: collaborating users' data-sets refer to the same objects but describe them in terms of distinct features

★ Privacy enhancing solutions:
Homomorphic encryption
Differential privacy





Research Questions

- What are the available technologies for vertical federated learning, and how are they implemented with respect to the performance vs. privacy trade-off?

- How do these methods compare in terms of efficiency, complexity, security and scenarios in that they would perform best?

Methodology

Multiple VFL technologies have been analysed in terms of:

1. Design
2. Computational overhead
3. Communication efficiency
4. Model accuracy
5. Security model and guarantees
6. Benefits and limitations
7. Potential improvements

Gathered information is used for comparing how these frameworks outperform each other and in which scenarios.

Study results

	Computational complexity	Convergence complexity	Communication costs	Model
FedBCD	$O(TKQ)$	$O(1/\sqrt{T})$	$O(\sqrt{T})$	SGD
MMVFL	$O(8Td^3/27)$	$O(1/T)$	$O(T)$	Classification
VAFL	$O(TK)$	$O(1/T)$	$O(T)$	SGD
Pivot	$O(Tncdbi)C_e$	$O(2/\sqrt{T})$	$O(2T)$	Decision Tree
FLOP	$O(TK)$	$O(1/T)$	$O(T)$	Classification

Table 1: Performance

	Threat model	Privacy techniques	Improvements
FedBCD	semi-honest	HE	-
MMVFL	semi-honest	-	FHE, FedBCD
VAFL	malicious	Gaussian DP	FedBCD
Pivot	semi-honest; cross-update colluding participants	HE(Paillier), DP	-
FLOP	semi-honest	-	FHE, MPC

Table 2: Privacy

Conclusion

- There is no 'one-fits-all' solution

- The VFL landscape is diverse and various problems require an educated balance between performance and privacy

- VFL allows parties to safeguard their data while achieving comparable costs & results to centralized alternatives

- There is continuous room for advancement and theoretical improvements have been proposed

Future work

- Combining technical contributions of similar frameworks may result in better security guarantees and higher performance. This should be experimentally pursued

- Theoretical enhancements proposed by this paper should be benchmarked in practice