

Strategic Data Sharing between Competitors

Nikita Tsoy Nikola Konstantinov

INSAIT, Sofia University
Sofia, Bulgaria

NeurIPS 2023

INSAIT

Motivation

- ML benefits operations

Motivation

- ML benefits operations
- ML heavily depends on data

Motivation

- ML benefits operations
- ML heavily depends on data
- Collaborative learning?

Collaboration Concerns

- Privacy (Kairouz et al., 2021)

Collaboration Concerns

- Privacy (Kairouz et al., 2021)
- Free-riding (Blum et al., 2021)

Collaboration Concerns

- Privacy (Kairouz et al., 2021)
- Free-riding (Blum et al., 2021)
- Heterogeneity (Donahue & Kleinberg, 2021)

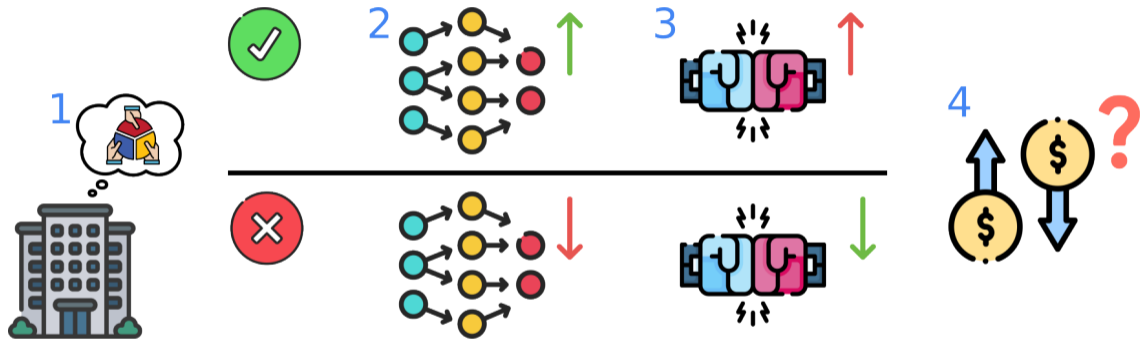
Collaboration Concerns

- Privacy (Kairouz et al., 2021)
- Free-riding (Blum et al., 2021)
- Heterogeneity (Donahue & Kleinberg, 2021)
- **Competition**

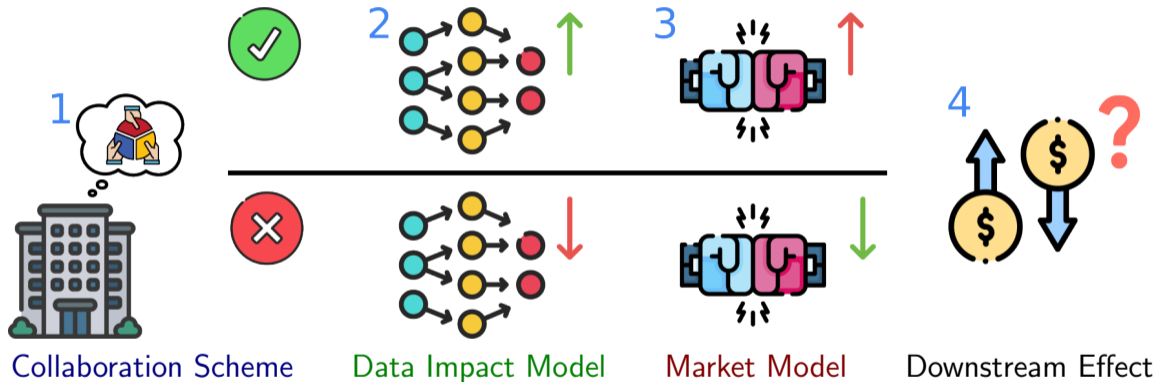
Research Question

When competitors can profit from collaborative learning?

Methodology



Methodology



Icons by Flaticon

General Framework

- 1 Collaboration Scheme
- 2 Data Impact Model
- 3 Market Model

Case Study

- 1 Full data sharing
- 2 Polynomial scaling law, $O(n^{-\beta})$
 - ▶ “Simplicity” parameter β
- 3 Conventional market model of Dixit (1979)
 - ▶ “Competitiveness” parameter γ

Results

Theorem (Informal)

$\exists x(\gamma, \beta)$ such that firms collaborate only if both of them have enough data

$$\forall i \frac{n_i}{n_1 + n_2} > x(\gamma, \beta).$$

- 1 $x(\gamma, \beta)$ is increasing in γ .
- 2 $x(\gamma, \beta)$ is increasing in β .

Results

- Collaborative learning is not always profitable

Results

- Collaborative learning is not always profitable
- Collaboration is more profitable when
 - ▶ The market is less competitive (small γ)
 - ▶ The learning tasks are harder (small β)

Results

We get similar results in different settings

Results

We get similar results in different settings

- Full data sharing between two firms

Results

We get similar results in different settings

- Full data sharing between two firms
- Partial data sharing between two firms

Results

We get similar results in different settings

- Full data sharing between two firms
- Partial data sharing between two firms
- Full data sharing between multiple firms

Conclusion

- Downstream effects are important

Conclusion

- Downstream effects are important
- Even if collaboration is easy, firms might not want to collaborate

Conclusion

- Downstream effects are important
- Even if collaboration is easy, firms might not want to collaborate
- Sometimes, firms want to collaborate even without incentivization

Thanks for your attention!

See us at Great Hall & Hall B1+B2 #1707 on Tue 12 Dec 5:15 p.m. CST

References I

- Blum, A., Haghtalab, N., Phillips, R. L., and Shao, H. One for one, or all for all: Equilibria and optimality of collaboration in federated learning. In *International Conference on Machine Learning*, pp. 1005–1014. PMLR, 2021.
- Dixit, A. A model of duopoly suggesting a theory of entry barriers. *Bell Journal of Economics*, 10(1):20–32, 1979.
- Donahue, K. and Kleinberg, J. Model-sharing games: Analyzing federated learning under voluntary participation. In *AAAI Conference on Artificial Intelligence*, volume 35(6), pp. 5303–5311, 2021.
- Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., and Cummings, R. Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning*, 14(1–2):1–210, 2021.