Efficient Shapley-based Data Valuation for Federated Trajectories

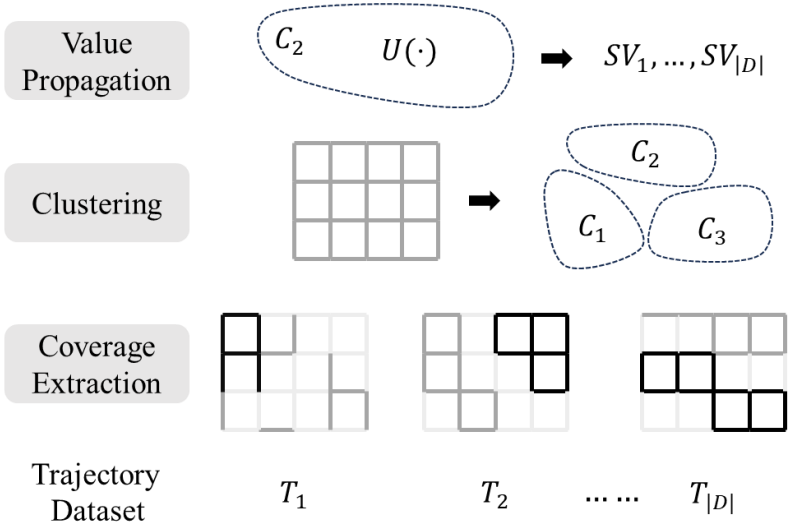
Yuxiang WANG, Shuyuan LI, Yongxin TONG, Shuyue WEI, Zimu ZHOU

Frontiers of Computer Science, DOI: 10.1007/s11704-025-51020-9

Problems & Ideas

- Problems of conventional centralized approaches:
  - In real-world applications, trajectory data is collected and owned by different entities, but data sharing remains limited due to the lack of effective incentives. To encourage contributions in a federated setting, it is critical to fairly compensate data owners for the value their data provides. While the Shapley value provides a provably fair foundation for this valuation, its exact computation is #P-complete, which makes it computationally intractable and impractical for large-scale trajectory datasets.

- Ideas:
  - The framework efficiently approximates the Shapley value using a clustering-based approach. It decomposes the total utility across road segment-time pairs, then clusters pairs with similar coverage patterns to avoid redundant computations by calculating the value once for each cluster's center and propagating it to other members.



Main Contributions

- Contributions:
  - We formally define the problem of Shapley-based data valuation for federated trajectories and prove that its exact computation is #P-complete;
  - We propose an efficient, clustering-based framework to approximate the Shapley value, which leverages decomposable utility across road segments to reduce redundant computations;
  - Experiments on real-world datasets show that our method achieves up to a 15.2x speedup and lower approximation error compared to existing baselines.

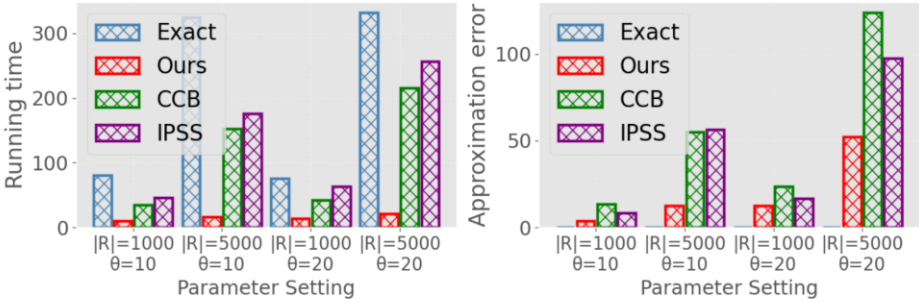


Fig. 1 Running time and approximation error of Shapley value computation for Beijing dataset with four parameter settings.

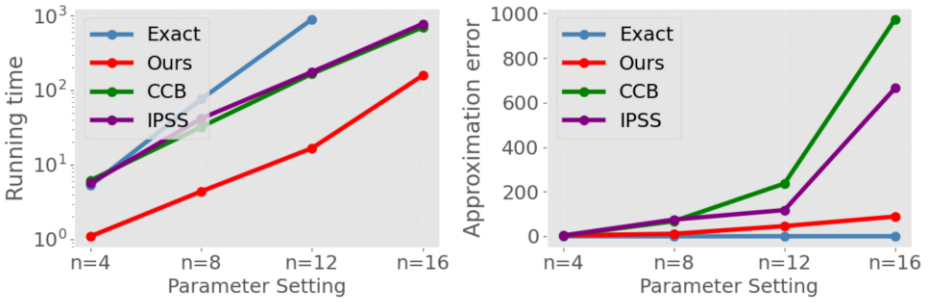


Fig. 2 Running time and approximation error of Shapley value computation for Chengdu dataset with different source number n.

We evaluated our method against three baselines: Exact, which computes the precise Shapley value according to its formal definition; CCB, a sampling-based estimation method; and IPSS, a stratified sampling-based algorithm for Shapley value estimation.