MobText-SISA: Efficient Machine Unlearning for Mobility Logs with Spatio-Temporal and Natural-Language Data

Haruki Yonekura[1,2], Ren Ozeki[1,3], Tatsuya Amano[1,2], Hamada Rizk[1,2,4], Hirozumi Yamaguchi[1,2]
[1]The University of Osaka, Japan
[2]RIKEN Center for Computational Science, Japan
[3]Japan Meteorological Agency, Japan
[4]Tanta University, Egypt
{h-yonekura, r-ozeki, t-amano, hamada_rizk, h-yamagu}@ist.osaka-u.ac.jp

Abstract

Modern mobility platforms have stored vast streams of GPS trajectories, temporal metadata, free-form textual notes, and other unstructured data. Privacy statutes such as the GDPR require that any individual's contribution be unlearned on demand, yet retraining deep models from scratch for every request is untenable. We introduce *MobText-SISA*, a scalable machine-unlearning framework that extends Sharded, Isolated, Sliced, and Aggregated (SISA) training to heterogeneous spatio-temporal data. *MobText-SISA* first embeds each trip's numerical and linguistic features into a shared latent space, then employs similarity-aware clustering to distribute samples across shards so that future deletions touch only a single constituent model while preserving inter-shard diversity. Each shard is trained incrementally; at inference time, constituent predictions are aggregated to yield the output. Deletion requests trigger retraining solely of the affected shard from its last valid checkpoint, guaranteeing exact unlearning. Experiments on a ten-month real-world mobility log demonstrate that *MobText-SISA* (i) sustains baseline predictive accuracy, and (ii) consistently outperforms random sharding in both error and convergence speed. These results establish MobText-SISA as a practical foundation for privacy-compliant analytics on multimodal mobility data at urban scale.

CCS Concepts

• **Networks** → *Location based services*; • **Security and privacy** → *Privacy Protection.*

Keywords

Privacy-preserving, Machine Unlearning, Right to be forgotten, Care Taxi

ACM Reference Format:
Haruki Yonekura, Ren Ozeki, Tatsuya Amano, Hamada Rizk, Hirozumi Yamaguchi. 2025. *MobText-SISA*: Efficient Machine Unlearning for Mobility Logs with Spatio-Temporal and Natural-Language Data. In *The 33rd ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL '25), November 3–6, 2025, Minneapolis, MN, USA*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3748636.3763226

SIGSPATIAL '25, Minneapolis, MN, USA

ACM ISBN 979-8-4007-2086-4/2025/11
https://doi.org/10.1145/3748636.3763226

1 Introduction

The profound demographic shift towards an ageing global population is placing unprecedented strain on health and social-care infrastructures. The World Health Organization (WHO) estimates that 142 million older adults are unable to live independently, with approximately two-thirds expected to require long-term care in their lifetime[11]. This demand is compounded by a projected global shortfall of 11 million health and care workers by 2030, further exacerbating the gap between need and supply[12]. To help mitigate this imbalance, care-oriented taxi services (Care Taxis) are being deployed worldwide to provide essential mobility assistance. For instance, Neighborly Care Network in Florida[1] facilitates attendance at senior day-care centers, while New Zealand's R&R Total Mobility[2] scheme subsidizes travel for older adults and persons with disabilities. Japan has similarly expanded its community-based transport services under its long-term-care insurance system.

The operational workflow of these services generates rich, longitudinal data. Providers dispatch nursing-care taxis, and each vehicle's onboard system logs timestamps for depot departures and arrivals at user residences. During dispatch, operators record the rider's profile (including age, gender, and mobility attributes like cane use or wheelchair dependence), along with free-text notes detailing special assistance needs. These multimodal logs constitute the dataset analyzed in this study.

Concurrently, the advent of large-scale foundation models is transforming geospatial analytics. Recent advancements like Google's Geospatial Reasoning models and the Trajectory-Powered Foundation Model of Mobility[3] promise to unlock insights from location traces for applications in public health and urban planning. However, these mobility traces contain highly sensitive information, such as home addresses and daily routines. This creates significant privacy risks, as machine learning models are vulnerable to attacks like membership inference (MIA)[10] and attribute inference (AIA)[13], which can expose private data from model outputs. The threat is amplified by recent jailbreak techniques that can bypass safety filters in large models, enabling malicious data extraction[4].

Against this backdrop of heightened risk, data protection regulations such as the GDPR have introduced the "Right to be Forgotten." This principle mandates that individuals can have their personal data erased and, critically, that its influence on trained models be eliminated. Machine unlearning is a computational paradigm designed to address this requirement by efficiently removing a specific user's contribution from a trained model.

[1]https://neighborly.org/
[2]https://randrmobility.co.nz/

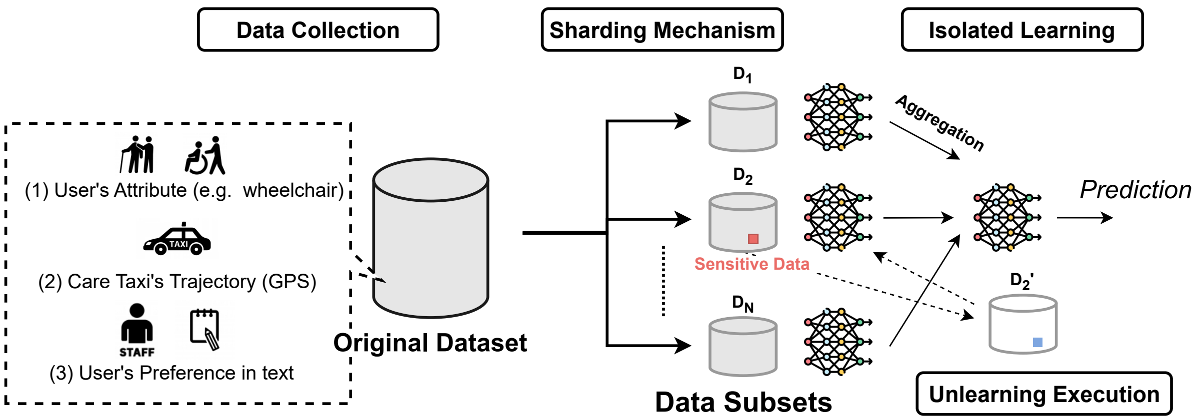

Figure 1: System Overview.

Among existing machine-unlearning frameworks, Sharded, Isolated, Sliced, and Aggregated (SISA) training [2, 6] offers a pragmatic solution. SISA first partitions the dataset into mutually exclusive shards and trains an independent constituent model on each shard. Each shard is then segmented into temporal slices; the constituent model is trained incrementally, with its parameters checkpointed immediately before a new slice is introduced. During inference, predictions from all constituents are aggregated to yield the ensemble output. When a data sample must be unlearned, only the constituent whose slice contains that sample is retrained from its last valid checkpoint, drastically reducing computational overhead compared with full-model retraining. Despite these efficiency gains, recent studies show that SISA's effectiveness hinges on the statistical similarity of the shards: when the data distributions differ markedly, constituent models learn shard-specific boundaries that fail to generalise, so the aggregated ensemble can underperform a monolithic baseline even when class proportions remain balanced[7]. In other words, performance degradation is driven by inter-shard distributional heterogeneity—loss of synergistic information, covariate shift, and feature sparsity—rather than by class imbalance alone. Moreover, partitioning inevitably reduces the amount of data available to each learner; small shards encourage underfitting or overfitting to idiosyncratic patterns. Yet little research has explored shard-partitioning strategies that preserve overall accuracy after partitioning. Proactive methods that explicitly minimise inter-shard divergence while maintaining unlearning locality therefore remain an open—and largely unexplored—research direction.

Leveraging real-world care-taxi trip logs, comprising location trajectories, mobility attributes, and textual notes, this study empirically investigates the consequences of data deletion on model performance and privacy. We specifically focus on how this inference capability changes after the corresponding user data is unlearned via the SISA framework, and examine how different sharding assignment strategies affect both unlearning efficiency and overall model fidelity. To this end, we introduce a similarity-aware shard-assignment strategy: before training, the multimodal records are clustered in a latent space, and each shard is populated with balanced samples drawn from every cluster. By homogenizing the latent data distribution across shards, this clustering-based allocation reduces inter-shard heterogeneity and thereby alleviates the accuracy loss identified in the preceding discussion.

Our contributions are summarized as follows. First, this is the first study to examine privacy threats and, in particular, deletion effects on machine learning models in the care-taxi domain from the perspective of machine unlearning. We discuss the application of the machine unlearning framework known as SISA and propose methodological extensions to realize machine unlearning efficiently. Second, in the context of machine unlearning, we evaluate practical applicability by handling multi-modal data comprising trip logs and associated textual information. Third, through a ten-month data collection and annotation process in collaboration with care providers, we have obtained a rich care-taxi dataset that can benefit the broader research community. Finally, we cast the empirical study as a supervised task that predicts the end-to-end pickup duration for each trip, leveraging the multimodal attributes collected by care providers.

2 Proposed Method

Figure 1 shows the overview of our proposed method. This method includes three modules and assumes an unlearning request.**Data Collection** is the process by which care service providers in Japan record data when picking up users using nursing care taxis. **Sharding Mechanism** partitions the collected dataset(Sharded dataset) to enable efficient learning in scenarios where unlearning is required. **Isolated Learning and Aggregation** refers to the procedure in which individual machine learning models are trained on each sharded dataset, and their outputs are aggregated to perform prediction with the overall model.

2.1 Data Collection

Care providers dispatch nursing-care taxis that follow home–facility and facility–home itineraries. For every trip, the onboard system records departure and arrival times as well as the precise geographic coordinates of the journey. Dispatch operators simultaneously store passenger profiles—age, gender, and mobility indicators such as cane use or wheelchair dependence—and optionally add free-text notes that describe special assistance (for example, "call five minutes before arrival" or "seat in the front passenger seat"). *Although such notes are rarely exploited in mobility analytics, we treat them*

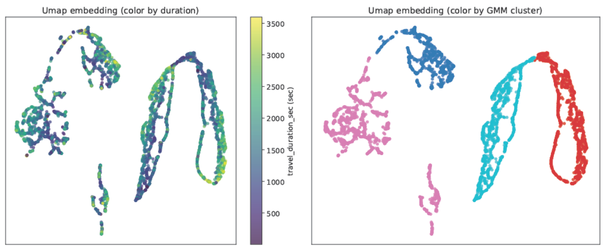

Figure 2: Umap Visualization[1] and Gaussian mixture model-based clustering.

as a first-class input modality and integrate them with the numeric features to enrich the representation of each trip instance. Such information is considered potentially useful for predicting the total duration required for the pickup process. We additionally include a naïve travel-time feature, computed by dividing the route length retrieved from OpenStreetMap's routing API[3] by the applicable speed limit.

2.2 Sharding Mechanism

The dataset is divided into k disjoint shards that can later be retrained independently when a right-to-be-forgotten request arrives. *The partitioning seeks two outcomes: first, any single deletion request should affect as few shards as possible; second, the aggregated predictor should retain high accuracy.* Every trip—numeric attributes plus free-text note—is mapped to a unified feature vector. The note is embedded with a frozen BERT encoder[5], then compressed to 32 dimensions via PCA; the result is concatenated with the scaled numeric fields. We project these vectors into a two-dimensional manifold with UMAP for visual inspection, as shown in Figure 2. A Gaussian mixture model[9] defined on this space provides a soft similarity metric among trips. *Clusters are traversed in a round-robin manner, one sample per shard, so highly similar trips land in distinct shards.* This spreads deletion impact while preserving representativeness, echoing the diversity-injection principle of CACTUS[8]. *Because each shard now receives only a fraction of the full dataset, sample scarcity can itself degrade model performance. Our balanced cluster-aware assignment mitigates this risk by ensuring that no shard is deprived of rare but informative patterns.*

2.3 Isolated Learning with Aggregation

Each shard feeds an identical model architecture that combines the structured fields with the cluster-aware text embedding described above. At inference time, the ensemble prediction is obtained through aggregation among constituent models. For all shard-specific models, we use a consistent baseline architecture consisting of a three-layer multilayer perceptron (MLP), and early stopping is applied to prevent overfitting. The classification task is defined at one-minute intervals, resulting in a large number of possible classes. To handle this fine-grained temporal granularity, we adopt an exponential label smoothing strategy. This approach assigns soft targets based on the exponential decay of distance from the true label in the time axis, enabling the model to learn smoother decision boundaries and tolerate minor deviations in prediction.

[3]https://project-osrm.org/

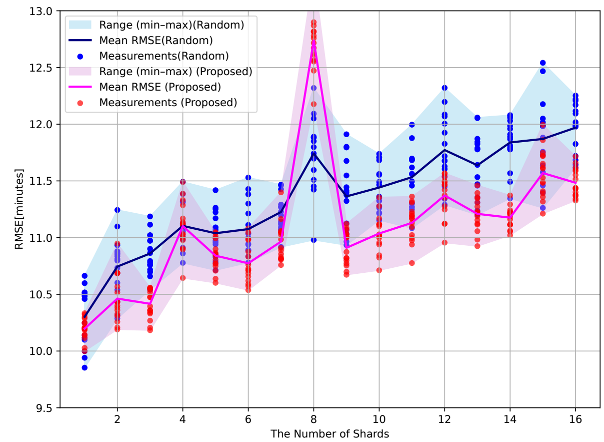

Figure 3: RMSE vs. The Number of Shards.

We choose to frame this task as classification rather than regression to improve learning stability. Regression is particularly sensitive to distribution shifts and heteroscedastic noise across clients, which can result in outlier-dominated updates and unstable optimization. In contrast, classification with normalized cross-entropy loss ensures bounded gradients and consistent aggregation, making it more robust under the data.

2.4 Unlearning Execution

When the system receives a right-to-be-forgotten request, it identifies the shard or shards that contain the target records and discards any checkpoints created after those records were incorporated. Only the affected constituent models are retrained, using exactly the retained data; all other shards and their models remain unchanged. *This procedure achieves exact unlearning while reducing computational cost compared to full retraining.* In the original SISA framework, this retraining process is further optimized through a technique called Slice, in which each shard is subdivided into multiple sub-datasets. Checkpoints are cached during training so that, upon a deletion request, retraining can resume from an intermediate state rather than starting from scratch. However, in this study, we do not employ the Slice mechanism or any caching strategy. Instead, we perform full retraining of the affected shard models from the initial state, without relying on any previously saved checkpoints. This allows us to isolate and evaluate the effects of our proposed shard partitioning strategy in a controlled and interpretable setting.

3 Evaluation

3.1 Dataset

To empirically validate our method, we utilize real-world pick-up and drop-off data provided by a specific care facility, spanning the period from October 1 to October 31, 2023, and consisting of 5193 unique records.

3.2 Evaluation of Prediction Performance

Figure 3 shows an evaluation of the relationship between the number of shards and the prediction error of the model. The evaluation

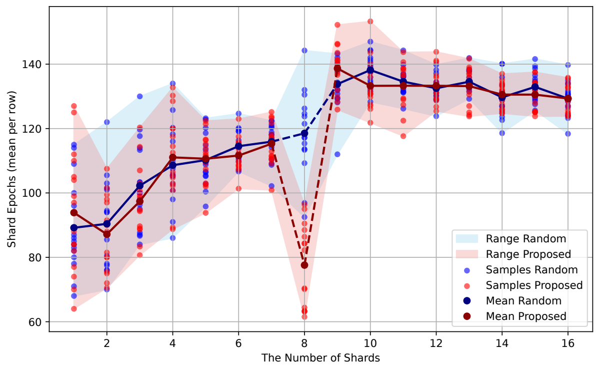

Figure 4: Training Epochs vs. The Number of Shards.

metric used in this study is the Root Mean Square Error (RMSE), and thus, the unit of error is expressed in minutes. It can be observed that the proposed method consistently outperforms the model trained on randomly partitioned datasets, regardless of the number of shards. A general trend is observed in which the overall prediction performance deteriorates as the number of shards increases. The proposed partitioning strategy mitigates the accuracy degradation that ordinarily accompanies an increasing number of shards by deliberately distributing dissimilar data samples across shards, thereby enabling each constituent model to learn more discriminative and complementary patterns. This facilitates robust classification even under limited data conditions, thereby alleviating the typical performance degradation associated with excessive data fragmentation.

3.3 Evaluation of Unlearning Scenario

Figure 4 shows the relationship between the number of shards and the average number of training epochs required for each partitioned model. In this evaluation, early stopping was applied based on the training data, enabling us to measure the convergence behavior of each model. In the context of machine unlearning, the number of training epochs can be interpreted as an indicator of how quickly a model can be retrained following a user's request for data deletion. While the overall difference between the proposed method and random partitioning is not substantial, the average and minimum number of epochs per shard are lower for the proposed method in a majority of cases, specifically, for shard counts ranging from 2 to 16, excluding the exceptional case when the number of shards is 8. This suggests that the proposed partitioning strategy is marginally more efficient than random partitioning in terms of retraining speed under deletion constraints.

Furthermore, across both Figures 3 and 4, the case with the 8-shard setting emerges as an outlier for the proposed method. This anomaly is likely attributable to two intertwined factors: (i) premature convergence of the constituent models to local optima, which triggers early termination of training, and (ii) the possibility that the Gaussian-mixture-model clustering failed to yield coherent clusters at this particular shard count, thereby degrading the subsequent partitioning quality.

4 Conclusion

In this study, we investigated a data shard partitioning strategy to enable efficient machine unlearning on mobility data containing linguistic information, using the SISA framework. Our method performs clustering on the two-dimensional coordinates obtained by projecting compressed input representations, via Uniform Manifold Approximation and Projection (UMAP), followed by Gaussian Mixture Modeling (GMM). The resulting clusters are then distributed across shards in a round-robin manner. This approach preserves the underlying data distribution within each shard, allowing models trained on reduced datasets to maintain predictive accuracy without significant degradation, even after partitioning.

While this work focuses on the data partitioning methodology, further research is warranted regarding the aggregation of machine learning models trained independently on each shard. Effective aggregation techniques may enhance both the performance and robustness of the overall system under deletion and retraining constraints.

Acknowledgments

This work was supported by JST, CREST Grant JPMJCR21M5, JST PRESTO Grant JPMJPR2361, and JST BOOST, Japan Grant Number JPMJBS2402.

References

[1] Etienne Becht et al. 2019. Dimensionality reduction for visualizing single-cell data using UMAP. *Nature biotechnology* 37, 1 (2019), 38–44.
[2] Lucas Bourtoule et al. 2021. Machine unlearning. In *2021 IEEE symposium on security and privacy (SP)*. IEEE, 141–159.
[3] Shushman Choudhury et al. 2024. Towards a Trajectory-powered Foundation Model of Mobility. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Spatial Big Data and AI for Industrial Applications* (Atlanta, GA, USA) *(GeoIndustry '24)*. Association for Computing Machinery, New York, NY, USA, 1–4.
[4] Gelei Deng et al. 2024. MASTERKEY: Automated Jailbreaking of Large Language Model Chatbots. In *NDSS*.
[5] Jacob Devlin et al. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*. 4171–4186.
[6] Yaron Kanza, Balachander Krishnamurthy, and Divesh Srivastava. 2024. A Geospatial Perspective on Data Ownership, the Right to be Forgotten, Copyrights, and Plagiarism in Generative AI. In *Proceedings of the 32nd ACM International Conference on Advances in Geographic Information Systems* (Atlanta, GA, USA) *(SIGSPATIAL '24)*. Association for Computing Machinery, New York, NY, USA, 477–480.
[7] Korbinian Koch and Marcus Soll. 2023. No matter how you slice it: Machine unlearning with sisa comes at the expense of minority classes. In *2023 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*. IEEE, 622–637.
[8] Mohammad Mehdi Rastikerdar et al. 2024. CACTUS: Dynamically Switchable Context-aware micro-Classifiers for Efficient IoT Inference. In *Proceedings of the 22nd Annual International Conference on Mobile Systems, Applications and Services* (Minato-ku, Tokyo, Japan) *(MOBISYS '24)*. ACM, New York, NY, USA, 505–518.
[9] Douglas Reynolds. 2009. Gaussian mixture models. In *Encyclopedia of biometrics*. Springer, 659–663.
[10] Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. 2017. Membership inference attacks against machine learning models. In *2017 IEEE Symposium on Security and Privacy (SP)*. IEEE, 3–18.
[11] World Health Organization. 2022. Long-term care. https://www.who.int/europe/news-room/questions-and-answers/item/long-term-care. Accessed: May 15, 2025.
[12] World Health Organization. 2025. Health workforce. https://www.who.int/health-topics/health-workforce. Accessed: May 15, 2025.
[13] Benjamin Zi Hao Zhao et al. 2021. On the (In)Feasibility of Attribute Inference Attacks on Machine Learning Models . In *2021 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE Computer Society, Los Alamitos, CA, USA, 232–251.