Scholar Publishing Group
International Journal of Business Management and Economics and Trade
https://doi.org/10.38007/IJBMET.2026.070104
ISSN 2790-0959 Vol. 7, Issue 1:28-35

Differentially Private High-Dimensional Business Data Publishing and Analysis Algorithm

Yiting Hong

Forecasting & Purchasing & QC Department, The Antigua Group, Peoria, 85382, Arizona, US

Keywords: Differential privacy, High dimensional positional data, Location based social networks (LBSNs), Dynamic privacy budget allocation, User preference modeling

Abstract: This research focuses on the balance between privacy protection and data practicality of high-dimensional location data in the mobile Internet and LBS scenarios. Traditional anonymization and perturbation methods suffer from precision loss, weak dynamic adaptability, and neglect of personalized user needs. Based on the rigor of differential privacy theory, two innovative solutions are proposed: firstly, the graph automatic encoding method integrates user social relationships and spatial behavior, and achieves joint protection of trajectory and topology through dynamic privacy budget allocation; Secondly, building personalized privacy configurations based on user preferences, allocating privacy budgets differentially through stopping point clustering and sensitivity scoring, and enhancing service accuracy while strengthening protection in sensitive areas. Experimental verification shows that both perform excellently in terms of privacy protection strength, data utility (improved by over 10%), and operational efficiency. Dynamic budget allocation and user preference modeling are key to balancing privacy and practicality. In the future, intelligent parameter adjustment, large-scale scene adaptation, and compliance verification will be explored.

1. Introduction

The research on high-dimensional business data publishing and analysis algorithm based on differential privacy[1] focuses on solving the problem of balancing privacy disclosure risk and data practicality faced by high-dimensional location data in the sharing process under the background of the rapid development of mobile Internet and location services (LBS)[2]. With the continuous expansion of the satellite navigation and location service industry, user location data is widely used in navigation, social, recommendation and other scenarios. However, traditional privacy protection technologies (such as data anonymization and perturbation methods) suffer from accuracy loss, insufficient dynamic adaptability, and neglect of users' personalized needs for privacy protection. Differential privacy, with its theoretical rigor and universality, has become the core mechanism for location privacy protection. However, existing research still has significant shortcomings in terms of feature encoding depth, dynamic response to user preferences, flexible regulation of privacy budgets, and user autonomy in decision-making. This study proposes two innovative directions: one

Copyright: © 2026 by the authors. This is an Open Access article distributed under the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (https://creativecommons.org/licenses/by/4.0/).

is to construct a graph automatic encoding method based on differential privacy, which integrates user social relationships and spatial behavior patterns through graph structure modeling, and introduces a dynamic privacy budget allocation mechanism in the embedding layer to achieve joint protection of trajectories and topology in location-based social networks (LBSNs); The second is to integrate user preference modeling, construct personalized privacy profiles through stopping point clustering and sensitivity scoring, design differentiated privacy budget allocation strategies, and improve service accuracy while ensuring privacy in highly sensitive areas. This method strikes a balance between privacy, availability, and robustness, providing theoretical support and technical path for the secure release and intelligent analysis of high-dimensional commercial data.

2. Correlation theory

2.1 Hybrid Design and Key Enhancement Mechanisms for Privacy Protection Architecture of LBS System

In the privacy protection system based on location-based services, the independent architecture[3] achieves efficient communication and reduces centralized attack risks through direct interaction between mobile terminals and LBS servers, but relies on terminal hardware performance and LBS server security; The centralized architecture [4] introduces a third-party trusted anonymous server to perform identity anonymization and query obfuscation, reducing the computational burden on terminals, but facing the risk of single point overload and failure; Distributed architecture adopts P2P network to distribute processing tasks, improving system scalability and robustness, but terminals need to bear higher computing loads and are vulnerable to malicious node attacks. The hybrid architecture combines the advantages of centralization and decentralization, dynamically switching modes to adapt to system loads and security requirements - small-scale scenarios adopt centralized guarantee for efficient management, while high-risk or large-scale scenarios switch to distributed and decentralized loads. To enhance the depth of privacy protection, this architecture further integrates graph structure features to capture high-order associations between user social relationships and spatial behavior, avoiding social graph re identification attacks; At the same time, introducing stop point information to identify high-frequency user stay areas (such as residential and workplace), combined with user sensitivity ratings to achieve differentiated allocation of privacy budgets, while ensuring privacy in highly sensitive areas and improving service response quality, forming a collaborative protection mechanism of dynamic switching at the architecture layer and personalized regulation at the semantic layer.

2.2 The core mechanism and theoretical characteristics of differential privacy model

Differential privacy, as the core technology of modern data privacy protection, reduces the impact of individual data on output by introducing random noise into query results, achieving a balance between privacy protection and data practicality. Its core definitions include adjacent datasets (only differing by one data point), ε - differential privacy[5] (limiting the output differences of adjacent datasets through probability ratios), and global sensitivity (quantifying the maximum impact of a single record on query results). In terms of implementation mechanism, Laplace mechanism is suitable for real value queries, and privacy requirements are met by adding Laplace noise with scale parameter $b=\Delta f/\varepsilon$; The index mechanism is designed for discrete scenarios and constructs a probability distribution based on the sensitivity Δu of the scoring function to select the optimal solution. In theory, differential privacy has composability (sequential combination of privacy budget accumulation, parallel combination taking the maximum value) and

post-processing immunity (any subsequent processing does not weaken privacy protection). These characteristics support its flexible application in multiple scenarios such as statistical queries, machine learning, recommendation systems, etc., providing a rigorous privacy protection framework for high-dimensional data publishing and analysis.

3. Research method

3.1 Location privacy protection technology system integrating automatic image encoding and differential privacy fusion

In the context of LBS and social network integration, graph autoencoder technology models user social relationships and location activity trajectories through graph structure, combines graph autoencoder (GAE) to achieve low dimensional representation learning of nodes, and relies on differential privacy or noise injection (such as Gaussian and Laplacian noise) to protect individual privacy during encoding and decoding while maintaining data analysis capabilities. Location privacy protection technology covers multidimensional strategies: Gaussian noise (suitable for continuous value smoothing scenarios) and Laplacian noise (suitable for real-time requirements scenarios) are used for location data perturbation, balancing privacy protection and data accuracy by adjusting noise parameters; Data aggregation and anonymization techniques utilize weighted averaging, group aggregation, density aggregation, k-anonymity, spatial partitioning, location blurring, and other methods to reduce individual identifiability through group characteristics, making them suitable for scenarios such as heat map generation and traffic analysis; Differential privacy location queries use a dynamic privacy budget allocation strategy to adjust the privacy budget (ε) based on query frequency, sensitivity, data distribution, and service accuracy requirements, maximizing query accuracy while protecting privacy. This technology system achieves a collaborative balance between privacy protection and data availability through graph structure feature fusion, noise intensity optimization, and adaptive mechanism, providing theoretical support and technical path for privacy security and intelligent services in LBSN.

3.2 Privacy Protection Mechanism of DP-GAE

With the evolution of mobile Internet and location technology, location information has become a core application element in social networks, but the sensitive information contained in graph structure data, such as node identity, connection relationship and behavior trajectory, faces privacy risks such as re identification attacks and trajectory inference. Traditional anonymization methods are difficult to effectively hide user identities in graph structures, while deep learning models (such as graph neural networks) can capture complex structural relationships, but there is a "memory effect" that may lead to privacy breaches. To this end, this chapter proposes a location map automatic encoding method based on differential privacy (DP-GAE), which models user location and social relationships through graph structure, combines graph attention network (GAT) to achieve low dimensional representation learning of nodes, and satisfies differential privacy constraints (such as ε - privacy budget) through gradient pruning and Gaussian noise injection in the encoding stage to ensure that gradient updates do not leak sensitive information; In the decoding stage, privacy protection and data availability are balanced by reconstructing node features [6](minimizing feature two normal form loss) and graph structure (maximizing similarity between adjacent node representations). The model problem is defined as a graph G=(V, E, X), where nodes represent dwell points (including latitude and longitude, dwell time, environmental features, etc.), edges are connected based on geographic proximity or time series, and the target learns a low dimensional latent representation Z to preserve topological structure and node features.

Experimental verification shows that this method effectively maintains the accuracy and robustness of location data analysis while resisting threats such as re identification attacks and trajectory inference, providing theoretical support and technical path for privacy protection in high-dimensional commercial data publishing.

3.3 Design of DP-GAE Differential Privacy Graph Autoencoder Position Protection Model

This study proposes the DP-GAE algorithm, which consists of three stages: location map structure construction, encoding, and decoding. In the construction of the location map structure, user dwell points S=(lat, lon, arr, dep) are extracted from dataset D, and dwell duration (duration=dep arr), daily period (period, identifying day/night), and environmental features (env, obtained by querying surrounding interest point types through OpenStreetMap Overpass API) are generated through feature engineering. Nodes represent dwell points, and edges are established based on geographic proximity (distance<500 meters) or user movement sequence (continuous positioning time interval<threshold), forming graph G=(V, E, X). In the encoding stage, Graph Attention Network (GAT) is used, and node i is represented at the l-th layer as

$$h_i^{(l)} = \sigma(\sum_{j \in N(i) a_{ij}} W h_j^{(l-1)})$$

to meet differential privacy, L2 norm clipping is applied to the gradient g-t and Gaussian noise is added, where the noise variance is determined by the privacy budget ϵ. The decoding stage reconstructs node features through reconstruction

$$L_{feat} = \frac{1}{N} \sum_{i=1}^{N} \left\| h_i^L - \hat{h}_1 \right\|_2^2$$

structure with diagram

$$L_{graph} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j \in N(i)} \left\| h_i^L - \hat{h}_J \right\|_2^2$$

optimize the model, with a total loss function of

$$L_{total} = \lambda_1 L_{feat} + \lambda_2 L_{graph}$$

regulating the ratio of node features to graph structure loss through hyperparameters. This design maintains the accuracy and robustness of location data analysis while protecting user privacy.

4. Results and discussion

4.1 The impact of privacy budget on DP-GAE performance and multi model comparison

This chapter validates DP-GAE performance on the Geolife and Gowalla datasets. The experiment adopts the GAT encoder decoder architecture and trains for 200 epochs (learning rate of 0.01, Adam optimizer, 10 cross validations). The evaluation indicators include privacy leakage degree (quantified through gradient backpropagation attack), differential privacy noise level, node feature reconstruction error, graph structure reconstruction error, training time, and convergence speed. As shown in Table 1

When ε =0.1, the privacy leakage risk is 3.2% but the reconstruction error is high (MSE=0.48). When ε =1.0, the optimal balance is achieved (leakage risk 9.5%, prediction accuracy 74.8%, graph structure preservation 85.6%). Comparison with baseline model (as shown in Table 2)

DP-GAE outperforms GAP (10.7%/72.4%), k-anonymity (4.5%/65.3%), and PPKS (6.8%/68.7%) in terms of privacy breach risk (5.2%) and prediction accuracy (74.8%), approaching unprotected GATE (20.3%/78.2%) but possessing privacy protection capabilities. Training efficiency analysis shows that when ε is small, the training time increases (520s vs 430s) and the

convergence speed slows down (80 vs 50 rounds). The privacy utility trade-off curve [7] shows that the optimal balance between privacy and data availability is achieved when ε =1.0. Experimental verification of ablation: Removing the DP mechanism increased the risk of leakage to 19.8%, removing the GAT structure reduced the accuracy to 70.2%, and using only the MLP encoder resulted in an accuracy of 72.1%, both inferior to the complete DP-GAE (5.2%/74.8%).

Title 1 Impact of Privacy Budget ε on Model Performance

ε Value	Privacy Leakage Risk (%)	Node Structure Reconstruction Error (MSE)	Graph Structure Retention (%)	Prediction Accuracy (%)
0.1	3.2	0.48	76.4	65.1
0.5	5.8	0.41	81.2	70.3
1.0	9.5	0.35	85.6	74.8
5.0	15.2	0.42	89.3	72.4
10.0	18.9	0.50	92.1	69.7

Title 2 Performance Comparison of Different Models in Location Data Tasks

Model Name	Privacy Leakage Risk (%)	Structure Retention (%)	Prediction Accuracy (%)	Reconstruction Error (MSE)
GATE	20.3	92.5	78.2	0.28
GAP	10.7	85.1	72.4	0.36
k-Anonymity	4.5	73.2	65.3	0.49
PPKS	6.8	78.9	68.7	0.43
DP-GAE	5.2	85.6	74.8	0.35

4.2 Model experiment

This study explores the emerging privacy protection needs brought about by the deep integration of spatial information technology and social services. The advancement of geographic information systems (GIS)[8], global positioning systems (GPS), and collaborative applications has significantly improved the accuracy and efficiency of geospatial data collection, promoted the development of location-based social networks (LBS), and achieved innovative functions such as geofencing social matching, spatiotemporal behavior sharing, and regional service recommendation. However, users face privacy and security challenges while enjoying precise location services; In a semi honest service architecture, attackers can use legitimate data interfaces for trajectory reconstruction attacks, thereby risking reverse parsing of user spatial behavior characteristics. The existing privacy protection paradigms, including l-diversity, k-anonymity, location ambiguity, and differential privacy, have limitations in both theory and practice: l-diversity focuses on discrete sensitive attributes, but ignores the risks of spatiotemporal behavioral dimensions; K-anonymity may dilute group characteristics, leading to identity re identification vulnerabilities in context aware attacks; Fuzzy positioning reduces positioning accuracy, but also lowers service matching efficiency; Although differential privacy is mathematically rigorous, it may lead to semantic distortion in small sample scenarios, and there is a need to improve the dynamic privacy budget allocation for personalized services. The current research gaps include: excessive emphasis on physical space protection in anonymization technology, lack of deep integration with user behavior modeling and context aware technology; There is a lack of dynamic balance model between personalized service maintenance and privacy protection intensity. To address these issues, an intelligent perception approach proposes a driven adaptive privacy protection paradigm that combines multidimensional user analysis and dynamic privacy parameter optimization to ensure sensitive spatial information is

protected while meeting the requirements of personalized location services and system response efficiency. This chapter introduces the User Preference and Differential Privacy Location Privacy Protection (UPDP-LPP) method, which has three key contributions: (1) a semantic based stopping point type detection clustering algorithm that uses data from a real POI dataset to determine the type of each stopping point; (2) Select candidate privacy aware location generalization algorithms based on the following factors, keep them within the radius threshold, allocate privacy budget for the radius, and inject Laplacian noise to protect location privacy; (3) Dynamic privacy budget allocation based on user privacy preferences to ensure the universal and effective use of reserved points. By combining behavioral semantic analysis with dynamic privacy regulation, this method strikes a balance between privacy protection and data utility and accuracy. The experiment on two real-world datasets validated its effectiveness in protecting user privacy while maintaining good data utility.

4.3 Effect analysis

This chapter evaluates the proposed User Preference and Differential Privacy Location Privacy Protection (UPDP-LPP) method through experimental analysis. The experimental environment is built on the PyCharm platform, using a social network dataset based on Geolife location and a point of interest (POI) dataset. The former contains 17621 movement trajectories of 182 users over three years, including attributes such as latitude, longitude, altitude, and timestamp. Data filtering is applied to focus on geographic ranges corresponding to city scale location services, ensuring consistency with typical application scenarios. The functional implementation includes three key algorithms: the stopping point extraction algorithm, which uses sliding window technology with time and distance thresholds (20 minutes and 200 meters) to identify meaningful positions; A semantic based stopping point type detection algorithm that integrates POI data through spatial clustering and weighted association to assign semantic labels (e.g. commercial services, spatial functions); And the UPDP-LPP algorithm itself, which dynamically allocates privacy budget based on user access frequency and injects Laplacian noise to summarize dwell points while maintaining type consistency. The performance comparison with three baseline methods (TLDP, DPLPA, LPPM) in terms of privacy protection level, data utility, and runtime indicators shows that UPDP-LPP achieves excellent privacy protection (reaching the highest privacy level under different privacy budgets and access frequencies), enhances data utility (increasing by more than 10% through dynamic budget allocation and noise optimization), and reduces computational overhead (with the shortest runtime due to effective privacy parameter tuning). The experimental results have verified that this method effectively balances privacy protection with service quality and accuracy. Future research directions include improving user behavior modeling through deep reinforcement learning, developing collaborative frameworks for privacy budgeting and geographic parameter optimization, exploring heterogeneous integration of multi-source privacy protection technologies to enhance adaptability in complex scenarios.

5. Conclusion

With the development of mobile communication and intelligent sensing technology, location aware services have been deeply integrated into multimodal application scenarios, building an intelligent service ecosystem. However, the increasing risk of location privacy breaches and user privacy concerns have constrained the healthy development of location-based services (LBS). This study focuses on high-dimensional commercial data publishing and analysis scenarios, and proposes two core methods based on differential privacy: one is the location map autoencoder method, which models user location information and social relationships into a graph structure through a graph

autoencoder (GAE), uses differential privacy to add noise to the graph embedding layer to achieve privacy protection, and dynamically adjusts the privacy budget to balance data privacy and model performance; The second is a personalized privacy protection method that integrates user preferences[9]. Through a privacy budget allocation strategy[10], the noise injection intensity is adjusted based on the user's importance preference for location, achieving a personalized balance between privacy protection and data availability. Experimental verification shows that both methods have significant effects in reducing privacy risks, improving the accuracy of graph data analysis, and enhancing user experience. Future research will explore intelligent dynamic adjustment of privacy parameters to adapt to preference changes, optimize the integration efficiency of graph neural network structure and differential privacy mechanism, and verify scalability and robustness in large-scale datasets and complex network environments. At the same time, it will combine privacy protection laws and ethical frameworks to ensure technical compliance, providing theoretical support and practical paths for high-dimensional commercial data security release and intelligent analysis in fields such as smart cities and big data analysis.

References

[1] *Ni G, Sun J. Differential privacy protection algorithm for large data sources based on normalized information entropy Bayesian network. 2024.*

[2] *Sheng J, Zhang L, Zhang Y, et al. A Localization Correction Algorithm of Location-Based Services Based on Point Clustering[C]//International Conference on Data Mining and Big Data. Springer, Singapore, 2024.*

[3] *Li L, Flynn T, Hoisie A. Learning Independent Program and Architecture Representations for Generalizable Performance Modeling. 2023.*

[4] *Sanwal S. Design and Architecture for a Centralized, Extensible, and Configurable Scoring Application. 2023.*

[5] *Huang, J. (2025). Research on Cloud Computing Resource Scheduling Strategy Based on Big Data and Machine Learning. European Journal of Business, Economics & Management, 1(3), 104-110.*

[6] *Lu, Z. (2025). AI-Driven Cross-Cloud Operations Language Standardisation and Knowledge Sharing System. European Journal of AI, Computing & Informatics, 1(4), 43-50.*

[7] *Alikhanifard P, Tsantalis N. A Novel Refactoring and Semantic Aware Abstract Syntax Tree Differencing Tool and a Benchmark for Evaluating the Accuracy of Diff Tools. ACM Transactions on Software Engineering and Methodology, 2025, 34(2).*

[8] *Zhang Z, Song Y. Spatial Big Data and Analysis Strategies Supporting Geographic Information System for Transportation (GIS-T) in Conceptual Design, Modelling, and Decision-making: A Review. ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, 2024, 10(4).*

[9] *Li, W. (2025). Research on Optimization of M&A Financial Due Diligence Process Based on Data Analysis. Journal of Computer, Signal, and System Research, 2(5), 115-121.*

[10] *Xu, H. (2025). Research on the Implementation Path of Resource Optimization and Sustainable Development of Supply Chain. International Journal of Humanities and Social Science, 1(2), 12-18.*

[11] *Li Y, Song X, Liu T M. GAPBAS: Genetic algorithm-based privacy budget allocation strategy in differential privacy K-means clustering algorithm. Computers & Security, 2024, 139(Apr.): 103697. 1-103697. 12.*

[12] *Lu, Z. (2025). Design and Practice of AI Intelligent Mentor System for DevOps Education. European Journal of Education Science, 1(3), 25-31.*

[13] Yu, X. (2025). Application Analysis of User Behavior Segmentation in Enhancing Customer Lifetime Value. Journal of Humanities, Arts and Social Science, 9(10).

[14] Zheng, H. (2025). Research on Lifecycle Configuration and Reclamation Strategies for Edge Nodes Based on Microservice Architectures. Advances in Computer and Communication, 6(5).

[15] Li, J. (2025). The Impact of Distributed Data Query Optimization on Large-Scale Data Processing.