Bidirectional LSTM-Based Privacy Preserving Method for Trajectory Generation

Xiangjie He, Tingting Gao, Yihan Yin, Wei Jiang
Institute of Computer Science and Information Engineering, Harbin Normal University, Harbin, China

Abstract: To ensure the privacy of trajectory data while improving its usability upon release, machine learning algorithms can be applied to process trajectory data, thereby enhancing its usability. Addressing the issue of trajectory data release usability, we propose a trajectory privacy protection scheme that combines Bidirectional Long Short-Term Memory (BILSTM) networks and differential privacy (DP). The scheme prepossess the trajectory data using BILSTM to improve its usability. For the generated trajectory data, the Laplace mechanism in differential privacy is applied to add noise, thereby achieving privacy protection. The generalized trajectory data-set obtained is then released. This scheme ensures good data usability and offers certain efficiency advantages.

Keywords: Trajectory Privacy; Neural Networks; Trajectory Generation

1. Introduction

With the rapid development of location-aware technology and wireless communication technology, location service providers can collect and store a large amount of valuable trajectory data. This data has important applications in various fields, especially in analyzing and predicting human movement patterns and the spread of the COVID-19 pandemic in recent years. Additionally, trajectory data is widely used in intelligent transportation and other fields. However, directly using real trajectory data for analysis and research poses a risk of user privacy leakage. To address this issue, researchers have proposed using synthetic trajectory data to replace real data for analysis [1].

Currently, research on generating synthetic trajectories mainly relies on machine learning methods. Deep generative models can effectively capture the long-range dependencies and complex time-series characteristics between locations in trajectories. The main principle is to train machine learning models with trajectory data, continuously adjusting the model parameters until the training is complete. The trained model can then generate a large number of synthetic trajectories given specific inputs[2]. Using deep generative models to generate synthetic trajectories has become a new trend.

2. Relevant Concepts

2.1 Differential Privacy

Differential Privacy is a technology designed to protect individual privacy by adding noise to make individual data in statistical analysis results difficult to identify. The core idea is to introduce random noise into the query results of a data-set to ensure that the overall statistical characteristics of the data-set remain unchanged while protecting the privacy of each individual.

Given two adjacent datasets D and D', if D and D' differ by only one element, differential privacy ensures that the probability distribution of the output of an algorithm A on D and D' is nearly identical. Specifically, an algorithm A satisfies ò - differential privacy if for any two adjacent datasets D and D', and any possible output $S \subseteq Range(A)$, the following holds:

$$P[A(D) \in S] \leq e^{ò} \cdot P[A(D') \in S] \quad (1)$$

where ò is a non-negative parameter called the privacy budget. A smaller ò value provides stronger privacy protection.

2.2 Laplace Mechanism

The Laplace mechanism is a method of implementing differential privacy by adding Laplace noise to query results to protect privacy. The distribution of Laplace noise is as follows:

$$Lap(b) \quad (2)$$

Among them, is the scale parameter of the

Laplace distribution.

For a given query function $f : D \to R$, the response generated by the Laplace mechanism is:

$$A(D) = f(D) + Lap(\frac{\Delta f}{\grave{O}}) \quad (3)$$

Among them, Δf is the sensitivity of function f , defined as:

$$\Delta f = \max_{D,D'} f(D) - f\left(D'\right)_1 \quad (4)$$

Sensitivity Δf represents the maximum change in query results between adjacent datasets. By adding Laplace noise with a scale parameter of $\frac{\Delta f}{\grave{O}}$, the Laplace mechanism can ensure differential privacy. The probability density function of the Laplace distribution:

$$Lap(x|\ b) = \frac{1}{2b} exp(-\frac{|x|}{b}) \quad (5)$$

Among them, $b = \frac{\Delta f}{\grave{O}}$ represents the scale parameter of Laplacian noise.

2.3 Recurrent Neural Networks

RNN (Recurrent Neural Network, RNN) is a neural network model designed to process sequential data. Unlike traditional feed forward neural networks, RNN adds temporal relationships between previous and subsequent time steps on top of fully connected neural networks, enhancing the internal memory of the network and thus better handling time-series related problems. Therefore, RNN can be seen as a process of jointly modeling the input information at each time step and the information from the previous time step. During the computation process, RNN uses the hidden state of the previous time step to encode historical information, and combines it with the input at the current time step to calculate a new hidden state and output vector. This recursive structure enables RNN to capture dependencies in sequential data well, making it more efficient and accurate in processing data.

The internal structure of RNN includes input layer, hidden layer and output layer, and what makes RNN can analyze the sequence data is that the hidden layer contains a cyclic connection, which can receive not only the information input at the current moment, but also the hidden state at the previous moment. In RNN, each layer shares the weight parameters, including the weight matrix U between the input layer and the hidden layer, the weight matrix v between the hidden layer and the output layer, and the weight matrix w from the state of the hidden layer in the previous moment to the input in the current moment, which reduces the number of parameters that need to be learned in the network and improves the learning efficiency of the network.

2.4 BILSTM Neural Network

LSTM, although its memory capacity has become stronger and alleviated the problem of vanishing gradients, still has the flaw that it can only pass messages from front to back and cannot utilize subsequent information, which is a limitation in many tasks. Bidirectional recurrent neural network adds the reverse operation on the basis of RNN, which is able to utilize both front-to-back and back-to-back information to predict the output at the current moment. Based on this idea, a reverse LSTM is added to the original LSTM to constitute the BILSTM.Specifically, the forward LSTM learns from the past information, while the backward LSTM learns from the future information, so that it can better capture the complete information about the past and the future of each moment in the input sequence, and make its output results more accurate.

2.5 Track Privacy Protection Related Technologies

At present, generalization [2,3], mix-zones [4], inhibition, and disturbance are the four most commonly used methods for protecting the privacy of location trajectories. Both generalization and mix-zones can be used to protect the privacy of users' trajectories, but their effectiveness is not ideal. Although inhibition and disturbance can effectively protect the privacy of users' trajectory data, there are certain security risks. Generalization is to generalize the location at each moment in the trajectory into a region; mix-zones are more suitable for the Internet of Vehicles, using intersection conversion pseudonyms to protect vehicle information; inhibition is to suppress the distribution of locations based on the size of sensitive areas where the user's actual location is located; disturbance is to add noise to each moment's location to generate interference for location distribution. However, if the location to

be protected is within a sensitive area, after using the above methods for trajectory protection, adversaries can still obtain personal privacy information through various types of privacy attacks such as combinatorial attacks and background knowledge attacks [5]. Differential privacy [6] can compensate for the defects of the above protection techniques due to its strict mathematical definition and quantitative standards. At the same time, differential privacy technology does not rely on the background knowledge that attackers have, and adversaries cannot distinguish whether individual records are included in the database. Therefore, using it as the main method for protecting the privacy of user trajectories has become a key research direction in the current privacy protection field [7,8]

With the development of machine learning-related technologies, more and more scholars are now trying to combine deep learning technology with differential privacy. Yan Yan et al. [9] proposed a deep learning-based location big data partitioning structure prediction method and differential privacy release method. This method constructs a deep learning prediction model based on spatiotemporal sequences, and extracts the time and spatial correlation characteristics of historical location big data statistics partition structure matrix to achieve effective prediction of the partition structure matrix, thereby solving the problems of unreasonable partitioning and release structure and inefficient release methods in traditional location big data statistics. Chen et al. [10] proposed a new trajectory release algorithm RNN-DP, which combines recurrent neural networks with differential privacy technology for trajectory release, improving data availability. In view of the inability of recurrent neural networks to process long-distance data and the inability of most privacy protection schemes to resist background knowledge attacks, this paper proposes a bidirectional long short-term memory-differential privacy (BILSTM-DP) protection scheme for trajectory data release using differential privacy and recurrent neural network technology, in order to achieve the goal of improving data availability while resisting background knowledge attacks. This scheme can better process long-distance trajectory data, reducing the impact of gradient explosion and gradient disappearance, and better adapt to different types of trajectory data.

3. BILSTM-Laplace Scheme

3.1 System Architecture

The system architecture of the BILSTM-DP program consists of 2 parts: trajectory prediction and trajectory noise addition. The scheme realizes privacy preservation for trajectory data release in the following steps.

Step 1 The trajectory prediction module applies BILSTM to predict the trajectory. Unlike general neural networks, BILSTM is not only suitable for processing time-series data, but also achieves higher prediction accuracy and better captures long-term dependencies than unidirectional recurrent networks.

Step 2 Assign a budget for trajectory privacy and add Laplace noise to the trajectory data based on the privacy budget value. When adding Laplace noise, it is necessary to balance the privacy needs and data accuracy requirements of the application scenarios to ensure that both user privacy can be effectively protected and the usability of the trajectory data in subsequent processing and analysis can be ensured.

3.2 Trajectory Prediction

Long Short Term Memory Network is a temporal recurrent neural network, which is specially designed to solve the long term dependency problem that exists in general RNN (Recurrent Neural Network). BILSTM is used to perform trajectory prediction and generate new trajectories, thus achieving the effect of hiding the original trajectories. In the process of model training, BILSTM refers to the data on both sides of the prediction point at the same time, therefore, more accurate trajectory data can be obtained.

3.3 Laplace Trajectory Noise Addition

The main role of Laplacian Noise (LN) in differential privacy is to protect data privacy by adding noise. Specifically, it protects individual privacy by introducing noise into the data, making it difficult to detect the impact of individual data points. This method is particularly suitable for scenarios where statistical data or analytical results need to be published, as it blurs information about the details of the data but still preserves the overall trend. In addition, Laplacian noise achieves the goal of differential privacy by fulfilling the requirement of differential privacy by making

the distribution of query results from neighboring databases (i.e, two databases differing in only one data point) similar. This noise follows the Laplace distribution and is added to sensitive query results, making the data both useful and effective in preventing re-identification attacks and preserving privacy.

4. Conclusion

Current trajectory privacy protection techniques can be summarized as fake data, anonymization, generalization and suppression methods. Fake data techniques obfuscate real trajectories by adding fake trajectories, but are less efficient in scenarios with large-scale data and high real-time requirements. Anonymization techniques achieve privacy protection by blurring the trajectory points, but reduce the usability of trajectory data. The generalization technique protects privacy by blurring the precision of trajectory points and is applicable to a variety of scenarios. Suppression techniques protect privacy by selectively publishing trajectory data, but suffer from the problem of incomplete trajectories.

In this paper, we propose a BILSTM-based trajectory data protection scheme that utilizes the bi-directionality of BILSTM to predict and process trajectory data, thus increasing the availability of trajectory data.BILSTM is able to estimate the current value by utilizing both the past and the future data simultaneously, which is well suited for processing static trajectory data. For dynamic trajectory data, the accuracy decreases as only past data can be utilized.

To address these issues, future research efforts will focus on how to improve data processing efficiency while ensuring data availability, and more in-depth study of the combination of ground machine learning and trajectory privacy protection techniques, especially the differential privacy trajectory protection method combining neural networks and clustering, with a view to continuously improving the privacy protection effect. Optimize the differential privacy mechanism, introduce more prediction mechanisms and dynamically adjust the noise addition strategy to improve the strength of privacy protection and data availability as well as the real-time and accuracy of trajectory data processing. Explore the comprehensive application of multiple privacy protection techniques in order to realize the optimal privacy protection effect in different application scenarios.

Acknowledgement

This present research work was supported by Harbin Normal University Higher Education Teaching Reform Research Project(No. XJGZ202409).

References

[1] Cao Xiaoqian. Research on trajectory synthesis method based on GAN [D]. Zhejiang Normal University, 2023.

[2] Gramaglia M, Fiore M, Furno A, et al. GLOVE: towards privacy-preserving publishing of record-level-truthful mobile phone trajectories[J]. ACM/IMS Transactions on Data Science (TDS), 2021, 2(3): 1-36.

[3] Mahdavifar S, Deldar F, Mahdikhani H. Personalized privacy-preserving publication of trajectory data by general-ization and distortion of moving points[J]. Journal of Network and Systems Management, 2022, 30(1): 10.

[4] Hou L, Yao N, Lu Z, et al. Tracking based mix-zone location privacy evaluation in VANET[J]. IEEE Transactions on Vehicular Technology, 2021, 70(10): 10957-10969.

[5] Hua J, Gao Y, Zhong S. Differentially private publication of general time-serial trajectory data[C]. 2015 IEEE Conference on Computer Communications (INFOCOM), 2015, 549-557.

[6] Dwork C, McSherry F, Nissim K, et al. Calibrating noise to sensitivity in private data analysis[C]. Theory of Cryp-tography: Third Theory of Cryptography Conference, 2006, 265-284.

[7] Cai S, Lyu X, Li X, et al. A trajectory released scheme for the internet of vehicles based on differential privacy[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(9): 16534-16547.

[8] Zhao X, Pi D, Chen J. Novel trajectory privacy-preserving method based on prefix tree using differential privacy[J]. Knowledge-Based Systems, 2020, 198: 105940.

[9] Yan Yan, Cong Yiming, Adnan Mahmood, et al. A deep learning based approach for statistical release and privacy protection of location big data[J]. Journal of Communication,2022,43(01):203-216.

[10] Chen S, Fu A, Shen J, et al. RNN-DP: A

new differential privacy scheme base on Recurrent Neural Network for Dynamic trajectory privacy protection[J]. Journal of Network and Computer Applications, 2020, 168: 102736.

[11] Shen Z, Zhang Y, Wang H, et al. BiGRU-DP: Improved differential privacy protection method for trajectory data publishing[J]. Expert Systems with Applications, 2024, 252: 124264.

[12] QIN Chengyi, WU Lei, WEI Xiaochao, et al. Research and progress of location trajectory correlation differential privacy protection technology[J]. Journal of Cryptography, 2023, 10(06): 1118-1139.

