Research on the Application of Deep Learning-based Differential Privacy Protection Models in Financial Big Data

Yanjun Ma[1], Rui Lu[2], Yu Zhang[3,*]

EMAIL: mayanjun0010@yeah.net (Yanjun Ma); luruilly@sina.com (Rui Lu); zhangyu02xinben@163.com (Yu Zhang)

[1]Liaoning Police College, 260 Yingping Road, Ganjingzi District, Dalian, 116036, China

[2]Liaoning Police College, 260 Yingping Road, Ganjingzi District, Dalian, 116036, China

[3]Dalian Medical University, No. 9 West Section of Lushun South Road, Dalian, 116044, China

Abstract: With the digital transformation of the financial industry and the widespread application of big data, privacy protection has become a critical challenge. The differential privacy protection model based on deep learning preserves the privacy of financial big data while maintaining data availability and accuracy. This paper investigates and analyses the application of differential privacy protection models based on deep learning in the field of financial big data. Firstly, this paper introduces the basic concepts and principles of differential privacy, as well as the significance of deep learning in the financial domain. Next, the working principles and main methods of the differential privacy protection model based on deep learning are elaborated, including privacy budget allocation, noise injection, and model optimization techniques. Finally, this paper explores the application of the differential privacy protection model based on deep learning in financial big data analysis. It is noted that these models can achieve efficient and accurate data analysis and prediction while protecting user privacy. Specific applications include individual credit assessment, fraud detection, anti-fraud measures, among others. The paper also discusses the advantages and potential challenges of the models and proposes future research directions.

Keywords: Deep Learning, Differential Privacy, Privacy Protection, Financial Big Data

1. Introduction

With the continuous growth of data in the financial industry and the continuous development of Internet technology, financial big data has become an important component of current social informatization. Financial institutions and enterprises can make more accurate decisions, improve operational efficiency, reduce risks, and increase profitability by deeply mining and analyzing massive financial data. However, the collection, storage, and processing of big data involve a large amount of personal privacy information, posing a significant challenge of how to protect user privacy in big data applications [1].

Privacy protection technology is a key means to address privacy leakage issues, and differential privacy, as an effective technique for protecting personal privacy, has been widely applied. Its main idea is to perturb the original data, thus protecting user privacy while ensuring data

FFIT 2023, July 07-09, Chongqing, People's Republic of China
Copyright © 2023 EAI
DOI 10.4108/eai.7-7-2023.2338057

accuracy. However, differential privacy also faces some challenges in practical applications, such as the balance between privacy protection and data analysis and the impact of random noise [2].

Deep learning, as a machine learning technique that can automatically learn and extract features from data, has achieved significant success in areas such as image recognition, speech recognition, and natural language processing [3]. In terms of privacy protection, deep learning can perturb data to prevent the leakage of private information while preserving the utility of the perturbed data, thus providing a certain level of data privacy protection.

Therefore, combining deep learning with differential privacy can provide more effective technical means for privacy protection in financial big data. This paper aims to explore the application of deep learning-based differential privacy protection models in financial big data, with the goal of improving the reliability and effectiveness of privacy protection while ensuring accurate data analysis. Through this research, we can provide more comprehensive data privacy protection solutions for financial institutions and enterprises, safeguard user privacy, enhance data security and confidentiality, and further promote the digital transformation and intelligent development of the financial industry.

2. Deep Learning and Differential Privacy Protection

2.1. Fundamentals of deep learning

Deep learning is a machine learning approach based on neural networks, aiming to predict output results by learning the features of input data [4]. Deep learning achieves modelling of complex nonlinear relationships through multi-layer neural networks and possesses powerful data analysis capabilities. The core of deep learning models is the neural network, which consists of three main parts: the input layer, hidden layers, and output layer. The hidden layers can include multiple layers.

2.2. Fundamentals of differential privacy protection

Differential privacy is a technique that protects data privacy by randomizing individual data through perturbation [5]. Its core idea is to add noise to conceal the real data, thus preventing the disclosure of sensitive information while ensuring data availability. Specifically, differential privacy techniques achieve randomization of individual data through methods such as adding noise, dimensionality reduction, filtering, and sampling while maintaining data utility.

Differential Privacy (DP) is a privacy protection technique proposed by Dwork in 2006 [6], which achieves data privacy by randomizing individual data through perturbation. The core idea of DP is to add noise to conceal the real data, thus preventing the disclosure of sensitive information while ensuring data availability. The amount of noise added by differential privacy is independent of the dataset size, requiring only a small amount of noise to achieve a high level of privacy protection in large-scale datasets. The definition of (ε, δ)-differential privacy is as follows [7]:

Given a random algorithm M, let P be the set of all possible outputs of algorithm M. For any pair of neighbouring datasets D and D', and for any subset S of P, if algorithm M satisfies Equation (1), then algorithm M is said to achieve (ε, δ)-differential privacy.

$$\Pr\left[M(D) \in S\right] \leq \Pr\left[M(D') \in S\right] \cdot e^{\varepsilon} + \delta \tag{1}$$

Among them, the parameter ε represents the privacy budget, which denotes the degree of privacy protection achieved by differential privacy techniques. A smaller value of ε indicates a higher level of privacy protection. δ represents the probability of violating strict differential privacy.

2.3. Integration of differential privacy and deep learning

Differential privacy techniques can be combined with deep learning models to achieve in-depth analysis and exploration of data while ensuring data privacy. There are two main types of differential privacy-based deep learning models: one is based on noise perturbation, which adds noise to the input data for privacy protection; the other is based on gradient descent, which adds noise during the gradient computation process for privacy protection.

Currently, differential privacy models based on deep learning have been widely applied in various fields, including social networks, healthcare, finance, etc., and have achieved good results in privacy protection and data analysis accuracy. In the financial domain, with the continuous accumulation and application of financial big data, protecting user privacy and sensitive information has become a crucial issue. Therefore, differential privacy models based on deep learning have significant significance in financial data analysis and applications.

3. Design of Deep Learning-based Differential Privacy Protection Models

3.1. Analysis of differential privacy protection requirements for financial big data

Financial data is experiencing explosive growth in terms of quantity and variety, including users' personal sensitive information such as transaction details, credit scores, ID numbers, phone numbers, etc. These data play a crucial role in decision-making, risk control, customer service, and other aspects of the financial industry [8]. However, with the frequent occurrence of security incidents such as network attacks and data breaches, financial data faces serious security threats. Therefore, privacy protection for financial data is particularly important.

Financial data exhibits the following characteristics.

(1) Large data volume: Financial data is vast, encompassing securities trading data, banking transaction data, insurance data, etc. These data contain a wealth of user information and transaction records.

(2) Diverse data types: Financial data includes structured and unstructured data, such as tables, text, images, etc.

(3) High data value: Financial data contains various customer-related information, such as personal details and transaction records, making the protection of customer privacy crucial.

(4) High data security requirements: Leakage of financial data can lead to severe consequences. Therefore, the security of financial data is of utmost importance.

Considering the characteristics and privacy protection requirements of financial big data, it is necessary to apply differential privacy protection to safeguard the data. The requirements for differential privacy protection of financial data mainly include the following aspects.

(1) Data privacy protection: Sensitive information pertaining to users, such as transaction records, credit ratings, ID numbers, and phone numbers, needs to be protected to prevent leakage.

(2) Data availability protection: To ensure the availability of financial data, it is necessary to apply noise perturbation techniques to the data to prevent data destruction.

(3) Data integrity protection: Ensuring the integrity of financial data requires the application of differential privacy protection to prevent data tampering.

(4) Data application protection: Protecting the application of financial data necessitates applying differential privacy protection to ensure the security of data usage.

3.2. Basic approach of deep learning-based differential privacy protection model

Deep learning is one of the most widely used machine learning techniques, capable of analysing and processing large volumes of data. However, when dealing with financial big data, it is necessary to introduce differential privacy protection mechanisms into deep learning models to safeguard data privacy. The basic approach of a deep learning-based differential privacy protection model involves adding random noise during the training process of the deep learning model to prevent the leakage of training data. Specifically, by incorporating differential privacy protection mechanisms into the training process of the deep learning model, it is possible to effectively protect the data without compromising the accuracy of the model.

The specific process of a deep learning-based differential privacy protection model generally includes the following steps.

(1) Data pre-processing: The original data is processed, including data cleaning, feature selection, and data normalization.

(2) Model selection and design: Choose a suitable deep learning model based on the specific situation, such as convolutional neural networks, recurrent neural networks, etc., and design the model structure and parameters.

(3) Introduction of differential privacy protection mechanism: Introduce the differential privacy protection mechanism during the model training process, typically achieved by adding noise.

(4) Model training and testing: Train and test the deep learning model with the incorporated differential privacy protection mechanism to obtain the final model performance.

3.3. Parameter tuning for differential privacy protection models

In the design of differentially private protection models based on deep learning, it is necessary to fine-tune the parameters of the model in order to improve its performance and effectiveness.

The noise mechanism is a key technique for achieving differential privacy protection, and the commonly used noise addition mechanisms are the Laplace mechanism and the exponential mechanism [9]. The Laplace mechanism achieves differential privacy protection by perturbing the true output values with noise generated from the Laplace distribution. The Laplace mechanism is defined as follows [10]:

Differential privacy protection of released data is achieved by adding Laplace noise. Let's assume there is a data table T, a function $f:T\to R^d$, and the sensitivity of function f is denoted as Δf. Then, $f(T)+Lap\left(\frac{\Delta f}{\varepsilon}\right)$ represents the added Laplace noise, satisfying ε-differential privacy. The probability density function of the Laplace distribution is denoted as $P(x/b)=\frac{1}{2b}\exp\left(-\frac{|x|}{b}\right)$.

Among them, $b=\Delta f/\varepsilon$。The sensitivity Δf is defined as follows:

Let $f:T\to R^d$ be a query function that takes as input any pair of neighboring data tables T_1 and T_2. The sensitivity of function f is defined as:

$$\Delta f=\max_{T_1,T_2}\left\|f(T_1)-f(T_2)\right\|_1$$

Where $\left\|f(T_1)-f(T_2)\right\|_1$ denotes a distance metric measuring the difference between the outputs of function f on neighboring data tables T_1 and T_2.

Figure 1 illustrates the optimization process of the Differential Privacy Model. In this figure, A represents any user in the dataset. Firstly, the datasets containing A and excluding A are analyzed separately. Subsequently, the processed results are fine-tuned by adding Laplace noise $f(T)+Lap\left(\frac{\Delta f}{\varepsilon}\right)$ to ensure that the differences between the output results do not exceed ε. In this model, the objective is to achieve differential privacy by adding Laplace noise that conforms to the output results. This approach guarantees data privacy while maintaining a high level of usability.

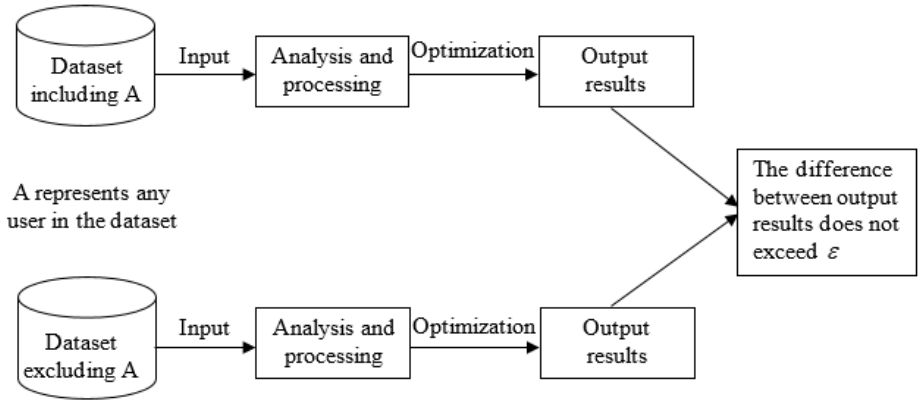

Figure 1: Optimization Illustration of Differential Privacy Model

4. Application of Differential Privacy Protection Models in Financial Big Data

After designing and implementing the differential privacy protection model based on deep learning, this paper proceeds to discuss its applications in the field of financial big data. Specifically, we explore two application scenarios in the areas of financial risk management and financial credit, and describe the specific methods and effects of these applications.

4.1. Application of differential privacy protection in financial risk management

Financial risk management is a crucial business domain that relies on extensive financial data to predict future risks and formulate appropriate decision strategies [11]. However, due to the involvement of personal privacy information, the privacy protection of financial data becomes particularly important. Traditional protection methods often employ techniques such as data encryption, but they suffer from issues like low information utilization. Therefore, the use of differential privacy protection methods can better safeguard data privacy while maintaining a high level of information utilization.

In the field of financial risk management, deep learning techniques have already been widely applied. The design of differential privacy protection models based on deep learning can enhance the privacy protection capabilities of these applications. Taking individual credit scoring as an example, applying differential privacy protection models can safeguard the privacy of individual credit scores while ensuring the model exhibits good predictive performance. Compared to traditional methods, this approach provides better protection of personal privacy while delivering more accurate prediction results, thereby offering more reliable foundations for financial risk management decisions.

4.2. Application of differential privacy protection in the financial credit domain

The financial credit domain is also a significant business area that relies on extensive personal financial and credit information for credit assessment and making informed loan decisions. However, personal privacy data carries high sensitivity in this domain, making the need for

differential privacy protection evident. In the field of financial credit, applying differential privacy protection models based on deep learning can provide dual guarantees of data privacy and data utility while addressing issues such as centralized data storage and privacy leakage.

In the financial credit domain, differential privacy protection models can be applied to safeguard users' personal information. For instance, in loan assessment, banks need to use users' personal information to evaluate their eligibility for loans. This personal information includes age, gender, income, credit history, among others. If this personal information is leaked, it can cause severe harm to user privacy.

By utilizing differential privacy protection models based on deep learning, accurate assessment and analysis of users' personal information can be conducted while preserving user privacy. For example, neural network models can be used to model users' personal information, and differential privacy techniques can be employed to protect user privacy. Throughout this process, all computations are conducted with the assurance of preserving user privacy, ensuring that users' personal information remains confidential.

In the financial credit domain, differential privacy protection models based on deep learning can be applied in the following areas.

(1) Loan assessment: Utilizing differential privacy protection models to model and evaluate users' personal information, determining their eligibility for loans. This aids banks in accurately assessing users' credit risks while safeguarding user privacy.

(2) Credit evaluation: Employing differential privacy protection models to model and evaluate users' credit history, determining their credit ratings. This helps banks accurately assess users' credit ratings while preserving user privacy.

(3) Risk assessment: Using differential privacy protection models to model and evaluate users' personal information, identifying potential risks. This assists banks in accurately assessing users' risk levels while protecting user privacy, enabling them to implement appropriate risk control measures.

(4) Anti-fraud measures: Utilizing differential privacy protection models to model and evaluate users' personal information, identifying potential fraudulent activities. This helps banks accurately detect fraudulent behaviour while preserving user privacy, allowing them to take necessary measures to prevent fraud.

The application of differential privacy protection models based on deep learning in the financial credit domain can enhance risk management effectiveness, improve compliance, and provide accurate credit risk assessments while safeguarding user privacy. This benefits both users and financial institutions.

5. Conclusion

This paper primarily investigates the application of deep learning-based differential privacy protection models in the field of financial big data. Firstly, this paper analyses the privacy protection issues and challenges faced by current financial big data applications, thoroughly discusses the current state of differential privacy techniques in the financial domain. Secondly,

this paper conducts research on deep learning-based differential privacy protection models, meticulously analyses the demands of financial big data for differential privacy protection, and designs deep learning-based differential privacy protection models suitable for the financial domain. Furthermore, this paper explores the specific methods and effects of deep learning-based differential privacy protection models in two application scenarios: financial risk control and financial credit, providing reliable and effective solutions for privacy protection in financial big data. Through this research, more comprehensive data privacy protection solutions are provided for financial institutions and enterprises, aiming to ensure user privacy, enhance data security and confidentiality, and further promote the digital transformation and intelligent development of the financial industry.

The proposed deep learning-based differential privacy protection models demonstrate good performance in financial big data, but they still have limitations. Firstly, the models still have privacy leakage risks, such as inference attacks on training data and model parameter reverse engineering attacks. Secondly, the models do not cover all application scenarios in the financial domain and need to be validated in more complex data environments. Future research can be conducted in two aspects: (1) Enhancing the privacy protection strength of the models by exploring more secure and effective privacy protection methods. (2) Expanding and validating application scenarios to explore the effectiveness in areas such as financial investment and financial derivatives.

Acknowledgements

The author acknowledges the Innovation Team Support Program of Liaoning Police College (LNPC2021KYTD001), Research on Innovative Practical Teaching Models of Artificial Intelligence Courses in the Context of Public Security Big Data (LNBKJG11432202101) and Key Laboratory of Intelligent Applications in Public Security Big Data of Liaoning Province.

References

[1] Abid Mehmood, ynkaran Natgunanathan, Yong Xiang, Guang Hua, Song Guo. (2016) Protection of Big Data Privacy. IEEE Access, 4:1821-1834. doi: 10.1109/ACCESS.2016.2558446.

[2] Acs G,Melis L, Castelluccia C, et al. (2018) Differentially private mixture of generative neural networks. IEEE Transactions on Knowledge and Data Engineering, 31(6):1109-1121. doi: 10.1109/TKDE.2018.2855136.

[3] Deng Li, Yu Dong. (2014) Deep learning: methods and applications. Foundations and Trends in Signal Processing, 7(3 4):197-387. http://dx.doi.org/10.1561/2000000039.

[4] Sohangir, S., Wang, D., Pomeranets, A. et al. Big Data: Deep Learning for financial sentiment analysis. J Big Data 5, 3 (2018). https://doi.org/10.1186/s40537-017-0111-6.

[5] Phan, N., Wu, X. & Dou, D. Preserving differential privacy in convolutional deep belief networks. Mach Learn 106, 1681–1704 (2017). https://doi.org/10.1007/s10994-017-5656-2.

[6] DWORK C, MCSHERRY F, NISSIM K, et al. Calibrating noise to sensitivity in private data analysis[C] // Theory of Cryptography, Third Theory of Cryptogra- phy Conference, TCC 2006, New York, NY, USA, March 4-7, 2006, Proceedings. Berlin, German : Springer, 2006 : 265-284.

[7] Dwork C, Roth A. The algorithmic foundations of differential privacy[J]. Found. Trends Theor. Comput. Sci., 2014, 9(3-4): 211-407.
[8] Regin R, Rajest S S, Shynu T. (2023) A Review of Secure Neural Networks and Big Data Mining Applications in Financial Risk Assessment. Central Asian Journal of Innovations on Tourism Management and Finance, 4(2): 73-90. https://doi.org/10.17605/OSF.IO/N8MH5.
[9] Cormode G, Procopiuc C M, Srivastava D,et al. Differentially private summaries for sparse data[C].Proceedings of 15th International Conference on Database Theory, Berlin, Germany, 2012: 299-311. https://doi.org/10.1145/2274576.2274608.
[10] Dwork C. Differential privacy. In: proceedings of the 33rd International Colloquium on Automata, languages and Programming. Springer, Berlin, 2006: 1–12. https://doi.org/10.1007/11787006_1.
[11] Hajiheydari N, Delgosha M S, Wang Y, et al. (2021) Exploring the paths to big data analytics implementation success in banking and financial service: an integrated approach. Industrial Management & Data Systems, 121(12): 2498-2529.doi: 10.1108/IMDS-04-2021-0209.