Research and Application Path Analysis of Deep Learning Differential Privacy Protection Method Based on Multiple Data Sources

Junhua Chen[1](✉) and Yiming Liu[2]

[1] Institute of Standardization Theory and Strategy, China National Institute of Standardization (CNIS), Beijing 100088, China
chenjunh@cnis.ac.cn

[2] Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing Technology and Business University, Beijing 100048, China

Abstract. The deep learning model will contain user-sensitive information during training. When the model is applied, the attacker can recover the sensitive information in the training data set through model inversion attacks, and directly or indirectly disclose the user-sensitive information. The existing methods can not solve the problem that the privacy budget accumulates with the increase of training times. This paper proposes a method of research on deep learning differential privacy protection method based on multiple data sources, aim at making privacy consumption independent of the number of training epochs to guarantee the potential to work with large datasets. First, we calculate the privacy budget upper bound to optimal experiment selection for parameter estimation. Second, we use the upper bound to determine the number of group, also to balance the number of group and the data size of the subdataset, avoiding data relying on a single model causes leakage of user sensitive information. Finally, we ensemble several models with majority voting, and perturb single model the traditional convolutional deep belief network (CDBN) objective functions, to descend the dependence of privacy budgets on the training deep learning model and improve machine learning results. We applied our model to a health social network dataset and MNIST dataset, and the results show that our method has high privacy protection ability than the existing method for sensitive information on the training dataset. Moreover, standardization can be a feasible path for the generalized application of the technique, which is beneficial for the stability of the application of differential privacy protection techniques and the subsequent feedback updates.

Keywords: Multiple data sources · Differential privacy · Deep learning · Aggregation · Deep Confidence Network

572022Y-9399 Research on the mechanism of national standards acceptance social organization standards and application of typical cases.

G. Guan et al. (Eds.): ICBDSS 2022, AHCS 8, pp. 299–310, 2023.
https://doi.org/10.2991/978-94-6463-064-0_34

1 Introduction

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. Some deep learning applications involve training data that is sensitive, such as the medical histories of patients in a clinical trial. A model may inadvertently and implicitly store some of its training data; careful analysis of the model may therefore reveal sensitive information. For example, Fredrikson et al. demonstrated a model-inversion attack that recovers images from a facial recognition system. With the recent upswing of privacy protection in deep learning, a new field of research, known as federated learning, has sparked global interest.

Differential privacy is an evaluation framework invented by Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith to evaluate the privacy guarantees provided by privacy-preserving mechanisms that address the limitations of previous approaches such as “k-anonymity”.Since Dwork et al. proposed differential privacy theory in ICALP (International Colloquium Automata, Languages and Programming) in 2006 (Dwork 2006), the theory 2006–2015 academics combined differential privacy with traditional machine learning algorithms, and in 2016 Google Research Institute Martín Abadi et al. combined differential privacy with deep learning for the first time, and in the same year combined semi-supervised learning algorithms with differential privacy, we can see that combining deep learning with differential privacy techniques is a hot spot for research and needs to go deeper, and the current state of research on combining differential privacy with deep models will be summarized in detail in the next section.

2 Our Approach

In this section, we introduce the specifics of the multi-group differential privacy protection method with upper bounds approach, which is illustrated in Fig. 1. The framework is divided into tow parts. The first one describes how the data is partitioned to train the deep learning model, and last one is describes how prediction made by this ensemble are noisily aggregated. The detail as follows: First, we calculate the upper bound of the corresponding privacy budget from the attacker’s perspective, with α and β are the adversary’s prior belief and posterior belief on $X = \omega$ given a query response. to reduce the scope of the experiment. Second, the single training model makes the data features dependent on only one model, which easily leads to leakage of user-sensitive information. Therefore, we are grouped the original dataset, and training models on each sub-data set can improve the accuracy of the results. we use the privacy budget upper bound to determine the number of packets, to ensures the balance in number of packets with the data size of the sub-dataset. Finally, we ensemble several models with majority voting, and perturb single model the traditional with Chebyshev polynomial convolutional deep belief network (CDBN) objective functions. The most important reason behind the usage of Chebyshev polynomial is that the upper and lower bounds of the error incurred by approximating activation functions and energy functions can be estimated and proved,

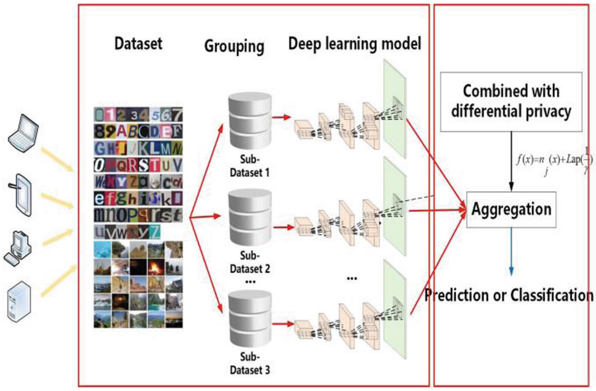

Dataset
Grouping
Deep learning model
Sub-Dataset 1
Sub-Dataset 2
Sub-Dataset 3
Combined with differential privacy
$f(x)=n_j(x)+Lap(\frac{1}{\gamma})$
Aggregation
Prediction or Classification

Fig. 1. Multi-group differential privacy protection method with upper bounds

we propose to preserve differential privacy in the model before applying Gibbs sampling. to descend the dependence of privacy budgets on the training deep learning model and improve machine learning results. Figure 1 shows the framework of the method.

3 Upper Bound of the Privacy Budget ε and Grouping

Differential privacy technology can provide a strong degree of privacy protection, the parameter ε measures the ability of the random algorithm A to resist attacks, and the smaller the parameter ε, the greater the privacy protection provided by it.

Definition 1 (ε-differentially private mechanism). A randomized mechanism A is ε-differentially private if for all data sets D and D^0 differing on at most one element, and all O $\subseteq$ Range(A) (Dwork 2014):

$$\Pr[A(D) \in O] \le \exp(\varepsilon) \times \Pr[A(D0) \in O] \tag{1}$$

The parameter ε is called the privacy budget.

Definition 2 (privacy budget). Privacy budget ε is a probability ratio of algorithm A to obtain the same output on two neighboring data sets, which in fact reflects the level of privacy protection A can provide.

In practical applications, ε usually takes a small value, the smaller, the higher the level of privacy protection. When ε is equal to 0, the protection level is maximized. For any adjacent data set, the algorithm outputs two identical results with probability distributions, and the results do not reflect any useful information about the data set. Therefore, the value of ε should be combined with the specific needs to achieve the output of the security and availability of the balance. As shown in Fig. 2.

3.1 Adversary Model

We assume a very strong adversary who has complete knowledge of the universe, i.e., full access to all records in the universe; thus each attribute value of all records in D

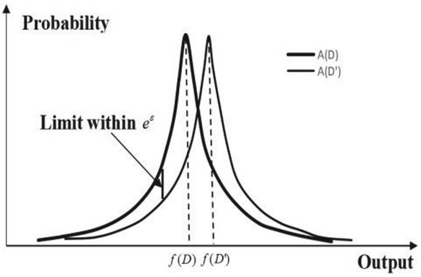


Fig. 2. Output Probability of Random Algorithm on Neighboring Datasets

is known to the adversary. The adversary knows everything about the universe except that which individual is missing in the database D^0(i.e., who is on academic probation). Assuming an adversary having complete knowledge about each individual since the database is not unrealistic because differential privacy is supposed to provide privacy given adversaries with arbitrary background knowledge.

In our model, the adversary has a database D consisting of N records, i.e., knowledge of the exact attribute values of each individual in D, and has an infinite computational power. Given a database D^0 with $N - 1$ records sampled from D(i.e., $D^0 \in D$ and $|D^0| = |D| - 1$), the adversary? goal is to figure out absence of a victim individual in D^0 by using knowledge of D. This is identical to find out other individuals? presences in D. With respect to our example, a privacy breach is to allow the adversary to guess absence/presence of an individual in D^0 correctly with high probability.

3.2 Attack Model

To determine membership in D^0, the adversary maintains a set of tuples hω, α, βi for each possible combination w of D^0, where α and β are the adversary's prior belief and posterior belief on X = ω given a query response. Let ψ denote the set of all possible combinations of D^0. For simplicity, we assume α is a uniform prior, i.e.

We refer to each possible combination w in ψ as a possible world. The posterior belief β is defined in Definition 3. Definition 3 (Posterior belief on $D^0 = \omega$). Given the query function f and the query response $\gamma = k_f(X^0)$, for each possible world w, the adversary's posterior belief on w is defined as:

The posterior belief $\beta(\omega)$ represents the adversary's changed belief on each possible world that the underlying database being queried against is w. To figure out which individuals are in the database, the adversary issues a query against D^0 and gets a noisy answer. After seeing the query response, the adversary computes the posterior belief for each possible world. Finally, the adversary selects one with the highest posterior belief as a "best guess". The confidence of the adversary's guess is calculated using Definition 4.

Definition 4 (Confidence level). Given the best guess, the adversary's confidence in guessing the missing element is defined as:

$$\mathrm{conf}(\omega 0) = \beta(\omega 0) - \alpha(\omega 0) \tag{2}$$

As the adversary's posterior belief on each possible world becomes large, the chances of disclosing any individual's presence in the database also become high, which makes disclosure of the statistics. This has an implication that the adversarys posterior belief on each possible world can be thought of as the risk of disclosure.

4 Model Training on Sub-datasets

In this section, we traning the model on the sub-datasets then calculate privacy budget. We base on the characteristic of differential privacy.

Theorem 1 Let $A_i$be an ε_i-differential private algorithm for $i \in [n]$ on dataset D. Then if $A_{[n]}$ is defined to be $A_{[n]} = (A_1\ (D), A(D)_2, \cdots, A_n(D))$, then $A_{[n]}$ is $(\sum^{\mathrm{n}}{}_{\mathrm{i=1}}\varepsilon_{\mathrm{i}})$-differential private.

Theorem 2 Let $A_i$be an ε_i-differential private algorithm for $i \in [n]$ on disjoint dataset $D_1, D_2, \cdots, D_n$. Then if $A_{[n]}$ is defined to be $A_{[n]} = (A_1\ (D), A(D)_2, \cdots, A_n(D))$, then A[n] is $(\max\varepsilon_i))$-differential private.

Existing methods cannot solve the problem that privacy budgets accumulate as the number of training steps. To solve this, we find a approach of using the Chebyshev Expansion to derive polynomial approximations of nonlinear energybased objective functions, such that differential privacy can be preserved by leveraging the functional mechanism. In principle, many polynomial approximation techniques, e.g., Taylor Expansion, Euler polynomial, Discrete Fourier transform, Hermite polynomial, Laguerre polynomial, and even the stateof-the-art techniques including spectral methods and Finite Element methods, can be applied to approximate non-linear energy functions used in CDBNs. There are two challenges in the traditional energy function $E\ (D,\ W)$ that prevent us from applying it for private data reconstruction analysis: (1) Gibbs sampling is used to estimate the value of every $h^k{}_{ij}$; and (2) The probability of every $h^k{}_{ij}$ equal to 1 is a sigmoid function which is not a polynomial function with parameters W^k. Therefore, it is difficult to derive the sensitivity and error bounds of the approximation polynomial representation of the energy function $E\ (D,\ W)$. With these challenging issues, Chebyshev polynomial really stands out.

In our method an input layer V and a hidden layer H. The layer of hidden units consists of K groups, each of which is an $N_H \times N_H$array of binary units. There are $N_H{}^2$ K hidden units in total. In addition, each group of hidden units has a bias b_k, and all visible units share a single bias c.

4.1 Construction of the Loss Function with Chebyshev Expansion

The most important reason behind the usage of Chebyshev polynomial is that the upper and lower bounds of the error incurred by approximating activation functions and energy functions can be estimated and proved, we propose to preserve differential privacy in the model before applying Gibbs sampling. The generality is still guaranteed since Gibbs sampling is applied for all hidden units. In addition, we need to derive an effective polynomial approximation of the energy function, so that differential privacy preserving is feasible. First, we propose to consider the probability $\mathrm{P}\left(h^k_{ij} = 1|\mathrm{v}\right) = \sigma(W^k * \mathrm{v}) + b_k$ instead of $h^k{}_{ij}$ in the energy function $E\ (D,\ W)$. The main goal of minimizing the energy function. Therefore, the generality of our proposed approach is still guaranteed. The energy function can be rewritten as follows:

$$E^1(D, W) = \sum\nolimits_{t \in D}\Big[-\sum\nolimits_{k=1}^{K=1}\sum\nolimits_{i,j=1}^{N_H}\sum\nolimits_{r,s=1}^{N_W}\sigma\Big(\Big(W^k * v\Big) + b_k\Big) \times W_{rs}^k v_{i+r\text{ - }1,\text{j + s - }1}^t - \\ \sum\nolimits_{i,j=1}^{N_V} b_k + \sum\nolimits_{r,s=1}^{N_W}\sigma(\Big(W^k * v^t)_{ij} + b_k\Big) - c\sum\nolimits_{i,j=1}^{N_V} v_{ij}^t\Big] \tag{3}$$

To solve the problem of sigmoid function, this paper constructs the sigmoid function using Chebyshev's inequality as follows:

$$\sigma\left(\frac{((W^k * v^t)_{ij} + b_k)}{(Z_ij^k)}\right) = \sum\nolimits_{l=0}^{\infty} A_l T_l\left(\frac{((W^k * v^t)_{ij} + b_k)}{(Z_ij^k)}\right) \tag{4}$$

Now, there is still a challenge that prevents us from applying the functional mechanism to preserve differential privacy in applying sigmoid function: The equation involves an infinite summation. To address this problem, we remove all orders greater than L. Based on the Chebyshev series, the polynomial approximation of the energy function E^1 $(\cdot)$ can be written as:

$$E^1(D, W) = \sum\nolimits_{(t \in D)}\left[-\sum\nolimits_{k=1}^{K=1}\sum\nolimits_{i,j=1}^{N_H}\sum\nolimits_{r,s=1}^{N_W}\left(\sum\nolimits_{l=0}^{L} A_l T_l\left(\frac{(W^k * v^t)_{ij} + b_k}{(Z_{ij}^k)}\right)\right) \times W_{rs}^k v_{i+r-1,\text{j + s}-1}^t \\ - \sum\nolimits_{i,j=1}^{N_V} b_k \sum\nolimits_{i,j=1}^{N_H}\sum\nolimits_{l=0}^{L} A_l T_l\left(\frac{(W^k * v^t)_{ij} + b_k}{(Z_{ij}^k)}\right) - c\sum\nolimits_{i,j=1}^{N_V} v_{ij}^t\right] \tag{5}$$

4.2 Construction of the Final Loss Function to Use Functional Mechanism

We employ the functional mechanism to perturb the objective function E^* $(\cdot)$ by injecting Laplace noise into its polynomial coefficients. The hidden layer contains K groups of hidden units. Each group is trained with a local region of input neurons, which will not be merged with each other in the learning process. Therefore, it is not necessary to aggregate sensitivities of the training algorithm in K groups to the sensitivity of the function E^* $(\cdot)$. Instead, the sensitivity of the function b E^* $(\cdot)$ can be considered the maximal sensitivity given any single group. As a result, the sensitivity of the function E^* $(\cdot)$ can be computed in the following lemma.

Theorem 3 (Phan, 2017) Let D and D^0 be any two neighboring datasets. Let E^* (D, W) and E^* (D^0, W) be the objective functions of regression analysis on D and D^0, respectively. α are Chebyshev polynomial coefficients. The following inequality holds:

$$\Delta \leq 2\max_{t,k}\sum\nolimits_{i,j=1}^{N_H}\sum\nolimits_{l=0}^{L}|\eta_l|\left[\left(\frac{\sum\nolimits_{r,s=1}^{N_W} v_{ij,rs}^{t,k} + 1}{z_{ij}^k}\right) + \sum\nolimits_{r,s=1}^{N_W}\left(\frac{\sum\nolimits_{r,s=1}^{N_W} v_{ij,r\prime s\prime}^{t,k} + 1}{z_{ij}^k}\right)^l \left|v_{ij,rs}^{t,k}\right|\right] + \sum\nolimits_{i,j=1}^{N_V} v_{ij,rs}^{t,k} \tag{6}$$

We use gradient descent to train the perturbed model E-$(\cdot)$. That results in private hidden layers. We stack multiple private hidden layers and max-pooling layers on top of each other. The pooling layers only play the roles of signal filters of the private hidden layers. Therefore, there is no need to enforce privacy in max-pooling layers.

5 Aggregation of the Deep Learning Model

In this section, we aggregation for multi-group training model for practical use. The privacy guarantees of this model ensemble stems from its aggregation. And the *m* be the number of classes in our task. The label count for a given class $j \in [m]$ and an input x^* is the number of model that assigned class j to input x^*: $n_j(x^*) = |\{i: i \in [N], f_i(x^*) = j\}|$. If we simply apply the largest count of ensemble's model decision may depend on a single model's vote. Indeed, when two labels have a vote count differing by at most one, there is a tie: the aggregated output changes if one model makes a different prediction. We add random noise to the vote counts nj to introduce ambiguity:

$$f(x) = \arg \underset{j}{max}\left\{n_j\left(x^*\right) + \text{Lap}\left(\frac{1}{\gamma}\right)\right\} \tag{7}$$

In this equation, γ is a privacy parameter and *Lap*(*b*) the Laplacian distribution with location 0 and scale *b*. The parameter γ influences the privacy guarantee we can prove. Intuitively, a large γ leads to a strong privacy guarantee, but can degrade the accuracy of the labels, as the noisy maximum f above can differ from the true plurality.

6 Experiments

6.1 Experimental Environment and Dataset

In this section, we will analyze, verify and explain the effect of multi-group differential privacy protection method with upper bounds algorithm through specific experiments. The experimental environment is Intel Xeon CPU E5–2603 v3@1.6 GHZ, 8 GB RAM, 2 TITAN X, Ubuntu 16.04 64bit os. The experiment uses TensorFlow1.0 framework, the algorithm is implemented by python.

Health Social Network Data were collected from Oct 2010 to Aug 2011 as a collaboration between Peace-Health Laboratories, SK Telecom Americas, and the University of Oregon to record daily physical activities, social activities (text messages, competitions, etc.), biomarkers, and biometric measures (cholesterol, BMI, etc.) for a group of 254 overweight and obese individuals. In total, we consider three groups of attributes:

Behaviors: #competitions joined, #exercising days, #goals set, #goals achieved, P(distances), avg(speeds);

#Inbox Messages: Encouragement, Fitness, Followup, Competition, Games, Personal, Study protocol, Progress report, Technique, Social network, Meetups, Goal, Wellness meter, Feedback, Heckling, Explanation, Invitation, Notice, Technical fitness, Physical;

Biomarkers and Biometric Measures: Wellness Score, BMI, BMI slope, Wellness Score slope.

The MNIST database of handwritten digits consists of 60,000 training examples, and a test set of 10,000 examples.

Each example is a 28 × 28 size gray-level image. The MNIST dataset is completely balanced, with 6,000 images for each category, with 10 categories in total.

6.2 The Effect of the Number of Groups on Accuracy on Two Database

As outlined in Sect. 3, this opinion is reflected by our data-dependent privacy analysis, which provides stricter privacy bounds when the model is too much. We calculate the number of votes for every possible result and measure the difference in votes between the first popular label and the second most popular label, such as, the gap. If the gap is small, introducing noise during aggregation might change the label assigned from the first to the second.

6.3 The Effect of the Dataset Size on Accuracy on Health Social Network Dataset

The every model includes two hidden layers. We trained 10 first layer bases, each variables v, and 10-s layer bases. The pooling ratio was 2 for both layers. In the paper, contrastive divergent algorithm (Hinton 2002) was used to optimize the energy function, and back-propagation was used to optimize the cross-entropy error function in the softmax layer. Verify the accuracy of algorithms for different data sizes, thus the method has been trained on daily and weekly datasets.

Competitive Models. We compare our method with two types of state-of-the-art models, as follows:

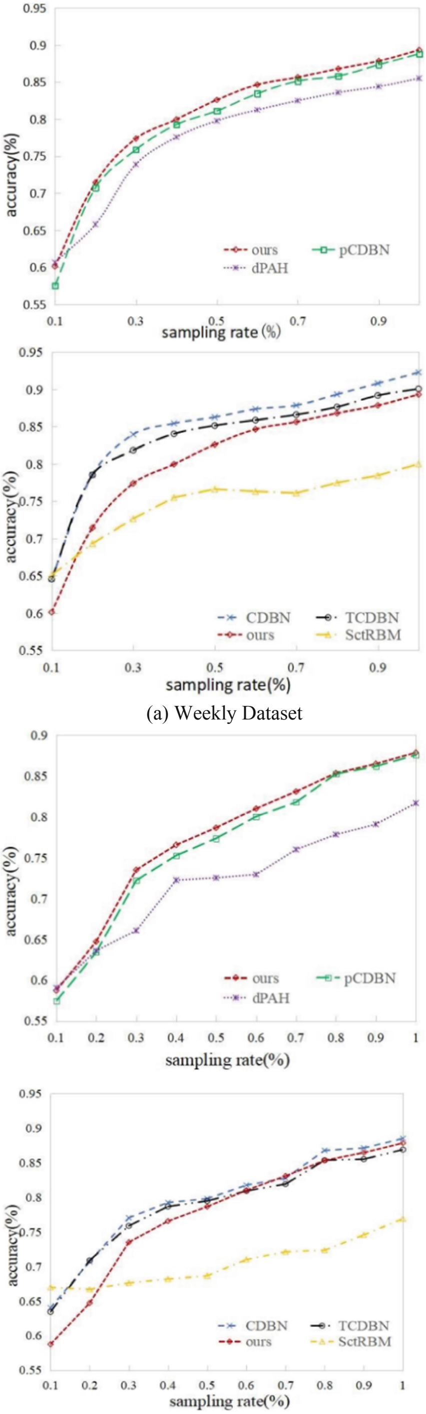
a) Deep learning models for human behavior prediction, such as CDBN, TCDBN, SctRBM.

b) Deep Private Auto-Encoder dPAH (Phan 2016).

c) Private convolutional deep belief network pCDBN.

Figure 3 shows the prediction accuracy of each algorithm as a function of the dataset cardinality. We vary the size of m, which also can be considered as the sampling rate of the dataset. In both datasets, there is a gap between the prediction accuracy of our method and the original convolutional deep belief network (CDBN). However, the gap dramatically gets smaller with the increase of the dataset cardinality. In addition, our method outperforms the state-of-the-art dPAH and pCDBN in most of the case, and the results are statistically significant.

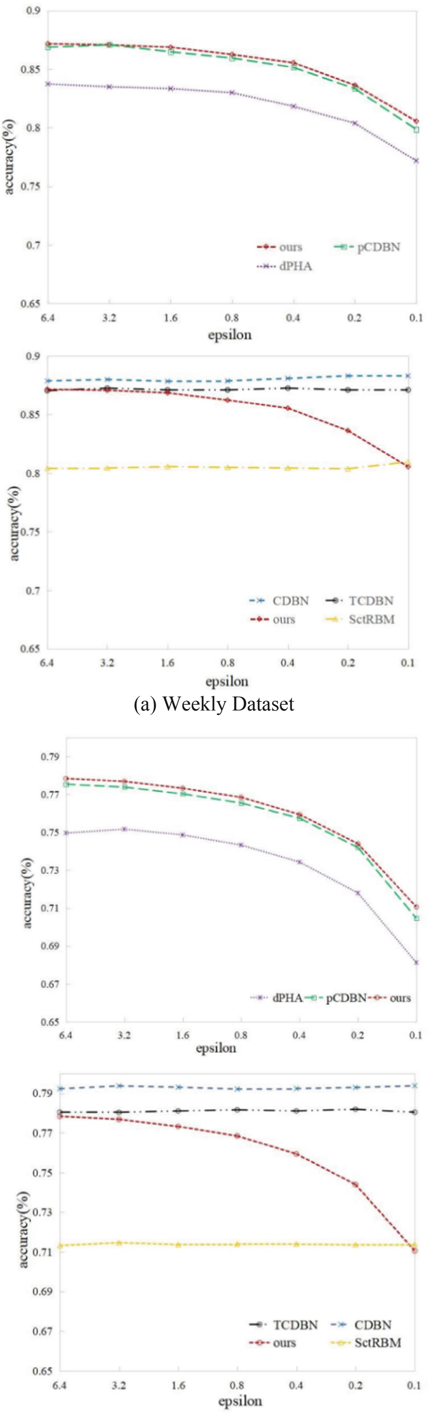
Figure 4 shows Our approach compare with no privacy protection. It has reduced compared to methods without privacy protection because differential privacy is achieved by adding noise and can result in a loss of accuracy.



(a) Weekly Dataset

(b) Daily Dataset

Fig. 3. Prediction accuracy vs. dataset size



(a) Weekly Dataset

(b) Daily Dataset

Fig. 4. Prediction accuracy vs. dataset size (compare with no privacy protection)

7 Exploring Applications in the Work of Adopting Group Standards

From the application area of the method, the deep differential privacy protection method based on multi-source data can be used for deep data mining and analysis in the field of standardization. For example, in the research work on the mechanism of adopting group standards for national standards, it is found that a large amount of group standard data needs to be used for collaborative filtering. According to the public data of the national group standard information platform, as of May 31, 2022, a total of 6276 social groups were registered in the national group standard information platform, and a total of 38712 group standards were published, which covered 19 national economic fields such as manufacturing, agriculture, construction, and education, thus covering a large amount of data both in terms of the number of group standards and the areas covered. Considering that group standards are standards developed by social groups in coordination with related parties, and their implementation and application are agreed to be used by their social groups themselves, privacy protection issues are inevitable in the process of data mining of group standards. Therefore, for preventing the privacy issues involved in the research of national standards adoption group standards from being leaked and maintaining a high privacy protection power, the deep differential privacy protection method based on multi-source data as a strict privacy definition can provide an effective solution to solve the privacy protection problems involved and has good application prospects.

8 Conclusions

In order to ensure that users' sensitive information is not leaked in deep model applications, this paper proposes a deep differential privacy protection method based on multi-source data. First, an upper limit of the corresponding privacy budget from the attacker's perspective. Second, the objective function of a conventional convolutional deep belief network (CDBN) is scrambled by an approximate polynomial representation obtained from Chebyshev's inequality. Finally, the experimental results show that the method achieves a better balance between user sensitive information protection and data availability in the training dataset, which ensures the correct rate and also effectively reduces information leakage. How to combine RNN and other sequence-based models to protect different types of datasets can be considered in the study of further work. As a new privacy protection method, there are still some difficulties in its theoretical derivation and practical application, and the future development direction still needs to be explored and studied in depth.

Acknowledgements. I am very grateful to my departmental leaders for their care, careful guidance and valuable advice. I am also grateful to my co-workers and friends for their patient listening and help, and for their care and concern in my life and studies.

References

1. Dwork C, F Mcsherry, Nissim K, et al. Calibrating Noise to Sensitivity in Private Data Analysis[J]. Proceedings of the VLDB Endowment, 2006.
2. Dwork C, Roth A. The Algorithmic Foundations of Differential Privacy [M]. 2013.
3. Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. arXiv:1406.2661, 2014
4. Hinton G E. Training products of experts by minimizing contrastive divergence. [J]. Neural Computation, 2002: págs. 1771–1800.
5. Phan N H, Wang Y, Wu X, et al. Differential Privacy Preservation for Deep Auto-Encoders: an Application of Human Behavior Prediction (AAAI-16) [oral presentation] [C]// Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI-16). AAAI Press, 2016.
6. Phan N H, Wu X, D Dou. Preserving Differential Privacy in Convolutional Deep Belief Networks[J]. Machine Learning, 2017, 106(9–10):1681–1704.
7. Radford A, Metz L, Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks[J]. Computer ence, 2015.
8. Shokri R, Shmatikov V. Privacy-preserving deep learning[C]// 2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton). 2015.