A Privacy-preserving Users' Power Load Prediction Method Based on Federated Learning

Kangqian Huang
Guangdong Power Exchange Center Co., Ltd
Guangzhou, China
huangkangqian@gd.csg.cn

Rui Zhou
Guangdong Power Exchange Center Co., Ltd
Guangzhou, China
zhourui@gd.csg.cn

Xin Hu
Guangdong Power Exchange Center Co., Ltd
Guangzhou, China
278271099@qq.com

Dejun Xiang
Guangdong Power Exchange Center Co., Ltd
Guangzhou, China
xiangdejun@gd.csg.cn

Yuhang Xie
School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen
Shenzhen, China
200110505@stu.hit.edu.cn

Jiaxin Lai
School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen
Shenzhen, China
200110515@ stu.hit.edu.cn

Jiaqi Deng
School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen
Shenzhen, China
200110717@ stu.hit.edu.cn

Yang Liu
School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen
Shenzhen, China
liu.yang@hit.edu.cn

***Abstract*—The new power system deeply integrates new digital technologies such as cloud computing, big data, Internet of Things, mobile Internet, artificial intelligence, etc., which enables different levels of the electricity grid to develop cooperatively and improves the ability to allocate power grid resources. Particularly, the smart meters located on the electricity consumers' side play an important role in collecting and aggregating users' data. Based on the collected data, users' power load data can be analyzed, which can efficiently allocate power resources and avoid resource waste caused by the limitation of power resource storage. However, the data collected by smart meters involves a large amount of users' private information. How to efficiently and legally use these data for joint data mining without violating users' privacy has become a challenge. In this paper, a privacy-preserving users' power load prediction method based on federated learning technology and RSA-AES encryption system is proposed. Experimental results show that our method can predict users' power load with high accuracy while protecting users' privacy.**

Keywords—Date privacy preserving, Smart Grid, Federated Learning

I. INTRODUCTION

Smart grids continually import information technology in the power system and become highly informative, automated, and interactive [1]. Compared with traditional power grids, smart grids enable two-way communication in all aspects from electricity consumption to power generation. The traditional power system is dominated by a top-down operation and management model, which often results in a certain waste of resources because the electricity demand of users does not reach the usage [2]. Power forecasting can to some extent provide the best power supply strategy for the power supplier to maximize resource utilization and avoid wastage. Today's smart grid systems have integrated, high-speed two-way communication networks [3]. Using hardware facilities such as smart meters and advanced equipment technologies, customer electricity consumption data can be collected so that the future electricity consumption of customers can be predicted based on these data. This enables power generators to dispatch resources more rationally according to demand, avoiding to a certain extent the waste of resources and making the whole distribution process more informative and intelligent.

Most of the power forecasting models already in the grid are centralized algorithms, the server-side collects information from device records by communicating with device terminals [4]. However, the data collected by smart meters and others contain a large amount of private user information. The process of collecting and transmitting electricity consumption data and other sensitive data from users by smart meters poses the risk of individual user privacy leakage [5]. Smart grids allow for a higher frequency of basic data collection than existing power systems [6]. There is a lot of data distributed around the network, and user privacy will be significantly impacted once an attacker intercepts it and extracts the plaintext data for analysis. Additionally, the attacker can also modify the stored data and communication messages by breaking the communication protocols, causing data integrity damage, which can easily cause system disorder and eventually lead to power paralysis [7]. Therefore, privacy protection techniques need to be utilized to protect the privacy of the process of data transmission.

Additionally, each regional power data center maintains its own data source, and for the sake of data privacy and data asset security, each party is hesitant to release its own data out of the data source. The resulting issue of "data silos" [8] creates challenges for collaborative data mining.

To address these problems, this paper introduces a federated learning architecture in the user power load prediction task, so that the data transmitted between the data centers in each region and the server are the predicted network model parameters, which do not involve sensitive user privacy data, and the network parameters are encrypted to further achieve the purpose of privacy protection.

This work is supported by the China Southern Power Grid Technological Project (No. GDKJXM20210105). Correspondence to: Yuhang Xie.

- Integrate Federated learning techniques, BiLSTM models, and RSA-AES encryption systems technologies and apply it to user power load prediction scenarios. The prediction task is accomplished while protecting the data privacy of individual data sources.
- Build a prototype system of the technique based on the above scheme and experimentally validate it on a real power data sets. The experimental results show that the average absolute percentage error of prediction under the federated learning architecture with a batch size of 64, selected client ratio of 0.2, and input of the previous seven days of electricity data as features is only 5.3%, which shows the related technique has good performance.
- Explore the possibility of applying privacy computing techniques represented by federated learning in the collaborative data mining of multi-source electricity market data, which provides a reference for promoting the common and compliant use of electricity market data.

II. Related Technologies

The user power load prediction scheme constructed in this paper mainly applies key technologies such as federated learning, LSTM, and RSA-AES encryption, and each related technology is briefly described below in turn.

A. Federated Learning

Federated learning was proposed by Google and first used to solve the problem of updating models locally by Android cell phone end users. It can carry out multi-party federated data mining and improve the efficiency of machine learning while ensuring data security and privacy protection.

B. LSTM and Variants

LSTM, proposed by Sepp Hochreiter et al [9], overcomes the problem of gradient explosion and gradient disappearance in backpropagation of RNN, is able to retain information for a longer period of time, and is suitable for long sequence data training. It is suitable for dealing with problems highly correlated with time series, such as speech recognition, machine translation, text generation, video tagging, etc., and is also applicable to power prediction problems.

BiLSTM consists of forward LSTM and backward LSTM, which are widely used in modeling temporal sequence models [10].CNN-LSTM extracts high-dimensional features with the help of CNN short sequence abstraction capability and then combines LSTM to synthesize high-dimensional features, which often has better performance in dealing with locally correlated temporal sequence data [11].

C. RSA-AES Encryption

RSA is a traditional asymmetric encryption algorithm and AES is a traditional symmetric encryption algorithm, both of which have high security. RSA is widely used in key distribution and management in large-scale networks, network digital signatures, etc. AES is used in network secrecy systems, financial secrecy, etc.

III. Overview of Prediction Schemes Based on Federated Learning and RSA-AES Cryptography

Based on the overview of related techniques, this paper gives a federated learning power load forecasting scheme based on the new power market.

In this paper, we use the classical Fedavg framework [12] in federated learning to study its applicability and effectiveness in the power load forecasting problem under the new power market. Meanwhile, considering the excellent performance of LSTM neural networks in time-series prediction, therefore we consider using LSTM networks and their variants as prediction model, and finally select BiLSTM as the final prediction model through experimental comparison.

For the privacy protection problem of data transmission between server and client, we select RSA-AES encryption system, which is described as follows. Under the multi-party collaborative data mining architecture of this paper, two main aspects of data transmission are carried out. One is the uploading of locally trained network parameters from the client to the server, and the other is the distribution of aggregated network parameters to each client by the server.

Among them, the data transfer from the client to the server is an N to 1 mode, so consider using the asymmetric encryption RSA algorithm, i.e., that is the client encrypts the network parameters through the shared public key distributed by the server, and the server decrypts the data through the private key and aggregates the network model. It can ensure the security of information transmission, prevent information theft among clients, and reduce the key management overhead.

For the process of data transmission from the server to the client, since the aggregated network model distributed by the server to each client is the same, there is no problem of mutual information theft among different clients. Meanwhile, symmetric encryption algorithm has faster encryption and decryption speed compared with asymmetric algorithm, therefore, the AES algorithm is used to encrypt the data transmitted from the server-side to the client-side. The final constructed prediction model based on federated learning and RSA-AES encryption technique is shown in Fig. 1.

IV. Power Forecasting Time-Series Modeling

This section introduces the construction and screening of feature combinations in power forecasting time series models, and the comparison and selection of forecasting models.

A. Feature Construction Method

In this paper, the Kaggle London Electricity Dataset [13] is used as the dataset for the experiment, which includes 112 block electricity consumption data, and the average value of daily electricity load is used as the prediction and processing index. Also according to the related literature [14], it is known that residential electricity consumption is mainly influenced by climate, economy, industrial characteristics and other factors. For example, the occurrence of high and low temperatures

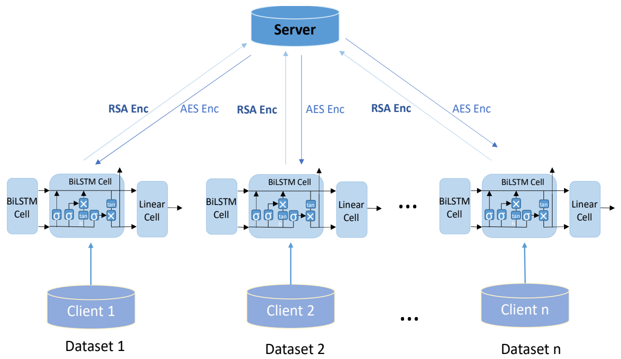


Fig. 1. Prediction model based on federated learning and RSA-AES Encryption Technology.

always leads to an increase in the use of high-powered electrical appliances such as air conditioners or heaters; the holding of large-scale events during holidays leads to an increase in electricity consumption, etc. Therefore, we consider adding climatic factors and holiday factors to the features of the training data for verification.

We consider climate factors, temperature indicators and comprehensive climate indicators respectively. The temperature indicators are constructed as follows, combining the highest apparent temperature and the lowest apparent temperature of the day for clustering, and finally the temperature indicators are clustered into 3 categories as shown below. The integrated climate indicators are constructed as follows, firstly, using the correlation coefficient matrix to filter out the climate indicators wind speed, humidity, temperature which have greater correlation with the power load, in order to comprehensively consider the impact of various weather factors on the power load, the above climate indicators are clustered, and finally a reasonable number of clusters is selected as 3 categories, so as to comprehensively measure the climate indicators. The effect of clustering under humidity and temperature indicators is shown in Fig. 2.

B. Prediction Model and Feature Combination Selection

To further choose the prediction models and features, Mean Absolute Percentage Error (MAPE) is considered as the prediction model measure and combined with the features of temperature feature, climate feature, holidays, etc., in three network architectures of BiLSTM, CNN-LSTM, CNN with 2 LSTM layers. The following experimental results are obtained for a block in the BiLSTM, CNN-LSTM, and CNN with 2 LSTM layers architectures, as shown in Table I.

Where the load is the power load data of the first seven days, temp is the real apparent temperature data, max & min temp are the maximum apparent temperature and minimum apparent temperature, temp cluster is the temperature clustering index constructed above, and month is the month information of the power data.

According to the experimental results, the prediction accuracy of each feature constructed is better on the BiLSTM model, and the best results are obtained by considering only holiday features and power load data. The combination of features considering climate indicators and power load data performs best on the remaining two prediction models, CNN-LSTM, CNN with 2 LSTM layers. On the contrary, adding too many features will make the prediction accuracy of the BiLSTM model decrease, intuitively because the introduced features are not very correlated with the power data, which will introduce noise to a certain extent and affect the final performance of the LSTM model.

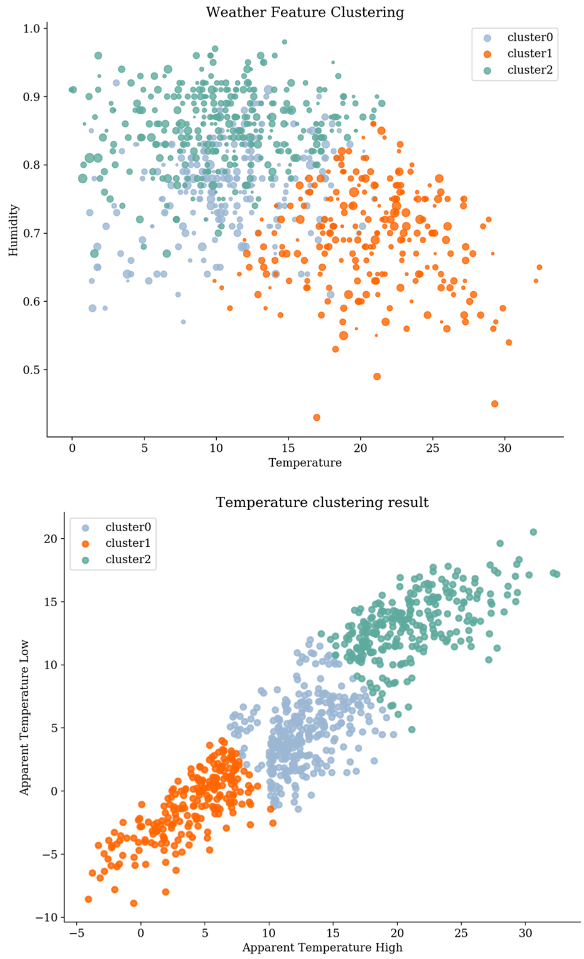


Fig. 2. Visualization of temperature and weather feature clustering

In summary, BiLSTM is selected as the final prediction model considering the comprehensive indexes of prediction accuracy, parameter size, and training difficulty. Meanwhile, to further consider the performance of the model on the federated architecture, four feature combinations load; load and holiday; load and weather; load, holiday and weather with better local prediction results further experiment under the Fedavg federated learning framework, and the conditions selected for the experiment are client-side ratio frac of 0.1. The experimental results are shown in Table II, where the number of server-side aggregation rounds is 10, the number of client-side local training rounds is 10, and the batch size is 10.

Finally, the combination of load, holiday and weather features with the least loss is selected as the final training feature.

TABLE I. RESULTS (WITH MAPE) OF FEATURE COMBINATIONS VS. PREDICTION MODELS

Item	Loss with MAPE		
	BiLSTM	*CNN-LSTM*	*CNN with 2 LSTM layers*
load	0.055 105	0.120 890	0.114 409
load, holiday	**0.053 979**	0.116 527	0.109 729
load, weather	0.058 405	0.114 559	0.105 318
load, holiday, weather	0.059 032	0.117 922	0.106 788
load, max & min temp	0.093 351	0.196 122	0.211 764
load, max temp	0.079 027	0.202 758	0.204 198
load, temp cluster	0.060 588	0.157 226	0.141 426
load,temp cluster,holiday	0.073 822	0.172 785	0.153 855
load, month	0.074 379	0.141 990	0.138 111
load, month, holiday	0.073 664	0.132 833	0.138 996

TABLE II. RESULTS (WITH MAPE) OF DIFFERENT FEATURE COMBINATIONS ON BILSTM

Feature Combinations	MAPE
load	0.057 6280
load, weather	0.068 3923
load, holiday, weather	**0.056 7945**
load, holiday	0.062 6221

V. FEDERATED LEARNING MODEL MODELING BASED ON FEDAVG ARCHITECTURE AND RSA AES

This section mainly introduces the construction details of RSA-AES encryption system under Fedavg federated learning framework and the algorithm flow of power load forecasting model.

A. *Encryption Algorithm Selection and Details*

For the privacy protection of data transmission between the clients and the server, RSA-AES is finally selected to encrypt network parameters by comparing various privacy protection schemes, as described below.

Since BiLSTM is selected as the prediction model, the mainstream algorithms such as Paillier's semi homomorphic encryption in federated learning [15] may fail in such tanh, $\sigma(\cdot)$ complex calculations. Meanwhile, it is found through experiments that the prediction results using the scheme of adding Laplace noise and performing gradient Clip limiting are poorer, and the average MAPE reaches 0.2264 in each client. this is due to the fact that in the power system scenario, the client datasets, the local training rounds, and the batch size are different, resulting in the uploaded gradient information is not synchronized. And due to the small number of clients, the addition of Laplace noise causes a certain offset in the final aggregated gradients and affects the update of the aggregated model. Finally, the scheme of aggregating the network parameters trained on the clients is chosen, and it is found that its average MAPE across the clients reaches 0.0568, which has a significant gain on the prediction model results compared to the gradient aggregation scheme. For the network parameter aggregation scheme, the RSA-AES encryption scheme is finally selected, of which the details are as follows.

- Information coding. In order to facilitate the encryption of network model data, it is necessary to encode each network parameter. In this paper, utf-8 encoding format is adopted to convert it into a plaintext format that can be encrypted by the RSA-AES encryption algorithm.
- Key generation. For AES encryption, a strong key with a length of 32 bytes is used for ECB mode packet encryption; For RSA encryption, 2048 bit RSA algorithm is used to generate corresponding public key and private key.
- Ciphertext generation. According to the key and the corresponding algorithm, the plaintext is encrypted to get the ciphertext
- Key distribution and holder. The client holds the AES key and the RSA public key distributed by the server, and the server holds the AES key and the RSA private key

B. *Security Analysis of Encryption Algorithm*

The federated learning framework provides privacy protection for participants, who only need to upload local model parameters without exposing local sensitive information. However, malicious attackers can still attack the federated learning process through poison attacks [16] and privacy attacks. Therefore, the model of this paper analyzes the security of these attacks.

- Data poisoning. It is mainly used to maliciously upload incorrect data to the client or server to slow down the convergence speed of the model or destroy the correctness of the model. In this scheme, data uploaded by the client is encrypted using the RSA public key. An attacker does not know the model parameter form cannot generate malicious data ciphertext based on the stolen public key and ciphertext, which can defend against poisoning attacks to a certain extent.
- Privacy attack. It means that an attacker steals the global model parameters shared by the central server, deduces the model parameter data of the participant, and then carries out further poisoning attacks. In this scheme, AES is used to encrypt the model parameters delivered by the server, so that the attacker who does not hold the AES key cannot obtain the plaintext parameters. Therefore, the parameters form and text of other clients cannot be inferred, and privacy attacks can be avoided to a certain extent.

C. *Fedavg Federated Learning Architecture Analysis*

1) Server

First, the server uses the RSA and AES key generation algorithms to generate keys and distribute the generated RSA public key and AES key to each client. And then encrypts and decrypts network parameters using the RSA public key and AES key in the process of parameters' transmission. Due to the limitations of the client's local upload environment, network transmission environment, transmission distance and so on [17], the server generally cannot obtain all the model parameters of the client during each round of aggregation. Therefore, in the experiment, a certain proportion of clients should be selected for local training in each round. After clients upload their own encrypted network parameters, the server uses the RSA private key to decrypt the uploaded parameters, then calls Fedavg algorithm to aggregate the model parameters to get the final corresponding mining network model. Finally, the server uses AES key to encrypt the aggregation model parameters and distributes them to all clients. The above steps were repeated until meeting the set round. The specific algorithm flow was shown in Fig.3.

2) Clients

For the clients, they first decrypt the encrypted network parameters from the server by AES key and load them to the local model. Then the server selects a certain proportion of clients for local update. The selected clients finally encrypts the trained local model parameters by RSA public key and transmits them to the server for aggregation. The detailed algorithm process is below.

The final architecture is shown in Fig. 4

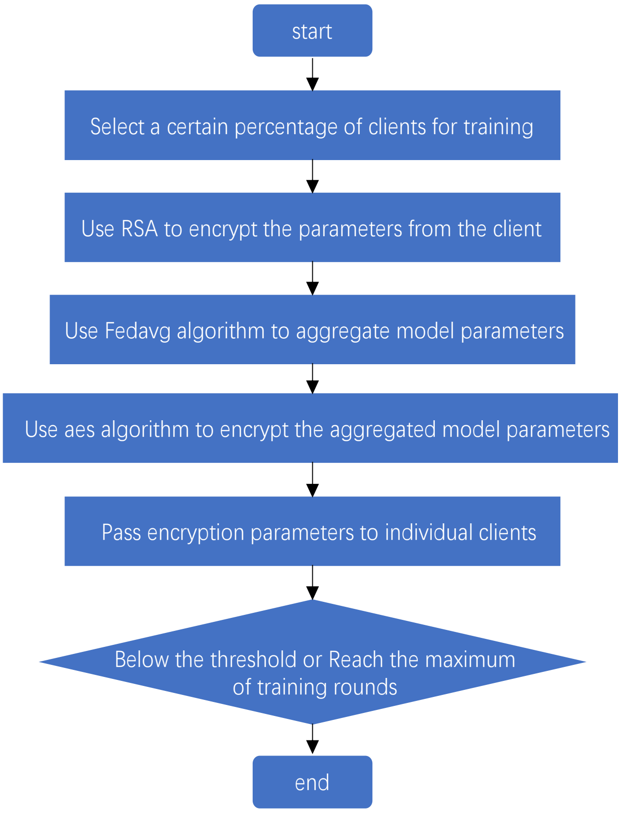

start
Select a certain percentage of clients for training
Use RSA to encrypt the parameters from the client
Use Fedavg algorithm to aggregate model parameters
Use aes algorithm to encrypt the aggregated model parameters
Pass encryption parameters to individual clients
Below the threshold or Reach the maximum of training rounds
end

Fig. 3. Flowchart of the server-side algorithm

Algorithm 1: Server

Input: N : the number of clients; $\{daatset_i|_{i=1}^{n}\}$: datasets of different clients; η : learning rate of clients' training; c : the frac of clients selected; m,n : the RSA and AES key generator bytes; L : Loss Function; M : model; $\boldsymbol{\Theta}$: parameters of neural network; tol_epochs: the total epochs of server's aggregation; local_epochs: the local epochs of clients' training

Output: M : the final aggregated model; $\overline{Loss}$: average loss of all clients

```
Initialize the network parameters $\boldsymbol{\Theta}$ and RSA-AES key
  $PubK_{rsa}, PrivK_{rsa} = \mathbf{Key\_Gen}_{rsa}(m)$;
  $K_{aes} = \mathbf{Key\_Gen}_{aes}(n)$;
Distribute $K_{aes}$, $PubK_{rsa}$ to each client
Pass encrypted network parameters $\widetilde{\boldsymbol{\Theta}} = E_{aes}(\boldsymbol{\Theta})$ to each client
while: $num \le$ tol_epochs do
    Randomly select $cN$ clients to call Client_update(i)
    Collect the network parameters and then decrypts them
      $\{D_{rsa}(\widetilde{\boldsymbol{\Theta}}_i)\} = \{\boldsymbol{\Theta}_i|_{i=1}^{cN}\}$
    Call Fedavg($\{\boldsymbol{\Theta}\}$) to aggregate $\{\boldsymbol{\Theta}\}$, use $PubK_{rsa}$ to encrypt it and
      then distribute it to all clients
end
    Return the final aggregated model and average loss on all clients.
```

Algorithm 2: Fedavg

Input: $\{\boldsymbol{\Theta}\}$: parameters from clients

Output: M : aggregation model

```
Calculate average network parameters $\overline{\boldsymbol{\Theta}}$
  $\overline{\boldsymbol{\Theta}} = \frac{\sum_{i=1}^{cN} \boldsymbol{\Theta}_i}{|\{\boldsymbol{\Theta}\}|}$
Load $\overline{\boldsymbol{\Theta}}$ into aggregation model $M$
Return $M$
```

Algorithm 3: Client_Update

Input: i: index of clients; $dataset_i$: dataset of client$_i$; γ : parameter of StepLR

Output: $\widetilde{\boldsymbol{\Theta}}$: the encrypted model parameters; $Loss$: loss on test data

```
Use AES key to decrypt the aggregated model from server and load it to local
  model $M_i$
Divide the $dataset_i$ into $\{\mathbf{D}_{train}\}, \{\mathbf{D}_{val}\}, \{\mathbf{D}_{test}\}$
while: $num \le$ local_epochs do
    $\eta_{num} = \mathbf{StepLR}(\eta_{num-1}, \gamma)$
    $\boldsymbol{\Theta}_{idx}^{(num)} = \mathrm{adam}(\boldsymbol{\Theta}_{idx}^{(num-1)}, \eta_{num})$
end
    Calculate $Loss$ on $\{\mathbf{D}_{test}\}$
    Use $PubK_{rsa}$ to encrypt final model parameters $\widetilde{\boldsymbol{\Theta}} = E_{rsa}(\boldsymbol{\Theta})$
    Return $\widetilde{\boldsymbol{\Theta}}$ and $Loss$
```

VI. Model Performance and Parameter Verification

This section mainly analyzes the performance and verifies the parameters based on the model built above.

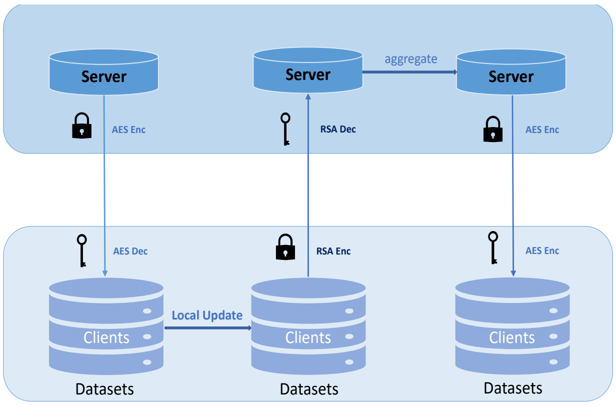


Fig. 4. Framework of federated learning based on RSA-AES Encryption Technology

A. Model Performance

Due to the limited local data of each client in smart meters, federated learning can be used to realize corresponding data mining under the premise of protecting user data privacy, so as to improve the model effect of each client. Its essence is similar to the bagging or stacking in ensemble learning, which aggregates the base learners to obtain a strong learner with better prediction effect and greater robustness.

To this end, We design an experiment to compare the difference between the federated learning model and the client local training model. Under Fedavg algorithm, the proportion of clients selected in each round is 0.2, the number of local training rounds is 10, and the number of aggregation rounds is 15. In the local training model, the local training epochs of the client is 50. And we compare results with the batch size for the client training of 16,32,64. The experimental results are shown in Figure 4. The selected evaluation index is MAPE, and the boxplot shows the losses on 112 blocks in Fig.5

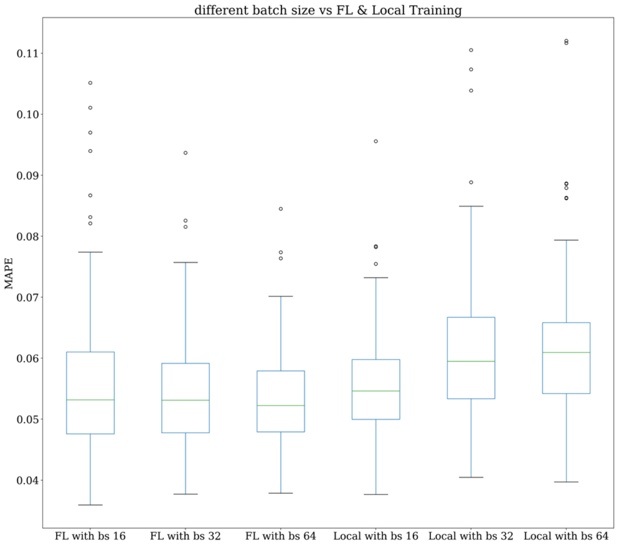


Fig. 5. MAPE on different batch size vs FL and local training.

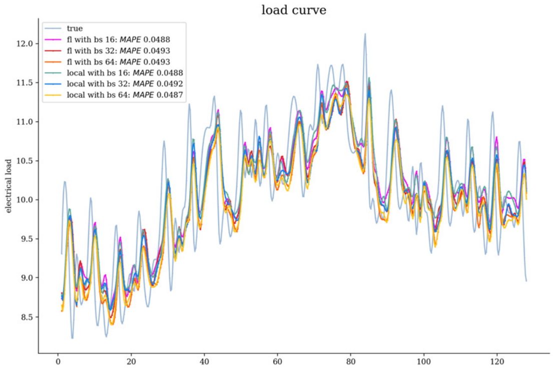
It can be observed that federated learning architecture reduces model losses to a certain extent compared with client-side local training, which indicates the effectiveness and feasibility of using federated learning architecture. At the same time, with the increase of batch size, local training model of loss is on the rise. According to the related literature [18], this is due to the fact that an increase in batch size under the same epoch condition will lead to a weaker model generalization and affect the prediction of the model, so more training rounds will be required to achieve the same accuracy as small batches, and the experimental results are consistent with this conclusion. However, by comparison, it is found that the federated learning architecture exhibits lower model prediction loss for larger batch size. This is due to the fact that using smaller batch size will result in client training models more closer to the local optimization and will retain more individual network parameters of each client when the server is aggregating training models, which will weaken the generalization ability of the aggregated model. This is verified experimentally in a later section.

To sum up, the federated learning architecture, not only can reduce the loss of model prediction, but also performance better on the larger batch size, less client training rounds, which can save a lot of training time and improve training efficiency. It shows the superiority of the model in prediction ability and computational cost. Specific experimental data for the conclusion is as shown in Table III.

To further observe the prediction effect, we select certain clients to observe the prediction effect under different training methods. The specific prediction effect is shown in Fig.6

TABLE III. COMPARISON OF INDICATORS BETWEEN FEDERATED LEARNING AND LOCAL TRAINING

Training Methods	**Loss with MAPE**		
	MAPE	***Client training rounds***	***Batch size***
FL with bs 16	0.056 268	3 300	16
FL with bs 32	0.054 779	3 300	32
FL with bs 64	**0.053 824**	3 300	64
Local with bs 16	0.055 377	5 600	16
Local with bs 32	0.061 213	5 600	32
Local with bs 64	0.062 112	5 600	64



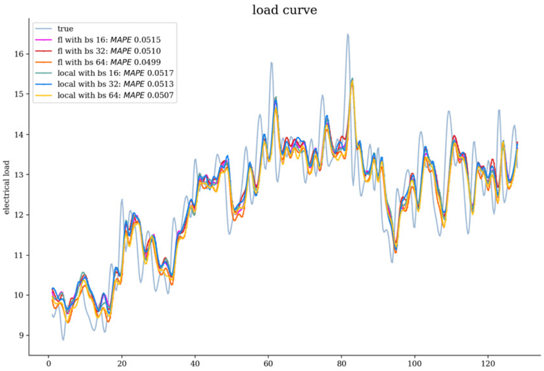


Fig. 6. Prediction effect of server's aggregation model on clients

B. *Parameter Validation*

According to the above results, the federated learning architecture has good results in improving the prediction accuracy and reducing the calculation cost. In order to further verify the influence of various super parameters on the model performance, multiple sets of experiments will be used to verify the parameters of the federated learning architecture and network model respectively. Meanwhile, in order to comprehensively verify the model performance from various aspects, we adopt three evaluation indexes which are commonly used in time series models, Mean Absolute Error (MAE) and Mean Square Error (Mean Square Error). MSE) and Root Mean Square Error (RMSE) to evaluate the performance of the prediction model.

1) *Federated Learning Parameter Validation*

According to relevant literature [19], in the federated learning architecture, the client proportion frac selected by the server in each round has a great impact on the model. Now, adjust it to study the effect of the prediction model.

The experiments were carried out under the given experimental conditions tol_epochs=10, local_epochs=10, and local_bs=10, as shown in Table IV.

According to the result, when the frac is small, the error indexes of the client test set are large, mainly due to the aggregation model can only aggregate a small number of client model features when the proportion is too small, which will lead to the weakening of the generalization ability of the final aggregation model and poor prediction effect on each client. As frac increases, model performance increases as the server is able to aggregate more client network parameters and thus improve model generalization. However, experiments show that when frac=0.3, the accuracy of the model decreases to a certain extent, mainly due to the randomness of client selection. When frac is small, the robustness of the model is not strong, but it still maintains a good prediction level. When frac increases to between 0.5 and 1, the prediction model converges basically. Thus in federal power forecast, select an appropriate proportion of the client can not only increase performance prediction models , but also consider the actual transmission environmental restrictions , which shows the value in practical applications.

TABLE IV. COMPARISON OF MODEL PERFORMANCE OF DIFFERENT FRAC

Frac	**Model Performance**		
	MSE	***MAE***	***RMSE***
frac=0.05	1.721 230	0.946 963	1.126 989
frac=0.1	0.975 922	0.733 500	0.907 802
frac=0.2	0.646 939	0.606 538	0.775 047
frac=0.3	0.816 110	0.675 138	0.851 554
frac=0.5	0.605 626	0.591 279	0.752 601
frac=1.0	**0.594 195**	**0.584 590**	**0.746 107**

TABLE V. COMPARISON OF MODEL PERFORMANCE OF DIFFERENT LOCAL BATCH SIZE

Local Batch Size	**Model Performance**		
	MAPE	***MAE***	***RMSE***
Local bs=8	1.393 269	0.755 277	0.931 049
Local bs=16	0.883 963	0.716 962	0.878 010
Local bs=32	0.746 664	0.649 366	0.821 717
Local bs=64	**0.630 847**	**0.598 589**	**0.768 157**

TABLE VI. COMPARISON MODEL PERFORMANCE OF DIFFERENT SEQUENCE LENGTH

Sequence Length	**Model Performance**		
	MAPE	***MAE***	***RMSE***
Seq len=1	0.978 752	0.737 163	0.930 145
Seq len=3	0.746 428	0.656 457	0.829 784
Seq len=7	**0.608 336**	**0.588 077**	**0.753 775**
Seq len=10	0.919 264	0.714 374	0.909 587
Seq len=14	0.702 965	0.635 727	0.809 039
Seq len=30	1.068 746	0.782 435	0.985 521

2) *Neural Network Validation*

In the process of network model parameter adjustment, the parameters that need to be verified mainly include the client training batch size local_bs and the number of days of input power data seq_len. The following results are obtained by adjusting the parameters.

According to the local training batch size local_bs, under the given experimental conditions frac=0.1, tol_epochs=10 and local_epochs=10, multiple experiments were conducted and the average values were taken, as shown in Table V.

Experiments show that in the federated learning architecture, larger batch size can not only bring faster training speed, but also enhance the accuracy of the prediction model to a certain extent, which confirms the above mentioned relevant content. At the same time, it further demonstrates the effectiveness of the federated learning architecture in the power load prediction.

For seq_len, 1,3,7,10,14, and 30 were selected as input to predict the power load data of the next day respectively. Experiments were conducted under the conditions of frac=0.5, tol_epochs=10, local_epochs=10 and local_bs=10. The results are shown in Table VI.

Through the experiment, the model loss reaches smallest when seq_len equals seven. Generally speaking, power data will show certain periodic changes. Too short an input sequence will lead to insufficient information acquisition and reduced model performance; too long an input sequence will introduce more noise to interfere with model prediction effect. Therefore, appropriate sequence length should be selected through experiments to improve the prediction effect of the model. To sum up, the model in this paper uses the sequence length with the best prediction effect in the experiment, namely the power data of the first seven days, as the final input.

VII. Conclusion

In this paper, we propose a power prediction scheme based on federated learning and RAS-AES encryption technology targeted at new power system. Then Kaggle London electricity data is used as the experimental dataset to validate the method. Our experiments on real-world data show that our framework with specific hype-parameters and federated structure is 13.3% lower than that of conventional local training in terms of loss. In addition, time consumption of our framework is much lower than conventional training process. To better apply it, we could consider to improve time efficiency and prediction accuracy in future work. Our framework may be combined with differential privacy techniques to reduce time consumption and more efficient data conversion methods such as carrying out feature extraction on time series data to improve prediction accuracy.

Acknowledgment

This work is supported by the China Southern Power Grid Technological Project (No. GDKJXM20210105). Correspondence to: Yuhang Xie.

References

[1] Siano P. Demand response and smart grids—A survey[J]. Renewable and sustainable energy reviews, 2014, 30: 461-478.

[2] Wu J, Ota K, Dong M, et al. Big data analysis-based security situational awareness for smart grid[J]. IEEE Transactions on Big Data, 2016, 4(3): 408-417.

[3] Lo C H, Ansari N. The progressive smart grid system from both power and communications aspects[J]. IEEE Communications Surveys & Tutorials, 2011, 14(3): 799-821.

[4] Ozkan M B, Karagoz P. A novel wind power forecast model: Statistical hybrid wind power forecast technique (SHWIP)[J]. IEEE Transactions on industrial informatics, 2015, 11(2): 375-387. processing [J]. Information Technology and Informatization, 2020(10): 217-219.

[5] Metke A R, Ekl R L. Smart grid security technology[C]//2010 Innovative Smart Grid Technologies (ISGT). IEEE, 2010: 1-7.

[6] Yu B, Mao W, Lv Y, et al. A survey on federated learning in data mining[J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2022, 12(1): e1443.

[7] Sepp Hochreiter, Jürgen Schmidhuber; Long Short-Term Memory. Neural Comput 1997; 9 (8): 1735–1780.

[8] S. Siami-Namini, N. Tavakoli and A. S. Namin, "The Performance of LSTM and BiLSTM in Forecasting Time Series," 2019 IEEE International Conference on Big Data (Big Data), 2019, pp. 3285-3292, doi: 10.1109/BigData47090.2019.9005997.

[9] Somu N, MR G R, Ramamritham K. A deep learning framework for building energy consumption forecast[J]. Renewable and Sustainable Energy Reviews, 2021, 137: 110591.

[10] Nilsson A, Smith S, Ulm G, et al. A performance evaluation of federated learning algorithms[C]//Proceedings of the second workshop on distributed infrastructures for deep learning. 2018: 1-8.

[11] https://www.kaggle.com/datasets/jeanmidev/smart-meters-in-london.

[12] Mawson V J, Hughes B R. Deep learning techniques for energy forecasting and condition monitoring in the manufacturing sector[J]. Energy and Buildings, 2020, 217: 109966.

[13] Min Z, Yang G, Sangaiah A K, et al. A privacy protection-oriented parallel fully homomorphic encryption algorithm in cyber physical systems[J]. EURASIP Journal on Wireless Communications and Networking, 2019, 2019(1): 1-14.

[14] Mothukuri V, Parizi R M, Pouriyeh S, et al. A survey on security and privacy of federated learning[J]. Future Generation Computer Systems, 2021, 115: 619-640.

[15] Li T, Sahu A K, Talwalkar A, et al. Federated learning: Challenges, methods, and future directions[J]. IEEE Signal Processing Magazine, 2020, 37(3): 50-60.

[16] Keskar N S, Mudigere D, Nocedal J, et al. On large-batch training for deep learning: Generalization gap and sharp minima[J]. arXiv preprint arXiv:1609.04836, 2016.

[17] McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data[C]. Artificial intelligence and statistics. PMLR, 2017: 1273-1282.

[18] Ke K, Hongbin S, Chengkang Z, et al. Short-term electrical load forecasting method based on stacked auto-encoding and GRU neural network[J]. Evolutionary Intelligence, 2019, 12(3): 385-394.

[19] Cai L, Lin D, Zhang J, et al. Dynamic sample selection for federated learning with heterogeneous data in fog computing[C]//ICC 2020-2020 IEEE International Conference on Communications (ICC). IEEE, 2020: 1-6.