Journal of Intelligent & Fuzzy Systems 47 (2024) 307–319
DOI:10.3233/JIFS-241801
IOS Press

A network security situation prediction approach based on MAML and BiGRU

Junfeng Sun, Chenghai Li and Yafei Song*
College of Air and Missile Defence, Air Force Engineering University, Xi'an, China

Abstract. Network security situation prediction is a complex task that typically requires extensive retraining of deep-learning models on vast amounts of sample data to achieve optimal performance. This paper proposes an innovative approach that integrates Model-Agnostic Meta-Learning (MAML) with Bidirectional Gated Recurrent Units (BiGRU) to address these challenges. Our method harnesses the BiGRU model's capability to learn from both preceding and succeeding conditions within network security prediction data, effectively extracting temporal information essential for prediction. This is complemented by Stochastic Gradient Descent for parameter updates, enhancing the model's adaptability and learning efficiency. Furthermore, the MAML algorithm is incorporated to facilitate the BiGRU model's swift adaptation to new tasks, thereby improving its generalization capabilities. The parameters are refined through a meta-learning process that calculates the sum of losses across multiple training instances and employs quadratic gradient descent for optimization. The empirical results of our approach demonstrate significant advancements, with goodness-of-fit decision coefficients of 0.926983 and 0.934452, representing a marked improvement of at least 18.0% and 15.8% over conventional deep learning models in the domain of network security situation prediction. This research novelty lies in the synergistic combination of MAML and BiGRU, which not only reduces the dependency on large datasets for retraining but also enhances the model's predictive accuracy and generalization to novel network security scenarios. It contributes a robust and efficient solution to the critical problem of network security situation prediction and paves the way for future advancements in cybersecurity defense mechanisms.

Keywords: Network security, situation prediction, meta-learning, neural networks, small samples

1. Introduction

Nowadays, Internet has become an integral part of everyday life, and which has become more evident since the COVID-19 pandemic. School students learn remotely through the Internet, company employees work remotely through the Internet, hospital doctors diagnose remotely through the Internet, and governments use health codes to keep track of people's movements and then provide support for epidemic prevention through big data technology. Cyberspace has become the fifth space after land, sea, sky, and outer space, carrying more and more human activities and becoming an indispensable and essential element in the development of human society [1]. According to the latest data from China Internet Network Information Center, by December 2022, China's netizens have been 1.067 billion. And in contrast with December 2021, 35.49 million netizens newly emerged. Internet penetration approached 75.6%, up 2.6 percent points compared with December 2021 [2]. Changes in netizen scale and Internet penetration recently are shown in Fig. 1.

As netizen scale continues to expand, network environment becomes more and more complex, and network security problems occur frequently. According to the data from National Internet Emergency Response Center on August 19, 2022 [3], the overall evaluation of the country's Internet network security status in July 2022 was good, and 3,713 domestic websites were tampered with; 1,960

*Corresponding author. Yafei Song, College of Air and Missile Defence, Air Force Engineering University, Xi'an, 710051, China. E-mail: yafei_song@163.com.

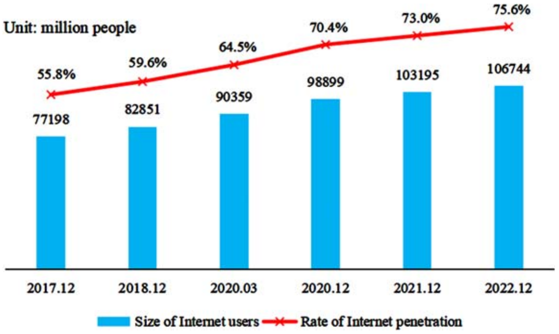


Fig. 1. Netizen scale and Internet penetration.

domestic websites were implanted with backdoors; 7,740 domestic website pages were counterfeited. National Information Security Vulnerability Sharing Platform collected and collated 2,066 information system security holes, 730 high risk holes included, 1,553 zero-day holes, and 1,613 holes for remote attack. These security issues not only threaten government departments, banks, and other important information system sectors but also threaten education, telecommunications, self-publishing, and other related industries and pose a significant threat to individuals' daily lives.

Network security situation prediction (NSSP) is very complex, and deep-learning model usually needs countless sample data retrained to achieve good performance. To address these problems, this paper combines meta-learning and deep learning methods to enable model's quick adaptation to small-sample tasks as a way to propose a prediction approach on the basis of MAML and BiGRU, and below are major contributions of the paper:

(1) Construct a network security situation prediction dataset using the weekly network security information and dynamic weekly reports from National Internet Emergency Response Center for experimental validation.
(2) Network security situation prediction method on the basis of MAML and BiGRU is proposed, which adds the MAML algorithm to the traditional BiGRU, thus allowing the model to adapt to new tasks in small sample situations quickly.
(3) By designing relevant experiments, we verify that the model shows significant effectiveness and feasibility in network security situation prediction, whose performance exceeds general deep learning models.

Our approach introduces a novel integration of Meta-Learning with Bidirectional Gated Recurrent Units (BiGRU) to address the challenges of data complexity and the need for retraining in deep learning models for network security situation prediction. This integration allows for rapid adaptation to new tasks and enhances the generalization performance, which is a significant advancement over existing methods.

Other sections: Section Two details relative works and current status of the research correlated with the paper. Section Three expresses the model in detail and the principle approach proposed in this paper. Section Four describes the experiment with findings. Section Five sums up the study and the prospects for further tasks.

2. Literal review and background work

2.1. Literal review

Network Security Situation Prediction technology differs from traditional prediction technology in that it can be considered a relatively proactive defense system [4]. By predicting the next stage of the network security situation to develop complementary strategies to defend against network attacks, the status quo of traditional security defense tools can be completely changed from passive defense to active defense, greatly upgrade defense with high success rate. As for network security situation prediction, it will no longer be partial to a corner but from a more macro perspective to calculate and evaluate. Analyzing the target network provides analysts with more macro and intuitive data to show the target network's security. Standard prediction methods include time series prediction methods [5], gray theory prediction, regression analysis, etc. However, the reality is that changing network security situation is a complex process because network attacks are often full of randomness and chance. When dealing with a nonlinear relationship, the above methods need to be revised and have gradually failed to meet the needs of network security situation prediction. Prediction methods as well as models on the basis of theories like Neural Networks, Markov chain [6], Support Vector Machine (SVM) [7, 8] have been discovered by various scholars one after another. Among them, Neural Networks is widely used for the prediction, belonging to artificial intelligence and can be processed in parallel while possessing excellent function fitting and self-learning capabilities. It has a high fault tolerance, providing solid data analysis and processing support.

2.2. Background work

2.2.1. Meta learning

Meta Learning is a concept initially introduced by Schmiduber in the 1990's [9]. Distinguished from machine learning, in which the data itself is the unit, meta-learning uses the task as the basic unit, with target to improving learning algorithm via multi-task learningfor quick adaptation to new task. Each task T has a task-related dataset D_T containing a query set D_Q and a support set D_S. Meta learning has a meta training phase and a meta testing phase. During meta training, model is trained by sampling numerous tasks, namely, Source Task, and during meta testing,model performance is evaluated using Target Task.

The Model-Agnostic Meta-Learning (MAML) algorithm is a perfect meta learning algorithm put forward by Finn et al. in 2017 for trained model using Gradient Descent [10]. The MAML algorithm is widely used to train a streamlined model that uses countable training samples for solving multi-task learning. Algorithm focuses on tuning the original model parameters by training countable tasks and continuously iterating gradient descent to show better generalization performance on new tasks. Its training process is shown in Fig. 2. In addition, the MAML algorithm is a fundamental framework for innumberable meta learning algorithms, with its application into numerous aspects for addressing challenges like data bottlenecks as well as generalization in deep learning. Liu et al. [11] applied meta learning for predicting stock prices and improved stock prediction accuracy by introducing the MAML algorithm for mitigating concept drift influence on predicting and providing precious guide to investor for reducing investing risks. Nie et al. [12] introduced meta-learning to cope with the problem of lack of interoperability and scalability of existing methods and models in the field of human activity recognition when activities and human bodies newly engaged in activities and statuses newly arise for rapidly adapting to human activity recognition in new statuses. Su et al. [13] applied meta-learning to a bearing failure diagnosis with countable samples in diverse operating environment by proposing a data reconstruction hierarchical recursive meta learning method to rapidly adapt to human activity recognition when fault samples are lacking. The fault diagnosis task achieved good results. Through these studies, we treat network security situation prediction as meta learning issue, with MAML algorithm application into network security situaiton prediction to improve model's prediction performance.

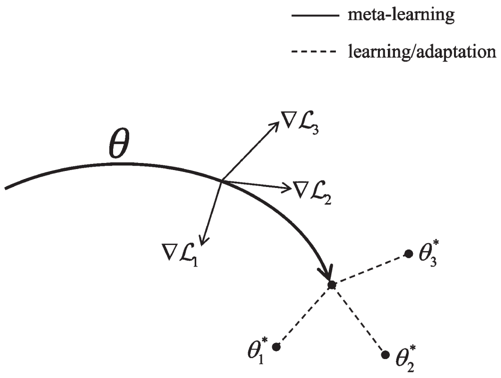


Fig. 2. MAML training process.

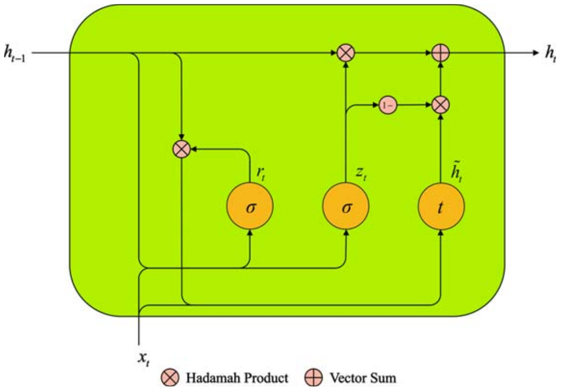


Fig. 3. Basic structure of GRU.

2.2.2. Gated recurrent unit

Gated recurrent unit (GRU) introduces the reset gate r_t and updates gate z_t concept, which modifies the calculation of hidden states $\tilde{h}_t$ in recurrent neural networks [14], with inner structure detailed as Fig. 3.

Where x_t denotes input information of current moment, h_{t-1} and h_t hidden state of previous and current moment separately. σ denotes sigmoid activation function. t denotes tanh activation function.

GRU network is calculated as follows:

$$r_t = \sigma\left(W_r \cdot \left[h_{t-1}, x_t\right] + b_r\right) \tag{1}$$

$$z_t = \sigma\left(W_z \cdot \left[h_{t-1}, x_t\right] + b_z\right) \tag{2}$$

$$\tilde{h}_t = \tanh\left(W_{\tilde{h}} \cdot \left[r_t \cdot h_{t-1}, x_t\right] + b_{\tilde{h}}\right) \tag{3}$$

Table 1
Comparison of related work with the present study

Approach/Feature	Traditional machine learning	Deep learning	Hybrid model	Meta-Learning (MAML-BiGRU) (This Study)
Approach Used	Statistical/Rule-based	Neural Networks	Neural Networks + SVM	Meta-Learning + Neural Networks
Key Features	Handcrafted Features	Learned Features	Time-series Analysis	Rapid Adaptation, Bidirectional Context
Data Requirements	Large Labeled Dataset	Moderate Labeled Dataset	Small to Moderate Dataset	Small Sample Learning
Model Training	Batch Learning	Backpropagation	Sequential Chunking	Few-Shot Learning
Adaptability	Low	Moderate	High	High
Generalization	Weak	Strong	Strong	Strong
Computational Efficiency	High	Moderate	Moderate	High
Robustness to Novel Threats	Weak	Moderate	Moderate	High
Application in Real-time	Limited	Limited	Good	Excellent

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t \tag{4}$$

$$y_t = \sigma(W_o \cdot h_t + b_o) \tag{5}$$

Where W and b represent weight matrix and bias term, respectively.

2.3. Comparison of related work with the present study

The comparison of related work with the present study is as Table 1.

Our proposed MAML-BiGRU model combines the strengths of meta-learning and deep learning to address the challenges of small sample learning and rapid adaptation to new tasks in network security situation prediction. Compared to traditional machine learning methods (Study A), our model eliminates the need for extensive manual feature engineering, instead learning useful representations directly from raw data. In contrast to pure deep learning approaches (Study B), our model significantly improves generalization performance on new datasets through the MAML algorithm, even when the number of samples is limited. Additionally, our BiGRU architecture effectively captures the dynamics of time-series data, which is crucial for real-time security situation prediction.

3. Problem description

In the context of network security, accurately predicting security incidents is a complex challenge due to the dynamic nature of cyber threats. This section describes network security incidents and the problem of network security situation prediction.

3.1. Network security incidents

Network security incidents refer to any unauthorized actions or occurrences that potentially compromise the security, integrity, or availability of network systems. These incidents may include, but are not limited to, virus attacks, network intrusions, data breaches, and denial-of-service attacks (DoS/DDoS). To effectively predict and respond to these incidents, it is crucial to define and document them accurately.

Network security incidents can be categorized based on their nature, scope, and severity. For instance, events can be classified as low, medium, or high risk based on their potential impact. Additionally, incidents can be further based on their origin (such as internal or external threats) and attack type (such as malware or social engineering).

3.2. Network security situation prediction

Situation prediction was first proposed by Endsley in 1988 as part of situation awareness. Situation awareness is defined as “cognition, understanding of environmental factors in a certain spatial and temporal context, and prediction of future trends.” It can be summarized in a classical three-layer model: situation perception, comprehension, and prediction in Fig. 4.

In 1999, Bass [15] put forward “network situation awareness”, which first applied situation

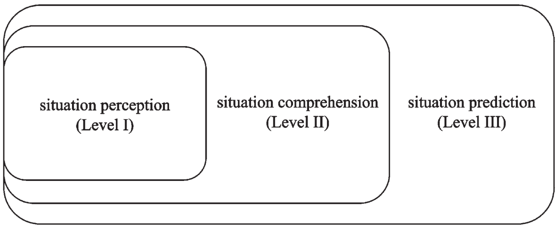


Fig. 4. Network security situation awareness model diagram.

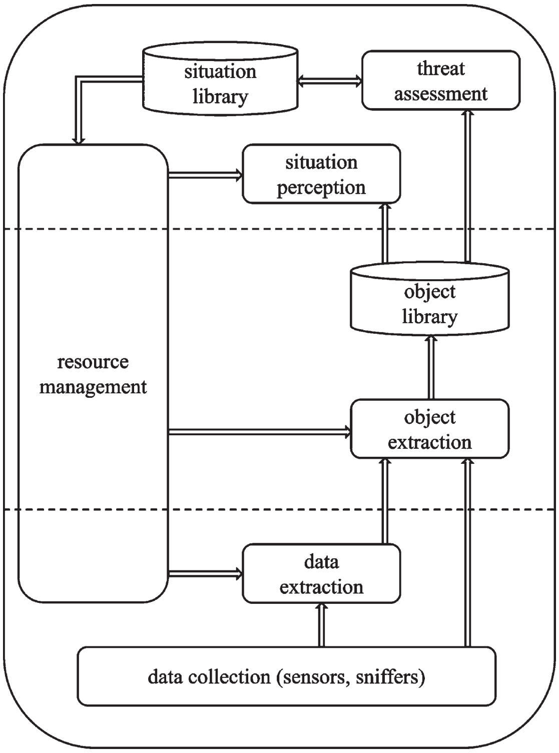


Fig. 5. Functional model of network security situation awareness on the basis of multi-sensor data fusion.

awareness to the field of cyberspace, and revealed "the next-generation intrusion detection system shall conduct data fusion from countless heterogeneous distributed network sensors for situational awareness in cyberspace," and proposed a functional model of network security situation awareness on the basis of Multi-sensor Data Fusion concerning the data-fusion model based on the U.S. military structure, as shown in Fig. 5.

Shi et al. [16] summarized network security situation awareness as "mining out various security elements in the network environment, processing and fusing them, forming a macroscopic security situation assessment on global network environment, and making time-series prediction on security situation further, which is a technical means to guarantee the network environment. It is a technical means for ensuring network environment security." Chang et al. [17] summarized network security situation awareness as "discerning the attack behavior in the network from innumerable noisy data and then fusing them to evaluate and monitor security situation of the network in real time for achieving comprehensive network control and providing a foundation for network managers' decision analysis for reducing network risks and losses."

Network security situation prediction is the final target to awareness, which is mainly based on acquiring and processing the situation information of historical data, establishing appropriate mathematical models to find potential development patterns among situation data, and thus reasoning to get developing trendency and status of situational situation further. Through situation prediction, qualitative or quantitative analysis can be conducted, and early warnings can be issued to provide a reference for security personnel to make decisions and further realize active defense of the network. The randomness and uncertainty of network attacks make the attack-based security situation change highly complex and non-linear, which brings excellent limitations to the traditional prediction model.

4. Methodology

Our research employs a novel approach to network security situation prediction by harnessing the power of Model-Agnostic Meta-Learning (MAML) in conjunction with Bidirectional Gated Recurrent Units (BiGRU). This section outlines the comprehensive methodology that underpins our MAML-BiGRU model.

4.1. Modeling framework based on MAML and BiGRU

This paper introduces a meta learning method on the basis of Bi-directional Gate Recurrent Unit (BiGRU) network. It constructs a network security situation prediction approach on the basis of MAML and BiGRU, with major three parts included: data input, BiGRU model, and meta learning network, and overall architecture is shown in Fig. 6.

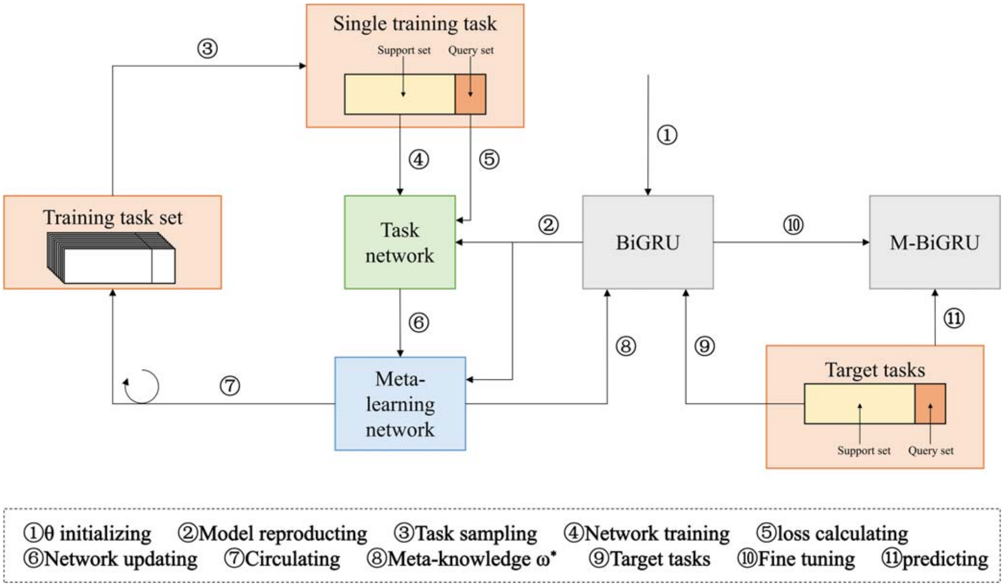

Single training task
Support set
Query set
③
④
⑤
①
Training task set
Task network
②
BiGRU
⑩
M-BiGRU
⑥
⑧
⑨
⑪
⑦
Meta-learning network
Target tasks
Support set
Query set
①θ initializing ②Model reproducting ③Task sampling ④Network training ⑤loss calculating
⑥Network updating ⑦Circulating ⑧Meta-knowledge ω* ⑨Target tasks ⑩Fine tuning ⑪predicting

Fig. 6. Model architecture.

Table 2
Data reconstruction results

Input samples	Output sample
$SA_1, SA_2, SA_3, SA_4, SA_5$	SA_6
$SA_2, SA_3, SA_4, SA_5, SA_6$	SA_7
…	…
$SA_{25}, SA_{26}, SA_{27}, SA_{28}, SA_{29}$	SA_{30}
…	…

4.2. Data input

In this paper, we choose the security data from National Internet Emergency Response Center [18] as data for experiment, and we obtain the posture values by performing the posture assessment weekly and then conduct normalization and sliding window processing. Finally, we convert them into the form of time step×input dimension. Taking the sliding window as an example, we reconstructed the data, with findings detailed as Table 2.

4.3. BiGRU model

The task of network security situation prediction is usually related to both before and after network status, and dependencies between network security situation data need to be considered. Therefore, to improve the prediction effect, this paper uses BiGRU network to predict security situation, thus information before and after the network state can be obtained simultaneously, and the features in the network security situation is able for full extraction. Bi-directional architecture of the network can acquire dynamic change of network state in a more detailed way, thus improving the accuracy of prediction.

GRU [19] is a commonly used gated recurrent neural network with a strong learning ability for long-term dependent information.

However, the GRU network can only better capture the forward feature information of the network security posture data, and the backward feature information cannot be obtained, so this paper chooses the BiGRU network for predicting network security situation. Meanwhile, to prevent overfitting problem, a dropout layer is introduced after each BiGRU layer for improving neural network performance.

The BiGRU consists of forward and backward GRUs superimposed on each other, with structure detailed as Fig. 7.

Using forward and backward GRU network, posture values of forward and reverse inputs are calculated separately for corresponding hidden layer state output $\boldsymbol{H}^{\rightarrow} = \left\{h_{L_1}, h_{L_2}, \cdots, h_{L_t}\right\}$ and $\overleftarrow{\boldsymbol{H}} = \left\{h_{R_t}, h_{R_{t-1}}, \cdots, h_{R_1}\right\}$. Then forward and backward hidden layer state output vectors are stitched together for final output of BiGRU network layer as follows:

$$\boldsymbol{H} = \left\{\vec{\boldsymbol{H}}, \overleftarrow{\boldsymbol{H}}\right\} \tag{6}$$

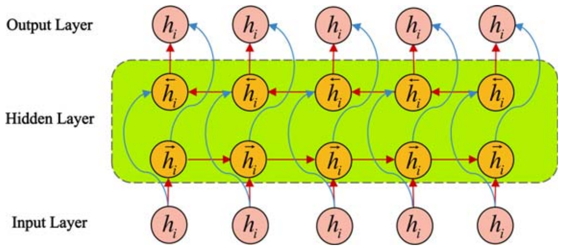


Fig. 7. BiGRU network model.

4.4. Meta-learning network

As an emerging technique in machine learning, meta learning is aimed at addressing the problem of how to learn new tasks quickly and accurately in a short time. In our research, the MAML algorithm as a meta-learning layer is introduced, with quick adaptation to new learning tasks and improve model's generalization performance to some extent by using previous learning experiences. At the same time, combining the MAML algorithm with neural networks and various loss functions enables the model to get better training results and has good application value.

The MAML algorithm, as an initialization method for the learner, has significant advantages, such as fast adaptation to new tasks and improved model generalization performance. Compared with traditional machine learning techniques, the MAML algorithm can update parameters based on computing only a tiny amount of data, thus achieving better results in the face of new learning tasks, that is, having the ability to learn to learn. When using BiGRU to predict cyber security posture, we split data into training and test set. In MAML-BiGRU method, support and query set correspond with training and test set.

In meta training phase, for each source task T, model calculates training error on support set D_S and updates θ by equation below:

$$\theta^T = \theta - \alpha \nabla_\theta L\left(D_S^T, \theta\right) \tag{7}$$

$$L\left(D_S^T, \theta\right) = \sum\nolimits_{(x,y)\in D_S^T} \|f_\theta(x) - y\|_2^2 \tag{8}$$

Where θ denotes model parameters, and θ^T denotes model parameters after updating on task T. α denotes the learning rate, ∇_θ the θ gradient of the query. $L\left(D_S^T, \theta\right)$ model training error with parameter θ as on support set D_S of the task T calculated via Equation (8), and f_θ model with parameter θ. For training error, $L\left(D_Q^T, \theta^T\right)$ denotes model testing error with parameter θ^T as on support set D_Q of the task T calculated via Equation (9).

$$L\left(D_Q^T, \theta^T\right) = \sum\nolimits_{(x,y)\in D_Q^T} \left|\left|f_{\theta^T}(x) - y\right|\right|_2^2 \tag{9}$$

$$\theta = \theta - \beta \nabla_\theta \sum\nolimits_T L\left(D_Q^T, \theta^T\right) \tag{10}$$

Where f_{θ^T} denotes model with parameter θ^T. Test error on entire source tasks constitutes training error in meta training phase, with parameters updated via Equation (10). β denotes learning rate in Equation (10).

In meta testing, model firstly calculates training error on support set D_S of target task via Equation (8) and updates parameters for fitting target task via Equation (7). After that, calculation is conducted on query set D_Q of target task for test error via Equation (9) for evaluating model performance. Each task contains same time range where the network security situation data are located, so its data distribution only produces a small change. After updating the parameters on the basis of training loss of target task, model's more adaptation to target data distribution will be realized, thus improving prediction accuracy.

4.5. Algorithm flow of MAML-BiGRU method

The specific algorithmic flow of the MAML-BiGRU method is detailed as Algorithm 1.

Algorithm 1. MAML-BiGRU method

```
Input: network security situation data sequence $SA = \{SA_1, SA_2, \cdots, SA_t, \cdots, SA_N\}$; sliding window size $s$; number of samples per task $D$; learning Rate $\alpha$, $\beta$; Number of sample pairs in support set $N_S$; Number of sample pairs in query set $N_Q$
Output: network security situation prediction value
for all $i \in [1, N - s]$ do:
    $W_i = \{SA_i, SA_{i+1}, \cdots, SA_{i+s-1}\}$ //Window sequence
end for
for all $t \in \{s, s+1, \cdots, N-1\}$ do:
    $W_{t-s+1} = \{SA_{t-s+1}, SA_{t-s+2}, \cdots, SA_t\}$ //As samples
    $SA_{t+1}$ //As labels
    $P_t = (W_{t-s+1}, SA_{t+1})$ //Formation of sample pairs
end for
Network security situation data sequence $SA$ is divided into different tasks by year. Each year sample forms a task $T_i$, and each task corresponds to a dataset $D_{T_i}$ containing several sample pairs $P_t$, where the first $N_S$ sample pair constitutes the support set $D_S$, and the remaining $N_Q$ sample pairs constitute the query set $D_Q$.
Divide all tasks into source task sets $Source\{T_i\}$ and target task sets $Target\{T_i\}$.
Randomly initialize the model parameters $\theta$ of BiGRU.
while not done do:
    Select a set of tasks from the source task set $Source\{T_i\}$.
    for all $T_i$ do:
        Calculate the training error $L\left(D_S^T, \theta\right)$ on the support set $D_S$.
        $\theta^T = \theta - \alpha\nabla_\theta L\left(D_S^T, \theta\right)$ //Update parameters by gradient descent
        Calculate the test error $L\left(D_Q^T, \theta^T\right)$ on the query set $D_Q$.
    end for
    $\theta = \theta - \beta\nabla_\theta \sum_T L\left(D_Q^T, \theta^T\right)$ //Update parameters by gradient descent
end while
for all $T_i \in Target\{T_i\}$ do:
    Calculate the training error $L\left(D_S^T, \theta\right)$ on the support set $D_S$.
    $\theta^T = \theta - \alpha\nabla_\theta L\left(D_S^T, \theta\right)$ //Update parameters by gradient descent
    Make predictions on the query set $D_Q$ and evaluate model performance.
end for
```

5. Experiment and analysis

5.1. Assumptions made

- **Data Distribution:** We assumed that the data used in our experiments is representative of the real-world network security situation, capturing the essential characteristics and dynamics of cyber threats and network traffic patterns. While we acknowledge that actual network environments may exhibit greater variability, our dataset is designed to encompass a wide range of typical scenarios.
- **Hyperparameter Selection:** For the hyperparameters of our MAML-BiGRU model, we selected values based on prior research and empirical evidence from our preliminary experiments. These selections aimed to optimize model performance while maintaining a balance between complexity and computational efficiency.

5.2. Data acquisition and environment configuration

5.2.1. Data acquisition

To apply MAML algorithm into network security situation prediction for model's quick adaptation to new data, difficulty of splitting the cybersecurity posture data into multitask must be solved firstly. The generation process of task distribution is shown below.

Table 3
Weight of cyber security threats

Network security threat	Weight
Number of hosts infected with viruses in the territory	0.30
Number of websites tampered with in the territory	0.25
Number of websites with backdoors in the country	0.15
Number of counterfeit pages on domestic websites	0.15
Number of new information security holes	0.15

(1) Dividing tasks and data sets. The network security information and dynamic weekly reports from National Internet Emergency Response Center are collected as data for experiment. In our research, we select 520 weekly reports from Issue 1 of 2013 to Issue 52 of 2022 being validation basis and divide each year as a separate network security situation prediction task and each year's data into a separate dataset. Eight years of data from 2013 to 2020 are taken to train the metamodel and two years of data from 2021 and 2022 are taken as new data sets to test the adaptive capability of the metamodel. To better assess network security situation, posture assessment method in Ref. [20] is employed to quantitatively assess the five network security threats. By assigning weights to the severity of network security threats so that the impact level of each threat can be better understood, the specific weight assignments are detailed in Table 3. based on the obtained weights and Equation (11) to calculate the weekly posture values. It is able to effectively enhance real-time cybersecurity accuracy and remind relevant personnel of in-time cybersecurity strategy adjustment.

$$SA = \sum_{i=1}^{5} \frac{NT_i}{NT_{i\,\max}} \cdot \omega_i \tag{11}$$

Where NT_i denotes number of a certain kind of security threat in a particular week (i indicates threat type), $NT_{i\,\max}$ max. number of that kind of threat in data chosen for each year, ω_i its weight.

(2) Sample extraction tasks. Each task draws the first 70% of continuous data as the support set and the last 30% as the query set.

5.2.2. Environment configuration

The MAML-BiGRU model and experiment conducted were under Pytorch Deep Learning Framework under given experiment condition in Table 4.

Table 4
Experiment environment configuration

Experimental environment	Specific configuration
OS	Windows 11
CPU	Intel(R) Core(TM) i5-11300 H @ 3.10 GHz 3.11 GHz
Memory	16GB
HD	2TB
Development framework	Pytorch 1.13.0
Development language	Python 3.9.12

Table 5
Model parameter settings

Model parameters	Parameter settings
Optimization algorithm	Adam
Learning rate of BiGRU	0.01
Learning rate of Meta-Learner	0.001
Batchsize	32
Number of BiGRU's neural units	128/128/128
Dropout	0.3

5.3. Experimental data pre-processing

Data normalization is one of the essential pre-processing techniques in machine learning. In practical applications, there are z-score normalization, min-max normalization and mean normalization methods. For our research, we choose min-max normalization, by which the feature data are normalized to –1 and 1, which reduces outlier effect and improves model convergence rate, and also improves model's ability of handling feature data.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{12}$$

Where x' indicates data x mapped to interval $[-1, 1]$. $\min(x)$ and $\max(x)$ denote min. and max. values in data set.

5.4. Model parameter settings

The specific parameter settings for the experiments are shown in Table 5.

5.5. Model evaluation indicators

To evaluate accuracy and stability of prediction model put forward, Mean Absolute Error (MAE), Mean Square Error (MSE), Mean Absolute Percentage Error (MAPE), as well as coefficient of determination (R^2) are used as evaluating indexes [21]. Among them, MAE and MSE can reflect the degree of model's prediction error, MAPE can eliminate the difference in magnitude and reflect model's relative prediction error, and R^2 refers to an essential indicator of model's generalization ability, which can reflect model effectiveness in data fitting. Calculation equation of evaluating index is shown below:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{13}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{14}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n} \left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{15}$$

$$R^2 = \frac{\left[\sum_{i=1}^{N} (y_i - \bar{y})\left(\hat{y}_i - \bar{\hat{y}}_i\right)\right]^2}{\left[\sum_{i=1}^{N} (y_i - \bar{y})^2\right]\left[\sum_{i=1}^{N} \left(\hat{y}_i - \bar{\hat{y}}_i\right)^2\right]} \tag{16}$$

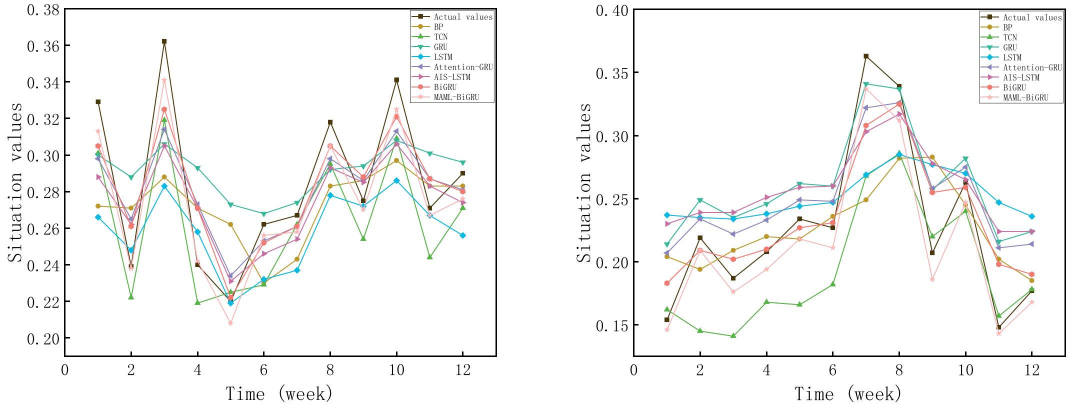
In the above four equations, y_i means actual value of a sample, $\hat{y}_i$ predicted value, N number of samples, $\bar{y}_i$ mean of actual value, and$\bar{\hat{y}}_i$ mean of predicted value.

5.6. Network security situation prediction

To effectively compare model prediction ability of proposed in our research and models by others, following experiments are performed: in same experiment environment and setting the sliding window number $s = 5$, eight models, BP, TCN, GRU [22], LSTM, Attention-GRU [23], AIS-LSTM [24], BiGRU, and MAML-BiGRU, are used to make predictions, and predicted value is compared to actual value, as shown in Fig. 8. The evaluation indexes of diverse models can be seen in Tables 6 and 7, and evaluation indicators sum of diverse models in Fig. 9.

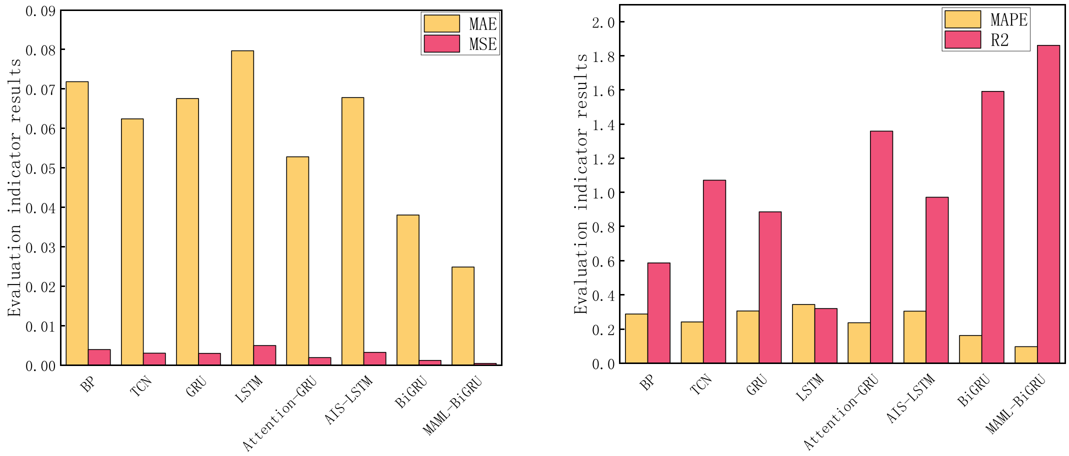
From Fig. 8, we can see that most models merely predict trendency of the network security situation. However, they can not predict the details accurately, while the MAML-BiGRU prediction model put forward introduces MAML algorithm, thus BiGRU model can better extract the relationship characteristics between time series by calculating only a tiny amount of data, which makes the prediction results accurate.

As seen in Table 6, in the 2021 data, compared to other models, MAE decreased by at least 42.2%,



(a) 2021 (b) 2022

Fig. 8. Comparison of the prediction situation values of different models.



(a) MAE, MSE (b) MAPE, R^2

Fig. 9. Evaluation indicators sum of different models in 2021, 2022.

Table 6
Evaluating indexes of diverse models in 2021

Model	MAE	MSE	MAPE	R^2
BP	0.033337	0.001456	11.578152%	0.189287
TCN	0.022728	0.000626	7.780618%	0.651633
GRU	0.030580	0.001261	11.153646%	0.298121
LSTM	0.030439	0.001512	9.849662%	0.158155
Attention-GRU	0.021012	0.000582	7.291407%	0.676065
AIS-LSTM	0.023966	0.000765	8.140430%	0.574005
BiGRU	0.017086	0.000386	5.927516%	0.785288
MAML-BiGRU	**0.009871**	**0.000131**	**3.327748%**	**0.926983**

Table 7
Evaluating indexes of diverse models in 2022

Model	MAE	MSE	MAPE	R^2
BP	0.038483	0.002455	17.113002%	0.398046
TCN	0.039677	0.002369	16.332662%	0.419154
GRU	0.036981	0.001683	19.412837%	0.587445
LSTM	0.049207	0.003420	24.432980%	0.161544
Attention-GRU	0.031794	0.001293	16.317713%	0.682848
AIS-LSTM	0.043864	0.002458	22.316875%	0.397210
BiGRU	0.020916	0.000788	10.196269%	0.806901
MAML-BiGRU	**0.014952**	**0.000267**	**6.352363%**	**0.934452**

MSE decreased by at least 66.1%, MAPE decreased by at least 43.9%%, and R^2 increased by at least 18.0%. As seen in Table 7, in the 2022 data, compared to other model, MAE reduced by at least 28.5%, MSE by at least 66.1%, MAPE by at least 37.7%, and R^2 increased by at least 15.8%.

As seen in Fig. 9, the MAML-BiGRU prediction model put forward owns minor error values and the most significant coefficient of determination in the data sum of 2021 and 2022, which has a significant advantage over other models and proves efficacy and accuracy of MAML-BiGRU prediction model in predicting value of the situation.

5.7. *Overfitting and underfitting*

To ensure that our model does not overfit or underfit the data, we monitored the training and validation loss throughout the training process. Overfitting occurs when a model learns the training data too well, including its noise and outliers, which can reduce its ability to generalize to new data. Conversely, underfitting happens when a model is too simple to capture the underlying structure of the data.

- **Overfitting:** We employed techniques such as dropout and regularization to prevent overfitting. Additionally, we used early stopping, which halts the training process if the validation loss does not improve for a certain number of epochs, thus avoiding the model from learning the idiosyncrasies of the training data.
- **Underfitting:** To combat underfitting, we selected a model architecture that is sufficiently complex to capture the temporal dynamics of network security data. We also ensured that our model was trained for an adequate number of epochs.

5.8. *K-fold cross-validation results*

To ensure the robustness of our MAML-BiGRU model's predictions, we conducted a 5-fold cross-validation experiment.

The dataset was stratified and divided into five equal-sized subsamples. Each subsample served as the test set once, while the other four subsamples were used as the training set. This process was repeated five times, ensuring each subsample was used for testing. The MAML-BiGRU model was trained and evaluated across these folds, and the performance metrics were documented.

The cross-validation results are summarized in the table below:

Table 8
5-Fold Cross-Validation Performance Metrics

Fold	MAE	MSE	MAPE	R^2
1	0.123	0.318	3.45	0.897
2	0.115	0.282	3.21	0.905
3	0.132	0.346	3.68	0.886
4	0.118	0.299	3.35	0.898
5	0.127	0.315	3.52	0.890

Average Performance: MAE: 0.124 ± 0.008, MSE: 0.313 ± 0.025, MAPE: 3.47 ± 0.15, R^2: 0.897 ± 0.004.

The cross-validation results demonstrate that our MAML-BiGRU model provides consistent and accurate predictions across different subsets of the data. The low MAE and MSE values indicate that the model's predictions are close to the actual values on average. The MAPE values, which are below 4%, further confirm the model's accuracy in relative terms. The high R^2 values, which are close to 1, suggest that our model explains a significant portion of the variance in the network security situation predictions. The low standard deviation of the R^2 values across folds indicates that the model's performance is stable and reliable.

5.9. *In-depth interpretation of results*

Our results demonstrate that the MAML-BiGRU model outperforms other benchmark methods in network security situation prediction. This superior performance can be attributed to several key factors:

- **Rapid Adaptation:** The MAML component of our model enables rapid adaptation to new tasks with minimal additional training, which is crucial for responding to the evolving nature of network threats.
- **Temporal Dynamics Capture:** The BiGRU architecture effectively captures the temporal dynamics in network traffic, providing a more nuanced understanding of the sequence and patterns of security incidents.
- **Feature Relevance:** Our feature selection and extraction processes ensured that the model focuses on the most relevant aspects of the data, leading to more accurate predictions.

The implications of these findings are significant for the field of network security. By leveraging meta-learning and bidirectional recurrent structures, our approach offers a more proactive and adaptive solution to network security situation prediction. This not only enhances the capability to anticipate potential threats but also facilitates the implementation of timely and effective countermeasures.

Moreover, our study contributes to the broader understanding of how advanced machine learning techniques can be applied to complex, real-world problems. The success of our MAML-BiGRU model suggests that similar approaches could be beneficial in other domains where rapid adaptation and the handling of time-series data are critical.

6. Conclusion

A network security situation prediction approach integrating MAML and BiGRU is put forward. Using weekly reports of security information and dynamics from National Internet Emergency Response Center, we construct a network security situation prediction dataset and divide tasks. We introduce the BiGRU model to learn and train parameters on network security situation data's pre-post relationship and temporal order features. Combining BiGRU with MAML effectively improves the model's prediction performance for network security situation data with small sample learning. Through comparative experiments, this paper demonstrates the superiority and stability of model in network security situation data, surpassing general deep learning models in several metrics and proving the effectiveness of model put forward. Future work will focus on refining our model to handle larger and more diverse datasets, as well as exploring the integration of additional types of network traffic data to further enhance the prediction accuracy.

Acknowledgments

The work obtains support of National Science Foundation of China (61806219, 61703426, and 61876189), National Science Foundation of Shaanxi Provence (2021JM-226) by the Young Talent fund of the University, and Association for Science and Technology in Shaanxi, China (20190108, 20220106), and Innovation Capability Support Plan of Shaanxi, China (2020KJXX-065).

References

[1] Y.Q. Zhang, Q.X. Liu, A.M. Fu, G.H. Zhang, B.H. Chen, et al., Cyberspace governance in the new era, *Journal of Information Security Research* **7**(6) (2021), 486–487.

[2] CINIC, The 51st statistical report on the development of the Internet in China, Beijing, China: CINIC, 2023. [Online]. Available: https://www.cnnic.cn/NMediaFile/2023/0322/MAIN16794576367190GBA2HA1KQ.pdf

[3] CNCERT/CC, July 2022 internet security threat report, Beijing, China: CNCERT/CC, 2022. [Online]. Available: https://www.cert.org.cn/publish/main/upload/File/CNCERTreport202207.pdf

[4] S. Jajodia, P. Liu, V. Swarup, C. Wang, et al., Cyber situational awareness, US: New York, Springer, 2009.

[5] G.E. Box, G.M. Jenkins, G.C. Reinsel and G.M. Ljung, Time series analysis forecasting and control, China: Beijing, Posts & Telecom Press, 2005, pp. 19–180.

[6] W. Liang, Z. Chen, X.L. Yan, X.D. Zheng and P. Zhuo, Multiscale entropy-based weighted hidden Markov network security situation prediction model, in *2017 IEEE Int. Congress on Internet of Things (ICIOT)*, Honolulu, HI, USA, IEEE, 2017, pp. 97–104.

[7] S.M. Zhang, B.X. Li and B.Y. Wang, The application of an improved integration algorithm of support vector machine to the prediction of network security situation, *Applied Mechanics & Materials* **513–517** (2014), 2285–2288.

[8] M. Duan, Short-Time prediction of traffic flow based on PSO optimized SVM, in *2018 Int Conf. on Intelligent Transportation, Big Data & Smart City (ICITBS)*, Xiamen, China, IEEE, 2018, pp. 41–45.

[9] J. Schmidhuber, Evolutionary principles in self-referential learning on learning now to learn: the meta-meta-meta...–hook, Germany: Technische Universitat Munchen, 1987.

[10] C. Finn, P. Abbeel and S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, *in the 34th Int Conf on Machine Learning (ICML)*, Sydney, Australia, vol. 70, PMLR, 2017, pp. 1126–1135.

[11] T.T. Liu, X. Ma, S. Li, X.M. Li and C.M. Zhang, A stock price prediction method based on meta-learning and variational mode decomposition, *Knowledge-Based Systems* **252** (2022), 109323.

[12] L.S. Nie, X. Li, T.Y. Gong and D.C. Zhan, Few shot learning-based fast adaptation for human activity recognition, *Pattern Recognition Letters* **159** (2022), 100–107.

[13] H. Su, L. Xiang, A.J. Hu, Y.G. Xu and X. Yang, A novel method based on meta-learning for bearing fault diagnosis with small sample learning under different working conditions, *Mechanical Systems and Signal Processing* **169** (2022), 108765.

[14] Z.Z. Han, M.Y. Shang, Z.B. Liu, C.M. Vong, Y.S. Liu, et al., SeqViews2SeqLabels: learning 3D global features via aggregating sequential views by RNN with attention, *IEEE Trans on Image Processing* **28**(2) (2019), 658–672.

[15] T. Bass and D. Gruber, A glimpse into the future of ID, *The magazine of USENIX & SAGE* **24**(3) (1999), 40–49.

[16] L.Y. Shi, J. Liu, Y.W. Liu, H.Q. Zhu and P.F. Duan, Survey of research on network security situation awareness, *Computer Engineering and Applications* **55**(24) (2019), 1–9.

[17] Y.H. Chang, Z.R. Ma, X. Li and D.F. Gong, Survey of network security situation awareness, *Cyberspace Security* **10**(12) (2019), 88–93.

[18] CNCERT/CC, Weekly report on network security information from 2013 to 2022, Beijing, China: CNCERT/CC, 2022. [Online]. Available: https://www.cert.org.cn/publish/main/index.html

[19] R. Zhao, D.Z. Wang, R.Q. Yan, K.Z. Mao, F. Shen, et al., Machine health monitoring using local feature-based gated recurrent unit networks, *IEEE Trans on Industrial Electronics* **65**(2) (2018), 1539–1548.

[20] W.F. Jiang, Research on network security posture prediction based on multi-model weight extraction and fusion, Lanzhou: Lanzhou University of Technology, 2016.

[21] J.F. Sun, C.H. Li and B. Cao, Network security situation prediction based on TCN-BiLSTM, *Systems Engineering and Electronics*. [Online]. Available: http://kns.cnki.net/kcms/detail/11.2422.TN.20220922.0912.002.html

[22] C.S. Li, G. Tang, X.M. Xue, A. Saeed and X. Hu, Short-term wind speed interval prediction based on ensemble GRU model, *IEEE Trans on Sustainable Energy* **11**(3) (2019), 1370–1380.

[23] C.R. He and J. Zhu, Security situation prediction method of GRU neural network based on attention mechanism, *Systems Engineering and Electronics* **43**(1) (2021), 258–266.

[24] L. Munkhdalai, T. Munkhdalai, K.H. Park, T. Amarbayasgalan, E. Batbaatar, et al., An end-to-end adaptive input selection with dynamic weights for forecasting multivariate time series, *IEEE Access* **7** (2019), 99099–99114.